PHD THESIS:

Small area estimation. An application to the estimation of the labour market variables in Galician counties.

Author:
M. Esther López Vizcaíno

Supervised by:
M. José Lombardía Cortiña
Domingo Morales González
Antonio Vaamonde Liste

UNIVERSIDADE DA CORUÑA

Facultade de Informática

Departamento de Matemáticas

February, 2014

Los abajo firmantes hacen constar que son los directores de la Tesis Doctoral titulada ”Estimación en áreas pequeñas. Una aplicación a la estimación de las variables del mercado de trabajo en las comarcas gallegas”. desarrollada por M. Esther López Vizcaíno en el ámbito del programa de doctorado de Estadística e Investigación Operativa ofertado por el Departamento de Matemáticas de la Universidade da Coruña, dando su consentimiento para que su autora, cuya firma también se incluye, proceda a su presentación y posterior defensa.

16 de Enero de 2014

Directores:

M. José Lombardía Cortiña

Domingo Morales González

Antonio Vaamonde Liste
Doctoranda:

M. Esther López Vizcaíno

A Sabela, Daniel, Óscar
e aos meus pais

Agradecimientos

En primer lugar quiero agradecer a los directores de esta tesis, la profesora Mª José Lombardía Cortiña, el profesor Domingo Morales González y el profesor Antonio Vaamonde Liste, por su ayuda en el desarrollo de este trabajo y su estímulo constate, sin ellos este trabajo no hubiera sido posible. Quiero extender este agradecimiento al Departamento de Estadística e Investigación Operativa de Santiago, en especial a M José Ginzo, a Beatriz Pateiro, Manuel Febrero y Manuel Oviedo por la ayuda en diferentes momentos de esta investigación y al Departamento de Matemáticas de la Universidade da Coruña, en especial a Rubén Fernández por la ayuda para hacer el paquete mme y a Borja Lafuente por echarme una mano con el inglés.

Quiero agradecer al Instituto Galego de Estatística, y en particular al director José Antonio Campo Andión, por haberme facilitado los datos para realizar este trabajo. Merecen una mención especial en estos agradecimientos mis compañeros de servicio en el IGE: Esther, Begoña, Noa, Antonio, Miki, Pili, Ana y Andrés, que me ayudan a creer todos los días que la estadística puede ayudar a cambiar la sociedad que nos rodea y por la amistad y buen ambiente que me brindan. A los compañeros del Servicio de Informática por su ayuda con los ordenadores para realizar las simulaciones de este trabajo y por lo agradable que es trabajar con ellos. En general a todos los compañeros en el IGE con los que he trabajado en los últimos años, por tantos buenos momentos: Estefanía, Sara, Marta, Trini, Barriuso, Raúl, Carlos, Rosa, Esteban, Isa, Ángel, Carmen, Loli, Solmary, Manuel, ...

Nada de esto hubiera sido posible sin el ánimo de Óscar y la comprensión de Sabela y Daniel para dejarme trabajar un poco delante del ordenador mientras ellos jugaban. Muchas gracias también a mis padres por confiar en mi y por darme la formación que me ha permitido llegar hasta aquí. Gracias también a Soly, mi mejor amiga estadística, por todos los buenos momentos, estadísticos y no, que me ha brindado desde que la he conocido, y por extensión, a Rober y a Leire.

Resumen

El desempleo es actualmente uno de los problemas más importantes de nuestra sociedad. Las medidas globales de lucha contra el desempleo no suelen ser satisfactorias para llevar a cabo políticas efectivas a nivel local, de ahí la necesidad de herramientas que permitan disponer con precisión de información del mercado de trabajo a nivel local. En España la estimación de los indicadores del mercado laboral se hace mediante la Encuesta de Población Activa que está diseñada para obtener información a nivel provincial. Los tamaños de muestra por debajo del nivel de desagregación de provincia son bajos y los estimadores directos en las comarcas o en los municipios suelen tener poca precisión. Por tanto, el objetivo de esta tesis es la estimación de indicadores del mercado laboral, tales como ocupados, parados y tasas de paro, en las comarcas gallegas. Utilizamos técnicas de estimación en áreas pequeñas con modelos multinomiales mixtos. El primer modelo que utilizamos para la estimación de los indicadores laborales es un modelo multinomial mixto a nivel de área basado en Molina et al. (2007). Estos autores consideraron un efecto aleatorio común para las dos categorías multinomiales (ocupados y parados). En el problema real esta situación no es apropiada por las distintas características de estos dos grupos poblacionales. Esta es la razón por la cual nosotros introducimos en el modelo dos efectos aleatorios, uno para cada una de las categorías multinomiales. Además, la disponibilidad de encuestas para distintos periodos de tiempo produce un importante incremento en la muestra en las áreas y nos permite introducir en el modelo efectos aleatorios independientes y correlados a lo largo del tiempo. La estimación de la precisión en los estimadores de áreas pequeñas es fundamental porque a menudo estos son sesgados. En este trabajo utilizamos diferentes métodos para la estimación del error cuadrático medio, mediante expresiones analíticas y mediante técnicas bootstrap.

Abstract

Unemployment is currently one of the most important problems of our society. Global measures to fight against unemployment are usually not satisfactory to carry out effective policies at the local level, hence the need of tools to provide accurate labor market information at local level. In Spain the estimation of labor market indicators is made by means of the Labour Force Survey that is designed to obtain information at the provincial level. Sample sizes below the provincial level of disaggregation are low and direct estimators in the counties or municipalities often have low precision. Therefore, the aim of this thesis is to estimate labor indicators, such as employed, unemployed and unemployment rates, in Galician counties. We use small area estimation techniques under area level multinomial mixed models. We will first use estimators based on the area level multinomial mixed model introduced by Molina et al. (2007). These authors considered a model with a common random effect for the two multinomial categories (employed and unemployed people). In the real data problem this may not be appropriate because of the very different characteristics of these two populations. This is the reason why we introduce in the model two random effects, one for each of the multinomial categories. In addition, the availability of surveys for different periods of time produces a significant increase of the domain samples and allows us to introduce in the model independent and correlated time random effects. The estimate of the accuracy of the estimators of small areas is a fundamental issue because these estimators are often biased. In this work we use different methods for estimating the mean squared error, by using analytical expressions and using bootstrap techniques.

Resumo

O desemprego é actualmente un dos problemas máis importantes da nosa sociedade. As medidas globais de loita contra o desemprego non adoitan ser satisfactorias para levar a cabo políticas efectivas a nivel local, de aí a necesidade de ferramentas que permitan dispor con precisión de información do mercado de traballo a nivel local. En España a estimación dos indicadores do mercado laboral faise mediante a Enquisa de Poboación Activa que está deseñada para obter información a nivel provincial. Os tamaños de mostra por debaixo do nivel de desagregación de provincia son baixos e os estimadores directos nas comarcas ou nos municipios adoitan ter pouca precisión. Polo tanto, o obxectivo desta tese é a estimación de indicadores do mercado laboral, tales como ocupados, parados e taxas de paro, nas comarcas galegas. Utilizamos técnicas de estimación en áreas pequenas con modelos multinomiais mixtos. O primeiro modelo que utilizamos para a estimación dos indicadores laborais é un modelo multinomial mixto a nivel de área baseado en Molina et al. (2007). Estes autores consideraron un efecto aleatorio común para as dúas categorías multinomiais (ocupados e parados). No problema real esta situación non é apropiada polas distintas características destes dous grupos poboacionais. Esta é a razón pola cal nós introducimos no modelo dous efectos aleatorios, un para cada unha das categorías multinomiais. Ademais, a dispoñibilidade de enquisas para distintos períodos de tempo produce un importante incremento da mostra nas áreas e permítenos introducir no modelo efectos aleatorios independentes e correlados ao longo do tempo. A estimación da precisión nos estimadores de áreas pequenas é fundamental porque a miúdo estes son nesgados. Neste traballo utilizamos diferentes métodos para a estimación do erro cuadrático medio, mediante expresións analíticas e mediante técnicas bootstrap.

Contents

Chapter 1

Introduction

The study of the national labour market is a critical issue in every kind of analysis of the economic structure of a country. It affects to different aspects of productive sectors and, in fact, it is a valuable aid for sectorial analysis. Therefore, the knowledge of the balances and imbalances between supply and demand in the labour market is important for economical researchers (Pérez-Infante, 2006).

Unemployment is one of the most important problems nowadays. Sociological surveys usually place it as one of the main concerns of citizens, and the fight against unemployment is a priority of the political action at all levels of Public Administration. Moreover, in the context of the crisis in which the European Union is involved, the impact on the Spanish labour market has been much more intense than in most advanced economies. The Spanish unemployment rate in the third quarter of 2013 reached 25.98%, more than 14 percentage points higher than in 2008. The labour market situation in Galicia is not very different, the unemployment rate is 21.6% and the number of unemployed people has reached 277000. At this point, politicians at all levels of the administration are planning and acting toward reducing unemployment.

Global political measures are not often satisfactory for local authorities, which can also develop their own strategies for employment. They need some tools to determine, with precision, reliability and acceptable punctuality, the main variables and labour market indicators in order to implement their strategies. Among the main labour market indicators, we can cite the totals of employed, unemployed and inactive people, the employment, unemployment and occupation rates, and the corresponding disaggregations by gender, age and economic activity.

In Spain, like in other European countries, the estimation of labour market indicators is made by means of the Labour Force Survey (LFS). The Spanish Labour Force Survey (SLFS) uses a stratified sampling design. The stratification variable is the size of the municipality (INE, 2009). As most municipalities are not represented in the sample and many of them are present with a very small sample size, the estimates at the municipal level are not accurate enough. Small sample sizes and, in some cases, no sample at all is the main problem when performing municipal estimations. In this situation the sample size could be enlarged but this, in addition to cause an increase in the costs and in the denial by the respondents to answer the sample questionnaire, can lead to other kind of damages due to delays in obtaining results and to the impact of non-sampling errors. Therefore the increase of the sample size is not always advisable and even sometimes unfeasible from an economic point of view.

The interest in developing small area estimation techniques to solve these problems in a reasonable way is growing among statisticians. The term "small area" is often used to refer to geographic areas but it can also be applied to other interesting areas with non geographical boundaries (domains), like age groups, economic activity sectors and so on. It is the small sample size in the domain, and consequently the large variance of the direct estimators, the key point defining the concept of small area. It is not the actual size of the area. In the small area estimation context, an estimator of a parameter in a given domain is direct if it is based only on the sample data of the specific domain. A drawback of these estimators is that they can not be calculated when there is no sample observations in an area of interest.

Small area estimation (SAE) is a part of the statistical science that combines sample information and inference in finite populations with statistical models. The use of small area statistics was originated centuries ago. Brackstone (1987) mentions the existence of these statistics in the eleventh century in England and in the seventeenth century in Canada. These early statistics were based on censuses and administrative registers. Over time, the sampling has been replacing the census as the most effective and economical tool to obtain information on a wide range of topics.

Early attempts of small area estimations with survey data can be found in the classic text Hansen et al. (1953), p.483-486, where regression estimators were proposed. Nevertheless, the popularity of these estimators is basically due to Ericksen (1974). The first reviews on small area estimation emphasize demographic methods for estimating the population in post-census periods, at this point Morrison (1971) writes a review of the demographic methods that exist before 1970. Purcell and Kish (1979) reviews demographic and statistical methods for estimation in small domains. Later, Zidek (1982) introduces a criterion to evaluate the relative performance of different methods for estimating the population in small areas. Platek et al. (1987) summarizes some techniques and applications, Rao (1986) and Chaudhuri (1994) show traditional techniques and methods, which at that time were more recent. Schaible (1996) reviews traditional and indirect estimators based on the models that were used in U.S. Federal programs. Also, a large number of conferences and workshops took place, a list of which is presented in the review of Ghosh and Rao (1994) and Rao (2003). A large number of reviews were published, such as Rao (1986, 1999, 2001), Chaudhuri (1994), Ghosh and Rao (1994), in addition to Marker (1999), Pfeffermann (2002), Lahiri and Meza (2002), Jiang and Lahiri (2006) and Pfeffermann (2012). Two books were written: Mukhopadhyay (1998); Rao (2003). All these activities have helped to spread the research in small area estimation and also they helped to understand the methodology and applications.

Generally small area estimation techniques can be divided into design-based methods and model-based methods. Model-based methods are used both from a frequentist point of view and from a Bayesian perspective. In some cases, as a combination of the two approaches, which in SAE literature is known as "empirical Bayes". Design-based methods often employ a model for the construction of the estimators, but the bias, variance and other properties of the estimators are calculated taking into account the probability distribution induced by the sampling design. However, the model-based methods make inference by taking into account the underlying model. The estimators based on these methods are useful because they give to practitioners an idea of how the data generation process is and how the different sources of information are incorporated.

Although these two methods have differences, they also have common characteristics. In both cases, they use auxiliary information obtained from samples and administrative or census registers. The use of auxiliary information in SAE is critical, because in many cases we find very small sample sizes. In small area estimation studies, even the most elaborate model will not produce accurate estimates if it cannot be feeded with proper auxiliary information about the target variable.

Depending on the availability of auxiliary information, small area models are classified into two main types:

- Area level models: when auxiliary information is available only at the area level. These models link the direct area estimator with an area specific covariate. It is assumed that the area-level direct estimator follows a population model.

- Individual level models: when auxiliary information is available on the units of the population. These models link the target variable with the covariates at the unit level. Unit-level data is assumed to be a realization of a population model and the sampling design distribution is not taken into account.

Model selection and validation play a vital role in the estimation process, because the properties of the model-based estimators rely on the assumed model distribution.

Mixed models are suitable for small area estimation due to its flexibility to make an effective combination of different sources of information and to its capacity to describe the various sources of error (Searle et al., 1992). These models incorporate random area effects that explain the additional variability that is not explained by the fixed part of the model. In addition, small area parameters such as the mean or the total can be expressed as a linear combination of fixed and random effects and they can be estimated by means of best linear unbiased predictors (BLUP). The BLUP minimizes the mean square error inside the class of linear unbiased estimators. Its derivation does not require the normality of the random effects and errors, but it assumes the existence of their second order moments. The model variances and covariances can be estimated, for example, by the maximum likelihood method (ML) or the restricted maximum likelihood method (REML). By plugging the estimations of these variance components in the expression of the BLUP we obtain the Empirical BLUP (EBLUP).

Mixed models have been used in the U.S. to estimate the per capita income in small areas (Fay and Herriot, 1979), the not included population in the census (Dick, 1995; Ericksen and Kadane, 1985) and the poverty in schoolchildren studies (Council, 2000). It is important to mention that using these estimators, the Department of Education of the United States allocates more than 7000 million dollars in general funds to the counties, and then the states distribute these funds among school districts (Rao, 2003). The use of these techniques is not restricted to socioeconomic data. The work of Battese et al. (1988) is an example of application in the field of agriculture using a linear mixed model to estimate the area under cultivation of corn and soy in 12 counties of North-Central Iowa. The inclusion of area random effects in the model is also a common practice in the literature of SAE (Herrador et al., 2009; Molina et al., 2007; Rao and Yu, 1994; Saei and Chambers, 2003). The random effects model the variability across the areas that is not explained by the auxiliary variables and, additionally, allow the correlation between them.

The objective of this thesis is the estimation of labour market indicators (totals employed and unemployed people and unemployment rate) in the counties of Galicia using small area estimation techniques under area level multinomial mixed models. The totals of unemployed and employed people can be estimated by using two separate linear mixed models relating the direct estimations of the respective proportions with other auxiliary variables. In that case we cannot ensure that the estimated proportions are in the interval $[0, 1]$, which is an important disadvantage. Another disadvantage is that these models do not take into account the natural relationship between the unemployed, employed and inactive population, since the sum of totals of the three categories is the total of the population aged 16 and over. These disadvantages can be overcome by using multinomial logistic models. These models have been discussed in the literature. See for example, Saei and Chambers (2003), Molina et al. (2007), Morales et al. (2007) and González-Manteiga et al. (2008b).

For estimating labour indicators we will first use estimators based on the area level multinomial logit mixed model introduced by Molina et al. (2007). These authors considered a model with a common random effect for the two multinomial categories (employed and unemployed people). In the real data problem we are dealing with, this may not be appropriate because of the very different characteristics of these two groups in Galicia. Alternatively, we introduce models with two random effects, one for each of the multinomial categories. Furthermore, the availability of time series produce a significant increase of the domain samples and led us introducing in the model independent and correlated time effects. This last idea is not new and has been developed in some papers (Pfeffermann and Burck, 1990; Rao and Yu, 1994; Saei and Chambers, 2003; Tiller, 1992; Ugarte et al., 2009a). As we will see, these models are naturally suited to the problem of interest, overcoming the disadvantages of previous proposals and allowing a simultaneous estimation of totals of employed, unemployed and inactive people. In this case, area-level models will be applied to the sample data. To illustrate the inferential process, Galicia's LFS data is used.

The estimation of the accuracy of the EBLUP is a fundamental issue in SAE because these estimators are often biased. Several approaches have been published in the literature. The first MSE simplification was obtained by Kackar and Harville (1981) assuming normality in the errors and in the random effects. In a second paper Kackar and Harville (1984) obtained an approximation of the MSE and proposed an estimator based on it. Prasad and Rao (1990) gave a new approach to covariance block diagonal matrix. Datta and Lahiri (2000) obtained analogous MSE estimator for models with covariance block diagonal matrices and variance components estimated by ML or REML. More recently Das et al. (2004) studied the prediction error approach in a wider class of models when the variance components are estimated using ML or REML methods. When there are not suitable estimators, the best option is to use resampling methods. Jiang et al. (2002) using jackknife give asymptotically unbiased estimators specifying the order of consistency. Pfeffermann and Tiller (2005) introduce parametric and nonparametric bootstrap procedures in order to estimate the MSE in state space models. Hall and Maiti (2006) present double bootstrap algorithms, González-Manteiga et al. (2008b) apply bootstrap procedures in mixed logistic regression models at area level and González-Manteiga et al. (2008a) introduce bootstrap procedures in mixed linear models at individual level. Their simulations show a reduction of bias compared with other estimators.

The fact that the estimators are biased has to be complemented with an accuracy gain. Hence, in this thesis different approaches to estimate the mean square error (MSE) are used, first through an analytical expression and second by bootstrap techniques. At this point it is desirable to take into account that in the statistics of labour the Office for National Statistics (ONS) in the UK considers that an estimate is publishable, and therefore official, if the coefficient of variation is less than 20% (ONS, 2004).

The remainder of the thesis is organized as follows. Chapter 2 describes the data sources on employment and unemployment in Spain and the direct estimators obtained from these sources. Chapter 3 introduces the general theory of the multivariate generalized linear mixed models (MGLMM) that we use in this work. Chapter 4 proposes a first multinomial mixed model and presents a simulation study and an application to LFS data from Galicia for the fourth quarter of 2008. Chapter 5 expands the model of Chapter 4 to include independent temporal effects and the corresponding application to data from the third quarter of 2009 to the fourth quarter of 2011. Chapter 6 considers correlated time effects, carries out the corresponding simulation study and gives the application to data for the period between the third quarter of 2009 and the fourth quarter of 2011. Chapter 8 presents the final conclusions. Finally, Chapter 7 gives a description of the mme package developed in the R programming language. This package implements the models discussed throughout this dissertation.

Chapter 2

Statistical sources for measuring the labour force: employment and unemployment

2.1 The Labour Force Survey

In Spain there are only two regular statistics that measure simultaneously the different situations of the labour force in the labour market (active or inactive people and employed or unemployed people): The Population Census and the Labour Force Survey (LFS), both conducted by the Spanisgh National Institute of Statistics (INE).

The Population Census is a comprehensive and decennial statistic operation, which since 1981, is elaborated in the years ending in one (until 1970 was made in the years ending in 0). The target population of the census is the whole population living in the country. The variables related to the labour market considered in population censuses are the active and inactive population, employed and unemployed people classified by variables such as sex, age, professional status, studies, economic sectors and with a geographical breakdown reaching the census section level.

Despite the enormous advantages at obtaining information with the Census, it has two major problems. The information obtained from the Census is very expensive and consists on a collection of demographic, social and/or economical variables. The time reference does not have, in many cases, the necessary update for users because of, among other reasons, the delay in obtaining results due to the large size of the operation.

Although it is still useful for analyzing long-term trends in the labour market, the Census is not the appropriate instrument to study the current situation and the evolution of this market. In addition, until the last census of 2011 the respondent defines him/herself as active, inactive, employed or unemployed. This self-assessment conducted by the respondents completing the questionnaire may be influenced by subjective criteria, since they had no objective definitions. This is not the actual case, where the determination of the labour classifications is done by the set of rules established by the International

Labour Organisation (ILO).

For all the above reasons, LFS is the only existing Spanish statistic with suitable periodicity to carry out the study of the labour market situation and evolution. It also permits international comparability because it uses ILO criteria for classifying people in different situations related to the activity. The LFS is a sample survey conducted by the INE since 1964 and is addressed to the population living in the country. From that date, and until the end of 1968, results were obtained quarterly. From 1969 to 1974, results were obtained every six months. Since 1975, it became quarterly once again.

In 1987, the survey questionnaire was modified and adapted to the latest international recommendations given in the International Conference of Labour Statisticians in Geneva 1982. As one year before Spain became member of the European Community, at that year the Spanish LFS was also adapted to the European Community Labour Force Survey. Further, retrospective series were recalculated with the new methodology beginning in the third quarter of 1976. This is the origin of the so-called "homogeneous series" of the LFS that is currently offered to users of INEBASE and other supports via the final survey files available from that period and up to date.

Until 1999 the Labour Force Survey interviews were conducted during 12 weeks of the 13 weeks of each quarter. Since 1999 the Labour Force Survey became a continuous survey, given that the interviews were conducted throughout the whole period of 13 weeks. In 2002, a new operative definition of unemployment was introduced, producing a break in the series of unemployed and active people. The impact of this modification was calculated by compiling a double estimation of both definitions throughout the year 2001.

Finally, the last substantial methodological change up to date was produced in 2005. It introduced a new questionnaire and a centralised control of the collection system, via a computer-assisted telephone survey. On that year, part of the survey variables were collected exclusively in an annual sub-sample that was representative of the average situation for the year, instead of obtaining them quarterly. These are the so-called annual sub-sample variables, and the corresponding results are disseminated annually. With the purpose of maintaining the homogeneity of the estimates, retrospective series were also calculated in 2005 for the period 1996-2004. These calculations were done under the new population base established for 2005. No more variations has been introduced since 2005. The current figures of the survey are encompassed in the methodology founded in 2005.

The main objective of the LFS is to reveal information about economic activities as regards their human component. It focuses on providing data on the main population categories related to the labour market (employed, unemployed, active and inactive population) and obtaining classifications of these categories depending on different characteristics. It also allows the creation of homogeneous time series of results. Finally, since all definitions and criteria used are in line with those established by international organisations dealing with labour-related topics, all data can be compared with information of other countries.

Detailed results are available for the whole country. As regards autonomous communities and provinces, information is provided on the main characteristics with the level of breakdown allowed by the variation coefficient of the estimators.

The LFS is aimed at the population living in family dwellings, namely, those used throughout the whole year (or most of it) as the habitual or permanent dwelling. Therefore, the survey does not consider group dwellings (hospitals, residences, barracks, etc.) or secondary or seasonal dwellings (used during holiday periods, at weekends, etc.). The survey does include families that, forming an independent group, reside in said group establishments (for example, the director or caretaker of the centre).

The reference period for the results of the Survey are quarters. The reference period for the information is the week (Monday to Sunday) just before the interview according to the calendar. The answers to the questionnaire will, therefore, always refer to said week. Nevertheless, some questions have special reference periods, such as seeking work methods, peculiarities of the working day and questions about studies refer to the four weeks prior to the interview and the availability to work refers to the two weeks subsequent to the Sunday of the reference week.

After the inclusion of Ceuta and Melilla in the second quarter of 1988, the LFS covers the whole of the Spanish territory.

Definitions are based on the recommendations endorsed by the ILO:

- Active population: All persons 16 years old and older who, during the reference week (week prior to the interview according to the calendar), fulfil all the conditions required to be included among the employed or unemployed persons, as defined below.

- Employed population: All persons 16 years old and older who, during the reference week, either were employed by others or performed freelance work, according to the next definitions:

  - Persons employed by others or wage-earners are all persons described in the following categories:

    * Working: Persons who worked for at least one hour during the reference week, even sporadically or occasionally, in exchange for a salary, wages or another form of remuneration in cash or in kind.
    * Employed but not working: Persons who, having worked in their current job, were absent from said job during the reference week, but are closely tied to the job.

  - Freelance or self-employed workers are all persons included in the following categories:

    * Working: Persons who worked for at least one hour during the reference week, even sporadically or occasionally, in exchange for personal gain or family earnings, in cash or in kind.
    * Employed but not working: Persons who should have worked during the reference week in exchange for personal gain or family earnings, but were temporarily absent from work, due to illness or accident, holidays, public holidays, bad weather or other similar reasons. According to this definition, the following persons perform freelance work: entrepreneurs, independent workers, members of cooperatives who work in said cooperatives and unpaid family workers (family assistance).

- Unemployed population: Unemployed persons are those persons 16 years old or older who combine the following conditions simultaneously:

  - without work, in other words, who have not been employed by others or have not freelanced during the reference week.
  - seeking work, in other words, who have taken specific measures to look for work employed by others or who have performed procedures to set up as freelancers during the previous month.
  - available to work, in other words, in conditions to start working within two weeks from Sunday of the reference week.

  Unemployed persons are also persons 16 years old and older who were without work during the reference week, who are available to work and who were not seeking work because they have found a job which they would be starting in the three months following the reference week. This case does not require the effectively seeking work criterion.

- Economically inactive population: comprises all persons 16 years old and older who do not classify as employed, unemployed or population counted separately during the reference week. This definition covers the following functional categories:

  - Persons who perform household chores: persons who perform household chores without performing an economic activity; for example, housewives and other family members looking after houses and children.
  - Students: persons who receive systematic instruction in any degree of education without performing an economic activity.
  - Retired or pre-retired persons: persons who have had a previous economic activity and who because of their age or other reasons have abandoned it, thereby receiving a pension (or some pre retirement income) because of their previous activity.
  - Persons currently perceiving a pension other than a retirement or pre retirement income.
  - Unpaid persons who perform social work, charitable activities, etc. (excluding family assistance).

  - Incapacitated to work.
  - Another situation: persons who, without exercising any economic activity, receive public or private aid and all those who are not included in any of the previous categories, for example the independently wealthy.

- Unemployment rate: Quotient between the number of unemployed persons and the active population.

Moreover, all defined characteristics refer to the national concept, not to the domestic concept, in line with the definitions of the European System of National and Regional Accounts (ESA-95, EUROSTAT-1996). This is due to the fact that information cannot be collected for the population who work in Spain and live abroad, since the Survey is aimed at the population resident in family dwellings on the Spanish territory.

The sample size is about 4000 dwellings in Galicia until the second quarter of 2009, which are uniformly distributed throughout the 13 weeks of each quarter (every week is interviewed the thirteenth part of the sample). In 2008, a collaboration agreement was signed between the INE and the Galician Statistics Institute (IGE) increasing the sample size in the Autonomous Community of Galicia. The goal of this sample increase is to allow the population analysis related to the economical activity with further breakdown of what was done until that time. As a result of this agreement, in the third quarter of 2009 additional sample is included, collected by the IGE (234 sections) using the same methodology and the same system of fieldwork that INE. In the second quarter of 2009 the sample in Galicia is about 8000 households.

The survey uses a two-stage sampling with primary sampling units stratification. Primary sampling units (PSU) are composed by census sections, which are geographical areas with a maximum of 500 dwellings or about 3000 people. Secondary sampling units are composed by main family dwellings and permanent accommodations. Sub-sampling is not carried out in secondary sampling units, information is collected on all persons who regularly live in the same.

Census sections are grouped in strata, according to the province and the type of municipality considering the following classification:

1. Stratum 1: Province capital municipality.
2. Stratum 2: Self-represented municipalities, important areas compared to the capital or municipalities with more than 100.000 inhabitants.
3. Stratum 3: Municipalities between 50.000 and 100.000 inhabitants.
4. Stratum 4: Municipalities between 20.000 and 50.000 inhabitants.
5. Stratum 5: Municipalities between 10.000 and 20.000 inhabitants.
6. Stratum 6: Municipalities between 5.000 and 10.000 inhabitants.

7. Stratum 7: Municipalities between 2.000 and 5.000 inhabitants.
8. Stratum 8: Municipalities under 2.000 inhabitants.

Census sections are grouped in substrata inside each strata, according to socioeconomic conditions.

Sample selection is performed to ensure that in each stratum all family dwellings have the same probability of being selected, in other words, to ensure that there are self-weighted samples in each stratum. First stage units are selected with a probability proportional to the number of main family dwellings. In each section selected in the first stage, a pre-set number of family dwellings with the same probability is selected by implementing a random start systematic sample. For this survey, 18 dwellings have been selected per section.

Let P be the population of individuals aged 16 years old and older residing in family households in Galicia. This population is divided in 4 provinces P_p, $p = 1, 2, 3, 4$. Each province is divided into 9 strata, denoted by P_{ph}, $p = 1, 2, 3, 4$, $h = 1, \ldots, 9$. The samples are independent in each province, therefore we simplify notation and used P_h. In addition, the provinces are divided into domains P_d, defined by sex groups and regions. These domains are not always nested in strata. Let S be the full sample and S_p, S_h, and S_d the sub-samples inside the province p, stratum h and domain d, respectively. Let V_{ha} be the number of dwellings in the PSU a and in the stratum h, V_h is the number of dwellings in the stratum h and m_h is the number of PSU selected in the stratum h. Therefore, the probability of selection of each dwelling V of the PSU a, belonging to the stratum h, can be calculated by

$$P(V_{ha}) = P(PSU_{ha})P(V_{ha}|PSU_{ha}) = m_h \frac{V_{ha}}{V_h} \frac{18}{V_{ha}} = \frac{18m_h}{V_h}.$$

As all residents in the dwelling aged 16 years old or older are interviewed, the probability π_j of selecting an individual j of dwelling v coincides with the probability of selecting the dwelling v. From the above formula it follows that this probability is constant within each stratum. Thus, the selection probability and the sample weights of the individual j in the stratum h are

$$\pi_j = \frac{18m_h}{V_h} \, w_j^{(1)} = \frac{1}{\pi_j r_h} = \frac{V_h}{18 m_h r_h} \cong w_h^{(1)}, \ \forall j \in S_h.$$

where r_h is the relative frequency of responses in the stratum h.

2.1.1 Estimators provided by the LFS

Let N_h be the size of the population aged 16 years old or older in the stratum h, according to the population projections given by the INE, and let n_h be the number of individuals of the sample in the stratum h. Until the year 2001, the total Y_p of a province p was calculated by using the ratio estimator

$$\widehat{Y}_p^{LFS} = \sum_{h \in P_p} \frac{N_h}{\widehat{N}_h} \sum_{v \in S_h} \sum_{j \in v} w_j^{(1)} y_j, \quad \text{where } \widehat{N}_h = \sum_{v \in S_h} \sum_{j \in v} w_j^{(1)} = w_h^{(1)} n_h.$$

The ratio estimator can also be written as a weighted sum of the values y_j, i.e.

$$\widehat{Y}_p^{LFS} = \sum_{j\in S_p} w_j^{(2)} y_j, \quad \text{where } w_j^{(2)} = \frac{N_h w_h^{(1)}}{\widehat{N}_h} = \frac{N_h}{n_h}, \forall j \in S_h.$$

Since the year 2002, calibration techniques are applied to the weights $w_j^{(2)}$ (Deville and Sarndal, 1992) in order to adjust the estimates of the survey with the information from external sources. New weights $w_j^{(3)}$ are obtained by minimizing the sum of the weighted differences between the old weights $w_j^{(2)}$ and the new ones $w_j^{(3)}$. This is to say, by minimizing

$$\sum_{j\in S} w_j^{(2)} G(\frac{w_j^{(3)}}{w_j^{(2)}})$$

in $w_j^{(3)}$, subject to

$$\sum_{j\in S} w_j^{(3)} x_{jk} = X_k, k = 1, \ldots, K,$$

where X_k, $k = 1, \ldots, K$ are known population quantities. INE uses the disparity function

$$G(z) = \begin{cases} \frac{1}{2}(z-1)^2 & \text{if } 0.1 \le z \le 10, \\ \infty & \text{otherwise} \end{cases}$$

Calibration equations match the totals of K known population variables with the corresponding weighted sums of the elements of the sample. The calibration variables are category indicators, so that restrictions are used to match the sum of the calibrated weights to the population sizes of

1. the sex-age groups in the autonomous region, with the age groups $16-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, \ge 65$, and
2. the provinces.

To simplify the notation, the final calibrated weights will be denoted by $w_j = w_j^{(3)}$, $j \in S$, so that the final expression of the estimator of totals Y_p in the p province is

$$\widehat{Y}_p^{LFS} = \sum_{j\in S_p} w_j y_j$$

In Galicia there are 53 counties but in this thesis we consider $D = 96$ domains. For all the quarters preceding the third quarter of 2009, the considered domains are obtained from crossing the 48 counties represented in the sample with the two sexes. Since the third quarter of 2009, $D = 102$ domains will be considered because on that period there were 51 counties with sample representation.

We divide the D domains P_d into subsets P_{d1}, P_{d2} and P_{d3} of employed, unemployed and inactive people. Our goal is to estimate the totals of employed and unemployed people and the unemployment rate, defined by

$$Y_{dk} = \sum_{j\in P_d} y_{dkj}, \quad R_d = \frac{Y_{d2}}{Y_{d_1} + Y_{d2}}, \qquad k = 1, 2,$$

where $y_{dkj} = 1$ if the j individual of the d domain is in P_{dk} and $y_{dkj} = 0$ in other case.

Official LFS data are not related to the domains (counties $\times$ sex) but the analogue direct estimators of the total Y_{dk}, the mean $\bar{Y}_{dk} = Y_{dk}/N_d$, the size N_d and the unemployment rate R_d are

$$\hat{Y}_{dk}^{dir} = \sum_{j \in S_d} w_{dj}\, y_{dkj}, \ \hat{\bar{Y}}_{dk}^{dir} = \hat{Y}_{dk}^{dir}/\hat{N}_d^{dir}, \ \hat{N}_d^{dir} = \sum_{j \in S_d} w_{dj}, \ \hat{R}_d^{dir} = \frac{\hat{Y}_{d2}^{dir}}{\hat{Y}_{d1}^{dir} + \hat{Y}_{d2}^{dir}}, \quad k = 1, 2. \tag{2.1.1}$$

where S_d is the sample domain and w_{dj}'s are the official calibrated sample weights that take into account the non-response.

2.2 Employment measurement: Registrations in the social security system

Besides LFS, which periodically estimates the unemployment and employment totals, there is another source of statistical information that provides regular information on the occupation: people registered in the social security system. The social security system (SSS) register is an administrative register that is not intended to estimate direct employment but people registered and paying to Social Security. It is published monthly with the data referring to the last day of each month and the average of the month. In this work we use the data at the last day of the month.

The registration data shows the number of workers enlisted, doing a labour activity (excluding unemployed or students) and are therefore forced to contribute to the public system for the protection of at least retirement situations, disability and death. The registration data, as all data from registers, is sensitive to law changes that affect it. The file can also contain fictitious contributors looking to get a pension in the future or the right to receive unemployment insurance. There are people employed according to LFS that are not required to pay into Social Security, as is the case with some family assistance (this category of workers must be registered in the Social Security only when they work daily, fundamentally and directly). There are also people who are required to be registered in the Social Security and they are not.

The registration file does not include government employees registered with their own mutual funds that do not perform any other activity that determine inclusion in any of the schemes of the Social Security System. It also does not include students who do not perform a work activity. Subsidized temporary agricultural workers are included in the file whether they work or not that month. The papal clergy is registered in the social security system but does not develop an economic activity in the strict sense, as discussed in the LFS. Workers with special agreement do not work and, although they are members in order to obtain benefits, with few exceptions, are not counted as such. Furthermore, it should be noted that when referring to statistical registrations and not the people registered, people with multiple jobs are considered more than once.

Although the two statistics, LFS and SSS, have important differences. They also have many similarities which will make us to use the variable registered in the social security system as an auxiliary variable in the models proposed in the following chapters.

2.3 Unemployment measurement: Registered unemployment

The two most important sources about unemployment in Spain are LFS and people registered as unemployed in the administrative register of employment claimants. There are important differences between the two sources, which can be of methodological or conceptual type and of dissimilarity of contents or considered population groups. Therefore the two data sources produce significantly different estimates of labour force indicators.

Registered unemployment consists of the total employment demands, registered by the administrative register of employment, existing on the last day of each month, excluding the ones that correspond to work situations described in the Ministerial Order of 11 March 1985 (B.O.E. of 14/3/85) by introducing statistical criteria for measuring registered unemployment.

Labour demand is the application for a job that makes a worker, unemployed or not, to an administrative register of employment. For measuring the registered unemployment, the unemployment register excludes all demands that, at the end of the reference month, are in any of the following situations:

- Claimants seeking other employment compatible with their current job (moonlighting). They are also excluded of the LFS unemployment.
- Claimants who, being employed, apply for a job to change it. They are also excluded of the LFS unemployment.
- Claimants with unemployment benefits participating in Social Collaboration works. They are also excluded of the LFS unemployment.
- Claimants who are retirement pensioners, pensioners with a huge absolute disability and claimants aged 65 years or older (Retired). They are not excluded from LFS unemployment.
- Claimants seeking employment for a period less than three months. They are not excluded from LFS unemployment.
- Claimants applying for a time working less than 20 hours a week. They are not excluded from LFS unemployment.
- Claimants who are studying for formal education if they are under 25 years old or overcoming this age are seeking their first job. They are not excluded from LFS unemployment.

- Claimants attending to occupational training courses when their teaching hours exceed 20 a week, with at least a scholarship to support and be seeking their first job. They are not excluded from LFS unemployment.

- Claimants with suspended request while remaining in this situation since the suspension of the request, usually processed at the request of applicant and good cause, interrupts the job search. If demand is suspended for not being available for work, they would not be LFS unemployed.

- Claimants with unemployment benefits which are compatibles by having a part-time job. They are also excluded from LFS unemployment.

- Claimants who are receiving the farming benefit or who, having it spent, has not elapse a period longer than one year from the date of entitlement (Benefited Agricultural Temporary Workers). They are not excluded from LFS unemployment, unless they had been working.

- Claimants who refuse job placement activities suited to their characteristics, as provided in Article 46, 1.2 of Law 8/88 of 7 April. The exclusion of LFS unemployment depends on the characteristics of each person included in this group.

- Claimants not immediately available for work or in incompatible situation with it, as claimants in situation of temporary disability or sick leave, claimants who are performing military service or alternative service, claimants registered to participate in a selection process for one particular job, job claimants exclusively for overseas, claimants of a home-based work, claimants who, under an employment regulation file, are in situation of suspension, and other causes.

  If the claimants are available to work within 15 days, they would be LFS unemployed

As both data sources use different definitions of employment and unemployment, they give different estimations of unemployment rates. On the one hand, in the LFS people can seek employment in various ways, not only by registering in the administrative register of employment; this fact would expect the LFS figure to be higher. On the other hand, enrollment in the administrative register of employment within the registered unemployment community does not include other conditions that are indeed necessary in the LFS to assign the condition of unemployed (such as availability for working in the next two weeks). This would favor a higher number of registered unemployment. The registered unemployment, like all figures from administrative registers, is strongly influenced by changes in legislation that affects it. Also the potential attractiveness of registration (ability to receive benefits, to receive training courses, etc.) has influence on registered unemployment figures. Moreover, data from the LFS, like any survey, reflect the statements made by the interviewees, which in some cases may not correspond to reality. Their estimates have also the associated sampling error inherent to any survey.

An approximation to the analysis of LFS unemployment can be done through registered unemployment. This is the reason why, in the models considered in Chapters 3, 4 and 5, this variable is included as an auxiliary variable for explaining of unemployment.

Chapter 3

Multivariate generalized linear mixed models

This chapter gives an introduction to multivariate generalized linear mixed models (GLMM) and their applications to small area estimation. The more simpler linear mixed models (LMM) and the linear models (LM) belong to the family of GLMM. To understand how the GLMM works in practice, it is better to start by describing the behavior of the LM and the LMM. Therefore, we first give some comments about the linear regression model. This model can be expressed as $\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{\epsilon}$, where $\boldsymbol{y}$ is a vector of observations, $\boldsymbol{X}$ is a matrix of known covariates, $\boldsymbol{\beta}$ is a vector of unknown regression coefficients and $\boldsymbol{\epsilon}$ is a vector of errors. In this model the regression coefficients are fixed unknown constants. However, there are some cases where it is better to assume that some of these coefficients are realizations of random variables. These cases typically are when the observations are correlated.

Let us see how a linear mixed model may be useful for modeling the correlations among observations. Consider, for example, the mortality in 53 counties of Galicia. Assume that the response variable y_{ij}, measured at individual i of county j, depends on a random effect u_j associated to the county j and whose value is unobservable. A linear mixed model may be expressed as $y_{ij} = x_{ij} + u_j + \epsilon_{ij}$, $i = 1, \ldots, m$, $j = 1, \ldots, 53$, where x_{ij} is a vector of known covariates, $\boldsymbol{\beta}$ is a vector of unknown regression coefficients, the random effects $u_1, \ldots, u_{53}$ are assumed to be i.i.d. with mean zero and variance σ^2 and the ϵ_{ij}'s are i.i.d. errors with mean zero and variance τ^2. This is a simple LMM. In a general form LMM may be expressed as

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{\epsilon},$$

where $\boldsymbol{y}$ is a vector of observations, $\boldsymbol{X}$ is a matrix of known covariates, $\boldsymbol{\beta}$ is a vector of unknown regression coefficients, which are often called fixed effects, $\boldsymbol{Z}$ is a known matrix, $\boldsymbol{u}$ is a vector of random effects and $\boldsymbol{\epsilon}$ is a vector of errors. Note that $\boldsymbol{u}$ is unobservable and $\boldsymbol{\beta}$ is unknown. This model assumes that the random effects and errors have mean zero and finite variances and that $\boldsymbol{u}$ and $\boldsymbol{\epsilon}$ are uncorrelated (Jiang et al., 2002). These models are used in small area estimation

Large-scale sample survey are usually designed to produce reliable estimates of various characteristics of interest for large geographic areas. However, for effective planning of health, social, and other services, there is a growing demand to produce similar estimates for smaller geographic areas for which adequate samples are not available. The usual design-based estimator are unreliable for these areas. This makes necessary to "borrow strength" from related areas for finding precise estimators without increasing the sample size. Such estimators are typically based on linear mixed models or generalized linear mixed models that provide a link to related small areas through the use of supplementary data. This data may be a recent census data or current administrative records.

The LMM have been widely used in situations where the observations are continuous. However, there are practical cases where the observations are discrete or categorical. For example, the number of forest fires in a country or the number of employed, unemployed and inactive people in a county. McCullath and Nelder (1999) proposed an extension of linear models, called generalized linear models (GLM). They pointed out that the key elements of a classical linear model are that the observation are independent, the mean of the observation is a linear function of some covariates and the variance of the observation is constant. The extension to GLM modifies the second and the third assumption. In GLM the mean of the observation is associated with a linear function of some covariates through a link function and the variance of the observation is a function of the mean. The GLM include as special cases, linear regression and analysis-of-variance models, logit and probit models for binary responses, log-linear models and multinomial response models for counts and some commonly used models for survival data. Therefore, these models are applicable to cases where the observations may not be continuous.

The GLM have in common with linear models that the observations area assumed to be independent. But, in many cases, the observations are correlated, as well as discrete or categorical. It is clear that it is necessary to extend the linear mixed models to cases where the responses are both correlated and, at the same time, discrete or categorical. Besides, many variables of interest in small area estimation are not normally distributed, and therefore cannot be adequately modeled via the linear mixed models. For such variables we can instead consider using a GLMM. Under this type of models, the distribution of the vector $\boldsymbol{y}$ of population values of the variable of interest is assumed to depend on a vector quantity $\boldsymbol{\eta}$ that is related to regression covariates and random components through the equation $\boldsymbol{\eta} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u}$. The linear predictor $\boldsymbol{\eta}$ is connected to $\boldsymbol{y}$ via a known function g, defined by $E(\boldsymbol{y}|\boldsymbol{u}) = g(\boldsymbol{\eta})$ (Saei and Chambers, 2003).

In the rest of the chapter we introduce the general theory of the multivariate generalized linear mixed models (MGLMM) that we use in this work. First, we present the model and then we describe the penalized quasilikelihood algorithm used to fit the model and the predictors. Finally, we introduce an explicit-formula procedure to estimate the mean squared error.

3.1 The model

Let $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m$ be independent vectors such that

$$\boldsymbol{u}_i \sim N_{\nu_i}(\boldsymbol{0}, \varphi_i \boldsymbol{\Sigma}_{ui}), \quad i = 1, \ldots, m,$$

where $\boldsymbol{\Sigma}_{u1}, \ldots, \boldsymbol{\Sigma}_{um}$ are known symmetric and positive-definite matrices. Let $\boldsymbol{u} = (\boldsymbol{u}'_1, \ldots, \boldsymbol{u}'_m)'$ and $\nu = \sum_{i=1}^{m} \nu_i$ be such that

$$\boldsymbol{u} \sim N_\nu(\mathbf{0}, \boldsymbol{V}_u), \qquad \text{with} \quad \boldsymbol{V}_u = \text{diag}(\varphi_1 \boldsymbol{\Sigma}_{u1}, \ldots, \varphi_m \boldsymbol{\Sigma}_{um}).$$

Let $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ be $q \times 1$ independent vectors whose densities, conditioned to $\boldsymbol{u}$, belong to the exponential family, i.e.

$$f(\boldsymbol{y}_j|\boldsymbol{u}) = c(\boldsymbol{y}_j) \exp\left\{\boldsymbol{\theta}'_j \boldsymbol{y}_j - b(\boldsymbol{\theta}_j)\right\}, \quad j = 1, \ldots, n,$$

where $\boldsymbol{\theta}_j \in \Theta$, $j = 1, \ldots, n$, are the unknown $q \times 1$ vectors of *natural parameters*. Let $\mu_j = \mu(\boldsymbol{\theta}_j)$ and $\Sigma_j = \Sigma(\boldsymbol{\theta}_j)$ be the mean vector and the covariance matrix of $\boldsymbol{y}_j$ conditioned to $\boldsymbol{u}$. If we use the matrix notation $\frac{\partial a}{\partial b} = \frac{\partial}{\partial b} \cdot a$, where "$\cdot$" denotes the product of the column vector $\frac{\partial}{\partial b} = (\frac{\partial}{\partial b_1}, \ldots, \frac{\partial}{\partial b_m})'_{m \times 1}$ by the row vector $a_{1 \times n}$, then

$$\boldsymbol{\mu}(\boldsymbol{\theta}_j) = E[\boldsymbol{y}_j|\boldsymbol{u}] = \frac{\partial b(\theta_j)}{\partial \theta_j}, \quad \boldsymbol{\Sigma}(\boldsymbol{\theta}_j) = \text{cov}[\boldsymbol{y}_j|\boldsymbol{u}] = \frac{\partial \boldsymbol{\mu}'_j}{\partial \boldsymbol{\theta}_j} = \frac{\partial^2 b(\boldsymbol{\theta}_j)}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}'_j}, \quad j = 1, \ldots, n.$$

Let $\boldsymbol{x}_j$ and $\boldsymbol{z}_j$ be matrices of sizes $q \times p$ and $q \times \nu$ respectively. The *linear predictors* are

$$\boldsymbol{\eta}_j = \boldsymbol{x}_j \boldsymbol{\beta} + \boldsymbol{z}_j \boldsymbol{u}, \quad j = 1, \ldots, n.$$

Let us consider an injective *link* function $g : M \mapsto R^q$, where $M \subset R^q$ is the subset of possible values of $\boldsymbol{\mu}(\boldsymbol{\theta}_j)$, and such that

$$g(\boldsymbol{\mu}(\boldsymbol{\theta}_j)) = \boldsymbol{\eta}_j, \quad j = 1, \ldots, n.$$

The dependency of the natural parameters from the random effects is explicitly given by the function $d = (g \circ \mu)^{-1}$, i.e.

$$\boldsymbol{\theta}_j = d(\boldsymbol{\eta}_j), \quad j = 1, \ldots, n.$$

The *natural link* is $g = \mu^{-1}$, so that

$$\boldsymbol{\theta}_j = \boldsymbol{x}_j \boldsymbol{\beta} + \boldsymbol{z}_j \boldsymbol{u}, \quad j = 1, \ldots, n.$$

In the next sections we present a penalized quasi-likelihood (PQL) method to estimate the model parameters and to predict the random effects of multivariate generalized linear mixed models (MGLMM) with natural link.

3.2 The Penalized quasi-likelihood algorithm

3.2.1 Estimation and prediction of $\boldsymbol{\beta}$ and $\boldsymbol{u}$

The joint probability density function of $\boldsymbol{y} = (\boldsymbol{y}'_1, \ldots, \boldsymbol{y}'_n)'$, conditioned to $\boldsymbol{u}$ is

$$f_1(\boldsymbol{y}|\boldsymbol{u}) = \left[\prod_{j=1}^{n} c(\boldsymbol{y}_j)\right] \exp\left\{\sum_{j=1}^{n} \left(\boldsymbol{\theta}'_j \boldsymbol{y}_j - b(\boldsymbol{\theta}_j)\right)\right\}$$

and the log-likelihood is

$$l_1(\boldsymbol{y}|\boldsymbol{u}) = c_1 + \sum_{j=1}^{n} \left(\boldsymbol{\theta}_j'\boldsymbol{y}_j - b(\boldsymbol{\theta}_j)\right).$$

The probability density function of $\boldsymbol{u}$ is

$$f_2(\boldsymbol{u}) = (2\pi)^{-\nu/2} |\boldsymbol{\Sigma}_u|^{-1/2} \exp\left\{\frac{-1}{2}\boldsymbol{u}'\boldsymbol{\Sigma}_u^{-1}\boldsymbol{u}\right\},$$

and the log-likelihood is

$$l_2(\boldsymbol{u}) = c_2 - \frac{1}{2}\left\{\log|\boldsymbol{\Sigma}_u| + \boldsymbol{u}'\boldsymbol{\Sigma}_u^{-1}\boldsymbol{u}\right\} = c_2 - \frac{1}{2}\left\{\sum_{i=1}^{m}\left(\nu_i \log\varphi_i + \log|\boldsymbol{\Sigma}_{ui}| + \varphi_i^{-1}\boldsymbol{u}_i'\boldsymbol{\Sigma}_{ui}^{-1}\boldsymbol{u}_i\right)\right\}.$$

The ML-PQL estimator of $\boldsymbol{\beta}$ and predictor of $\boldsymbol{u}$ (see Breslow and Clayton (1993)) maximizes the joint likelihood

$$l(\boldsymbol{y}, \boldsymbol{u}) = l_1(\boldsymbol{y}|\boldsymbol{u}) + l_2(\boldsymbol{u}).$$

In this section we derive a Newton-Raphson algorithm to calculate ML-PQL estimator of $\boldsymbol{\beta}$ and predictor of $\boldsymbol{u}$.

Proposition 3.2.1. Under the MGLM model with natural link defined in the Section 3.2.1, the score vector and the Fisher information matrix of $l(\boldsymbol{y}, \boldsymbol{u})$ are

$$\boldsymbol{S}(\theta) = \begin{bmatrix} \boldsymbol{S}_\beta(\boldsymbol{\theta}) \\ \boldsymbol{S}_u(\boldsymbol{\theta}) \end{bmatrix}, \quad \boldsymbol{F}(\boldsymbol{\theta}_s) = \begin{bmatrix} \boldsymbol{F}_{\beta\beta}(\boldsymbol{\theta}) & \boldsymbol{F}_{\beta u}(\theta) \\ \boldsymbol{F}_{u\beta}(\boldsymbol{\theta}) & \boldsymbol{F}_{uu}(\theta) \end{bmatrix},$$

where

$$\begin{array}{ll} \boldsymbol{S}_\beta(\boldsymbol{\theta}) = \sum_{j=1}^n \boldsymbol{x}_j'[\boldsymbol{y}_j - \boldsymbol{\mu}_j], & \boldsymbol{S}_u(\boldsymbol{\theta}) = \sum_{j=1}^n \boldsymbol{z}_j'[\boldsymbol{y}_j - \boldsymbol{\mu}_j] - \boldsymbol{\Sigma}_u^{-1}\boldsymbol{u}, \\ \boldsymbol{F}_{\beta\beta}(\boldsymbol{\theta}) = \sum_{j=1}^n \boldsymbol{x}_j'\boldsymbol{\Sigma}_j\boldsymbol{x}_j, & \boldsymbol{F}_{\beta u}(\boldsymbol{\theta}) = \sum_{j=1}^n \boldsymbol{x}_j'\boldsymbol{\Sigma}_j\boldsymbol{z}_j, \\ \boldsymbol{F}_{u\beta}(\boldsymbol{\theta}) = \sum_{j=1}^n \boldsymbol{z}_j'\boldsymbol{\Sigma}_j\boldsymbol{x}_j, & \boldsymbol{F}_{uu}(\boldsymbol{\theta}) = \sum_{j=1}^n \boldsymbol{z}_j'\boldsymbol{\Sigma}_j\boldsymbol{z}_j - \boldsymbol{\Sigma}_u^{-1}. \end{array}$$

Proof. Let $l_j = \boldsymbol{\theta}_j'\boldsymbol{y}_j - b(\boldsymbol{\theta}_j)$ and remind that $\boldsymbol{\theta}_j = \boldsymbol{x}_j\boldsymbol{\beta} + \boldsymbol{z}_j\boldsymbol{u}$, $j = 1, \ldots, n$. Then

$$\begin{aligned} \boldsymbol{S}_\beta &= \frac{\partial l(\boldsymbol{y},\boldsymbol{u})}{\partial\boldsymbol{\beta}} = \frac{\partial l_1(\boldsymbol{y}|\boldsymbol{u})}{\partial\boldsymbol{\beta}} = \sum_{j=1}^n \frac{\partial\boldsymbol{\theta}_j'}{\partial\boldsymbol{\beta}}\frac{\partial l_j}{\partial\boldsymbol{\theta}_j} = \sum_{j=1}^n \boldsymbol{x}_j'\left(\boldsymbol{y}_j - \frac{\partial b(\boldsymbol{\theta}_j)}{\partial\boldsymbol{\theta}_j}\right) = \sum_{j=1}^n \boldsymbol{x}_j'\left(\boldsymbol{y}_j - \boldsymbol{\mu}_j\right), \\ \boldsymbol{S}_u &= \frac{\partial l(\boldsymbol{y},\boldsymbol{u})}{\partial\boldsymbol{u}} = \frac{\partial l_1(\boldsymbol{y}|\boldsymbol{u})}{\partial\boldsymbol{u}} + \frac{\partial l_2(\boldsymbol{u})}{\partial\boldsymbol{u}} = \sum_{j=1}^n \frac{\partial\boldsymbol{\theta}_j'}{\partial\boldsymbol{u}}\frac{\partial l_j}{\partial\boldsymbol{\theta}_j} - \boldsymbol{\Sigma}_u^{-1}\boldsymbol{u} = \sum_{j=1}^n \boldsymbol{z}_j'\left(\boldsymbol{y}_j - \boldsymbol{\mu}_j\right) - \boldsymbol{\Sigma}_u^{-1}\boldsymbol{u}, \end{aligned}$$

The second order partial derivatives are

$$\begin{aligned} \boldsymbol{H}_{\beta\beta} &= \frac{\partial\boldsymbol{S}_\beta'}{\partial\boldsymbol{\beta}} = -\sum_{j=1}^n \frac{\partial\boldsymbol{\mu}_j'}{\partial\boldsymbol{\beta}}\boldsymbol{x}_j = -\sum_{j=1}^n \frac{\partial\boldsymbol{\theta}_j'}{\partial\boldsymbol{\beta}}\frac{\partial\boldsymbol{\mu}_j'}{\partial\boldsymbol{\theta}_j}\boldsymbol{x}_j = -\sum_{j=1}^n \boldsymbol{x}_j'\Sigma_j\boldsymbol{x}_j, \\ \boldsymbol{H}_{u\beta} &= \frac{\partial\boldsymbol{S}_\beta'}{\partial\boldsymbol{u}} = -\sum_{j=1}^n \frac{\partial\boldsymbol{\mu}_j'}{\partial\boldsymbol{u}}\boldsymbol{x}_j = -\sum_{j=1}^n \frac{\partial\boldsymbol{\theta}_j'}{\partial\boldsymbol{u}}\frac{\partial\boldsymbol{\mu}_j'}{\partial\boldsymbol{\theta}_j}\boldsymbol{x}_j = -\sum_{j=1}^n \boldsymbol{z}_j'\Sigma_j\boldsymbol{x}_j, \\ \boldsymbol{H}_{uu} &= \frac{\partial\boldsymbol{S}_u'}{\partial\boldsymbol{u}} = -\sum_{j=1}^n \frac{\partial\boldsymbol{\mu}_j'}{\partial\boldsymbol{u}}\boldsymbol{z}_j - \boldsymbol{\Sigma}_u^{-1} = -\sum_{j=1}^n \frac{\partial\boldsymbol{\theta}_j'}{\partial\boldsymbol{u}}\frac{\partial\boldsymbol{\mu}_j'}{\partial\boldsymbol{\theta}_j}\boldsymbol{z}_j - \boldsymbol{\Sigma}_u^{-1} = -\sum_{j=1}^n \boldsymbol{z}_j'\Sigma_j\boldsymbol{z}_j - \boldsymbol{\Sigma}_u^{-1}. \end{aligned}$$

Finally $\boldsymbol{F}_{\beta\beta} = -\boldsymbol{H}_{\beta\beta}$, $\boldsymbol{F}_{u\beta} = \boldsymbol{F}'_{\beta u} = -\boldsymbol{H}_{u\beta}$ and $\boldsymbol{F}_{uu} = -\boldsymbol{H}_{uu}$.

Algorithm A. Let $\varphi_1, \ldots, \varphi_m$ be known. The ML-PQL estimator and predictor of $\boldsymbol{\beta}$ and $\boldsymbol{u}$ can be calculated by applying the following Newton-Raphson algorithm:

(A.1) *Set initial values:* Do $r = 0$, $\boldsymbol{\beta}^{(0)} = \boldsymbol{\beta}^{inicial}$ and $\boldsymbol{u}^{(0)} = \boldsymbol{u}^{inicial}$.

(A.2) *Iteration $r+1$:* Calculate

$$\boldsymbol{\theta}_j^{(r)} = \boldsymbol{x}_j\boldsymbol{\beta}^{(r)} + \boldsymbol{z}_j\boldsymbol{u}^{(r)}, \quad \boldsymbol{\mu}_j^{(r)} = \boldsymbol{\mu}(\boldsymbol{\theta}_j^{(r)}), \quad \boldsymbol{\Sigma}_j^{(r)} = \Sigma(\boldsymbol{\theta}_j^{(r)}), \quad j = 1, \ldots, n.$$

Update $\boldsymbol{\beta}^{(r)}$ and $\boldsymbol{u}^{(r)}$ by means of

$$\begin{bmatrix} \boldsymbol{\beta}^{(r+1)} \\ \boldsymbol{u}^{(r+1)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\beta}^{(r)} \\ \boldsymbol{u}^{(r)} \end{bmatrix} + \left(\boldsymbol{F}^{(r)}\right)^{-1} \begin{bmatrix} \sum_{j=1}^n \boldsymbol{x}'_j(\boldsymbol{y}_j - \boldsymbol{\mu}_j^{(r)}) \\ \sum_{j=1}^n \boldsymbol{z}'_j(\boldsymbol{y}_j - \boldsymbol{\mu}_j^{(r)}) - \boldsymbol{\Sigma}_u^{-1}\boldsymbol{u}^{(r)} \end{bmatrix}$$

with

$$\boldsymbol{F}^{(r)} = \begin{bmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^n \boldsymbol{x}'_j\boldsymbol{\Sigma}_j^{(r)}\boldsymbol{x}_j & \sum_{j=1}^n \boldsymbol{x}'_j\boldsymbol{\Sigma}_j^{(r)}\boldsymbol{z}_j \\ \sum_{j=1}^n \boldsymbol{z}'_j\boldsymbol{\Sigma}_j^{(r)}\boldsymbol{x}_j & \sum_{j=1}^n \boldsymbol{z}'_j\boldsymbol{\Sigma}_j^{(r)}\boldsymbol{z}_j + \boldsymbol{\Sigma}_u^{-1} \end{bmatrix}$$

and

$$\left(\boldsymbol{F}^{(r)}\right)^{-1} = \begin{bmatrix} F^{11} & F^{12} \\ F^{21} & F^{22} \end{bmatrix} = \begin{bmatrix} (F_{11} - F_{12}F_{22}^{-1}F_{21})^{-1} & -F^{11}F_{12}F_{22}^{-1} \\ -F_{22}^{-1}F_{21}F^{11} & F_{22}^{-1} + F_{22}^{-1}F_{21}F^{11}F_{12}F_{22}^{-1} \end{bmatrix}.$$

(A.3) *End:* Repeat step (A.2) until convergence of $\boldsymbol{\beta}^{(r)}$ and $\boldsymbol{u}_i^{(r)}$, $i = 1, \ldots, m$.

3.2.2 PQL estimation of the variance components

Let us write $l_1(\boldsymbol{\beta}, \boldsymbol{u}) = l_1(\boldsymbol{y}|\boldsymbol{u})$ as a function of $\boldsymbol{\beta}$ and $\boldsymbol{u}$. We expand $l_1(\boldsymbol{\beta}, \boldsymbol{u})$ in Taylor series around the values $\boldsymbol{\beta}^\circ$ and $\boldsymbol{u}^\circ$ that maximizes $l_1(\boldsymbol{\beta}, \boldsymbol{u})$. We obtain

$$\begin{aligned} l_1(\boldsymbol{\beta}, \boldsymbol{u}) &\approx l_1(\boldsymbol{\beta}^\circ, \boldsymbol{u}^\circ) + \left(\frac{\partial l_1(\boldsymbol{\beta}^\circ, \boldsymbol{u}^\circ)}{\partial \boldsymbol{\beta}}, \frac{\partial l_1(\boldsymbol{\beta}^\circ, \boldsymbol{u}^\circ)}{\partial \boldsymbol{u}}\right) \begin{pmatrix} \boldsymbol{\beta} - \boldsymbol{\beta}^\circ \\ \boldsymbol{u} - \boldsymbol{u}^\circ \end{pmatrix} \\ &+ \frac{1}{2}(\boldsymbol{\beta}' - \boldsymbol{\beta}^{\circ\prime}, \boldsymbol{u}' - \boldsymbol{u}^{\circ\prime}) \begin{pmatrix} \frac{\partial^2 l_1(\boldsymbol{\beta}^\circ, \boldsymbol{u}^\circ)}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'} & \frac{\partial^2 l_1(\boldsymbol{\beta}^\circ, \boldsymbol{u}^\circ)}{\partial \boldsymbol{\beta}\partial \boldsymbol{u}'} \\ \frac{\partial^2 l_1(\boldsymbol{\beta}^\circ, \boldsymbol{u}^\circ)}{\partial \boldsymbol{u}\partial \boldsymbol{\beta}'} & \frac{\partial^2 l_1(\boldsymbol{\beta}^\circ, \boldsymbol{u}^\circ)}{\partial \boldsymbol{u}\partial \boldsymbol{u}'} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta} - \boldsymbol{\beta}^\circ \\ \boldsymbol{u} - \boldsymbol{u}^\circ \end{pmatrix} \end{aligned} \quad (3.2.1)$$

Let $\boldsymbol{\eta} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u}$ and $\boldsymbol{\eta}^\circ = \boldsymbol{X}\boldsymbol{\beta}^\circ + \boldsymbol{Z}\boldsymbol{u}^\circ$, then

$$\frac{\partial l_1}{\partial \boldsymbol{\beta}} = \frac{\partial \boldsymbol{\eta}'}{\partial \boldsymbol{\beta}}\frac{\partial l_1}{\partial \boldsymbol{\eta}} = \boldsymbol{X}'\frac{\partial l_1}{\partial \boldsymbol{\eta}}, \ \frac{\partial l_1}{\partial \boldsymbol{u}} = \frac{\partial \boldsymbol{\eta}'}{\partial \boldsymbol{u}}\frac{\partial l_1}{\partial \boldsymbol{\eta}} = \boldsymbol{Z}'\frac{\partial l_1}{\partial \boldsymbol{\eta}}, \ \frac{\partial^2 l_1}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'} = \boldsymbol{X}'\frac{\partial^2 l_1}{\partial \boldsymbol{\eta}\partial \boldsymbol{\eta}'}\boldsymbol{X},$$
$$\frac{\partial^2 l_1}{\partial \boldsymbol{\beta}\partial \boldsymbol{u}'} = \boldsymbol{X}'\frac{\partial^2 l_1}{\partial \boldsymbol{\eta}\partial \boldsymbol{\eta}'}\boldsymbol{Z}, \ \frac{\partial^2 l_1}{\partial \boldsymbol{u}\partial \boldsymbol{u}'} = \boldsymbol{Z}'\frac{\partial^2 l_1}{\partial \boldsymbol{\eta}\partial \boldsymbol{\eta}'}\boldsymbol{Z}.$$

As $\boldsymbol{\beta}^\circ$ and $\boldsymbol{u}^\circ$ maximize $l_1(\boldsymbol{\beta}, \boldsymbol{u})$, the linear term of (3.2.1) is null. Therefore

$$\begin{aligned} l_1(\boldsymbol{y}|\boldsymbol{u}) &\approx c + \frac{1}{2}(\boldsymbol{\beta}' - \boldsymbol{\beta}^{\circ\prime}, \boldsymbol{u}' - \boldsymbol{u}^{\circ\prime}) \begin{pmatrix} \boldsymbol{X}' \\ \boldsymbol{Z}' \end{pmatrix} \left(\frac{\partial^2 l_1}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}'} \right) (\boldsymbol{X}, \boldsymbol{Z}) \begin{pmatrix} \boldsymbol{\beta} - \boldsymbol{\beta}^\circ \\ \boldsymbol{u} - \boldsymbol{u}^\circ \end{pmatrix} \\ &\approx c - \frac{1}{2}(\boldsymbol{\eta}^\circ - \boldsymbol{\eta})' \boldsymbol{W} (\boldsymbol{\eta}^\circ - \boldsymbol{\eta}) \doteq \ell_1(\boldsymbol{\eta}^\circ|\boldsymbol{u}), \end{aligned}$$

where

$$\boldsymbol{W} = -E\left[\frac{\partial^2 l_1(\boldsymbol{y}|\boldsymbol{u})}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}'}\right]\Bigg|_{\eta=\eta^\circ}. \tag{3.2.2}$$

For the natural link we have $\boldsymbol{\eta}_j = \boldsymbol{\theta}_j$ and

$$\boldsymbol{W} = -E\left[\frac{\partial^2 l_1(\boldsymbol{y}|\boldsymbol{u})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}\right]\Bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^\circ} = \operatorname{cov}(\boldsymbol{y}|\boldsymbol{u})|_{\boldsymbol{\theta}=\boldsymbol{\theta}^\circ} = \operatorname*{diag}_{1 \le j \le n}(\boldsymbol{\Sigma}(\boldsymbol{\theta}_j^\circ)).$$

Remark. Under the natural link it holds that $\boldsymbol{\eta}_j = \boldsymbol{\theta}_j$ and the partial derivatives of $l_1(\boldsymbol{y}|\boldsymbol{u})$ with respect to $\boldsymbol{\theta}$ are

$$\begin{aligned} \frac{\partial l_1(\boldsymbol{y}|\boldsymbol{u})}{\partial \boldsymbol{\theta}_j} &= \frac{\partial(\boldsymbol{\theta}_j' \boldsymbol{y}_j - b(\boldsymbol{\theta}_j))}{\partial \boldsymbol{\theta}_j} = \boldsymbol{y}_j - \frac{\partial b(\boldsymbol{\theta}_j)}{\partial \boldsymbol{\theta}_j}, \quad j = 1, \dots, n, \\ \frac{\partial^2 l_1(\boldsymbol{y}|\boldsymbol{u})}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}_j'} &= -\frac{\partial^2 b(\boldsymbol{\theta}_j)}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}_j'}, \quad j = 1, \dots, n, \\ \frac{\partial^2 l_1(\boldsymbol{y}|\boldsymbol{u})}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}_k'} &= 0 \quad \forall j \ne k, \quad j, k = 1, \dots, n. \end{aligned}$$

Therefore

$$\frac{\partial^2 l_1(\boldsymbol{y}|\boldsymbol{u})}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}'} = \frac{\partial^2 l_1(\boldsymbol{y}|\boldsymbol{u})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} = -\operatorname{diag}(\boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_n) = -\boldsymbol{\Sigma} \quad \text{and} \quad \boldsymbol{W} = E[\boldsymbol{\Sigma}] = \boldsymbol{\Sigma},$$

because the expectation is taken with respect to $f(\boldsymbol{y}|\boldsymbol{u})$ and $\boldsymbol{u}$ is fixed.

In summary, we have got that

$$l_1(\boldsymbol{y}|\boldsymbol{u}) \approx \ell_1(\boldsymbol{\eta}^\circ|\boldsymbol{u}) = -\frac{1}{2}(\boldsymbol{\eta}^\circ - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{Z}\boldsymbol{u})' \boldsymbol{W} (\boldsymbol{\eta}^\circ - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{Z}\boldsymbol{u}).$$

This is to say, $l_1(\boldsymbol{y}|\boldsymbol{u})$ can be approximated to the likelihood function of multivariate linear mixed model. The marginal probability distribution function should also be approximately the same, i.e. $l_1(\boldsymbol{y}) \approx \ell_1(\boldsymbol{\eta}^\circ)$. Therefore, the values of $\varphi_1, \dots, \varphi_m$ maximizing $l_1(\boldsymbol{y})$ should be approximately the same to those maximizing $\ell_1(\boldsymbol{\eta}^\circ)$. On the other hand $\ell_1(\boldsymbol{\eta}^\circ)$ can be obtained from $\ell_1(\boldsymbol{\eta}^\circ|\boldsymbol{u})$ and $l_2(\boldsymbol{u})$, which are both normal log-likelihoods. Therefore, we have

- $\ell_1(\boldsymbol{\eta}^\circ)$ is the loglikelihood of $\boldsymbol{\eta}^\circ \sim N(\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}_y)$, with $\boldsymbol{\Sigma}_y = \boldsymbol{Z}'\boldsymbol{\Sigma}_u\boldsymbol{Z} + \boldsymbol{W}^{-1}$,
- For $\boldsymbol{\eta}^\circ$ we assume the model $\boldsymbol{\eta}^\circ = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u} + \boldsymbol{e}$, where $\boldsymbol{e} \sim N(\boldsymbol{0}, \boldsymbol{W}^{-1})$ and $\boldsymbol{u} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_u)$ are independent.

- The values $\varphi_1, \ldots, \varphi_m$ maximizing $\ell_1(\boldsymbol{\eta}^\circ)$ and $\ell_1(\boldsymbol{\eta}^\circ|\boldsymbol{u}) + l_2(\boldsymbol{u})$ should be approximately the same.

McGilchrist (1994) proposed to maximize $\ell_1(\boldsymbol{\eta}^\circ)$ for obtaining an updating equation for $\varphi_1, \ldots, \varphi_m$. The PQL algorithm of McGilchrist for $\varphi_1, \ldots, \varphi_m$ can be implemented by applying the ML or the REML method to $\ell_1(\boldsymbol{\eta}^\circ, \boldsymbol{u}) = \ell_1(\boldsymbol{\eta}^\circ|\boldsymbol{u}) + l_2(\boldsymbol{u})$.

The ML-PQL approach

The penalized maximum likelihood estimation of $\varphi_1, \ldots, \varphi_m$ maximizes $\ell_1(\boldsymbol{\eta}^\circ)$ after approximating $l_1(\boldsymbol{y}|\boldsymbol{u}) + l_2(\boldsymbol{u})$ by $\ell_1(\boldsymbol{\eta}^\circ|\boldsymbol{u}) + l_2(\boldsymbol{u})$. We assume the approximate model

$$\boldsymbol{\eta}^\circ = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{e},$$

where $\boldsymbol{e} \sim N(\boldsymbol{0}, \boldsymbol{W}^{-1})$ and $\boldsymbol{u} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_u)$ are independent. Therefore $\boldsymbol{\eta}^\circ \sim N(\boldsymbol{X\beta}, \boldsymbol{\Sigma}_y)$, with $\boldsymbol{\Sigma}_y = \boldsymbol{Z\Sigma}_u\boldsymbol{Z}' + \boldsymbol{W}^{-1}$ and $\boldsymbol{\Sigma}_u = \mathrm{diag}(\varphi_1\boldsymbol{\Sigma}_{u_1}, \ldots, \varphi_m\boldsymbol{\Sigma}_{u_m})$. The marginal density and log-density of $\boldsymbol{\eta}^\circ$ are

$$\begin{aligned}
f_1^\circ(\boldsymbol{\eta}^\circ) &= (2\pi)^{-nq/2}|\boldsymbol{\Sigma}_y|^{-1/2}\exp\left\{-\frac{1}{2}(\boldsymbol{\eta}^\circ - \boldsymbol{X\beta})'\boldsymbol{\Sigma}_y^{-1}(\boldsymbol{\eta}^\circ - \boldsymbol{X\beta})\right\}, \\
\ell_{ML}(\boldsymbol{\eta}^\circ) &= -\frac{nq}{2}\log 2\pi - \frac{1}{2}\log|\boldsymbol{\Sigma}_y| - \frac{1}{2}(\boldsymbol{\eta}^\circ - \boldsymbol{X\beta})'\boldsymbol{\Sigma}_y^{-1}(\boldsymbol{\eta}^\circ - \boldsymbol{X\beta}).
\end{aligned}$$

Let $\boldsymbol{G}_i = \boldsymbol{Z}_i\boldsymbol{\Sigma}_{ui}\boldsymbol{Z}_i'$. By taking derivatives with respect to φ_i, $i = 1, \ldots, m$, we get

$$\begin{aligned}
\frac{\partial\boldsymbol{\Sigma}_y}{\partial\varphi_i} &= \boldsymbol{Z}_i\boldsymbol{\Sigma}_{ui}\boldsymbol{Z}_i' = \boldsymbol{G}_i, \\
\boldsymbol{S}_{\varphi_i} &= \frac{\partial\ell_{ML}(\boldsymbol{\eta}^\circ)}{\partial\varphi_i} = -\frac{1}{2}\,\mathrm{tr}\left(\boldsymbol{\Sigma}_y^{-1}\frac{\partial\boldsymbol{\Sigma}_y}{\partial\varphi_i}\right) + \frac{1}{2}(\boldsymbol{\eta}^\circ - \boldsymbol{X\beta})'\boldsymbol{\Sigma}_y^{-1}\frac{\partial\boldsymbol{\Sigma}_y}{\partial\varphi_i}\boldsymbol{\Sigma}_y^{-1}(\boldsymbol{\eta}^\circ - \boldsymbol{X\beta}) \\
&= -\frac{1}{2}\,\mathrm{tr}(\boldsymbol{\Sigma}_y^{-1}\boldsymbol{G}_i) + \frac{1}{2}(\boldsymbol{\eta}^\circ - \boldsymbol{X\beta})'\boldsymbol{\Sigma}_y^{-1}\boldsymbol{G}_i\boldsymbol{\Sigma}_y^{-1}(\boldsymbol{\eta}^\circ - \boldsymbol{X\beta}), \\
\frac{\partial^2\ell_{ML}(\boldsymbol{\eta}^\circ)}{\partial\varphi_i\partial\varphi_j} &= \frac{1}{2}\,\mathrm{tr}(\boldsymbol{\Sigma}_y^{-1}\boldsymbol{G}_j\boldsymbol{\Sigma}_y^{-1}\boldsymbol{G}_i) - (\boldsymbol{\eta}^\circ - \boldsymbol{X\beta})'\boldsymbol{\Sigma}_y^{-1}\boldsymbol{G}_i\boldsymbol{\Sigma}_y^{-1}\boldsymbol{G}_j\boldsymbol{\Sigma}_y^{-1}(\boldsymbol{\eta}^\circ - \boldsymbol{X\beta}).
\end{aligned}$$

We recall that

$$E[\boldsymbol{y}'\boldsymbol{Q}\boldsymbol{y}] = \mathrm{tr}\left\{\boldsymbol{Q}\,v(\boldsymbol{y})\right\} + E[\boldsymbol{y}]'\boldsymbol{Q}E[\boldsymbol{y}].$$

The Fisher amount of information associated to φ_i is

$$\begin{aligned}
\boldsymbol{F}_{\varphi_i\varphi_j} &= E\left[-\frac{\partial^2\ell_{ML}(\boldsymbol{\eta}^\circ)}{\partial\varphi_i\partial\varphi_j}\right] = -\frac{1}{2}\,\mathrm{tr}(\boldsymbol{\Sigma}_y^{-1}\boldsymbol{G}_j\boldsymbol{\Sigma}_y^{-1}\boldsymbol{G}_i) + \mathrm{tr}(\boldsymbol{\Sigma}_y^{-1}\boldsymbol{G}_i\boldsymbol{\Sigma}_y^{-1}\boldsymbol{G}_j\boldsymbol{\Sigma}_y^{-1}\boldsymbol{\Sigma}_y) \\
&= \frac{1}{2}\,\mathrm{tr}(\boldsymbol{\Sigma}_y^{-1}\boldsymbol{G}_i\boldsymbol{\Sigma}_y^{-1}\boldsymbol{G}_j).
\end{aligned}$$

For $\boldsymbol{\varphi} = (\varphi_1, \ldots, \varphi_m)$ the updating equation of the Fisher-scoring algorithm is

$$\boldsymbol{\varphi}^{k+1} = \boldsymbol{\varphi}^k + \boldsymbol{F}^{-1}(\boldsymbol{\varphi}^k)\boldsymbol{S}(\boldsymbol{\varphi}^k).$$

The REML-PQL approach

The REML log-likelihood is

$$l_{reml}(\boldsymbol{\eta}^{\circ}) = -\frac{1}{2}(qn-p)\log 2\pi - \frac{1}{2}\log|\boldsymbol{K}'\boldsymbol{\Sigma}_y\boldsymbol{K}| - \frac{1}{2}\boldsymbol{\eta}^{\circ\prime}\boldsymbol{P}\boldsymbol{\eta}^{\circ},$$

where

$$\boldsymbol{P} = \boldsymbol{\Sigma}_y^{-1} - \boldsymbol{\Sigma}_y^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Sigma}_y^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}_y^{-1}, \quad \boldsymbol{K} = \boldsymbol{W} - \boldsymbol{W}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{W}.$$

The first order partial derivatives of $l_{reml}(\boldsymbol{\eta}^{\circ})$ with respect to φ_i are

$$S_{\varphi_i} = -\frac{1}{2}\mathrm{tr}(\boldsymbol{P}\boldsymbol{G}_i) + \frac{1}{2}\boldsymbol{\eta}^{\circ\prime}\boldsymbol{P}\boldsymbol{G}_i\boldsymbol{P}\boldsymbol{\eta}^{\circ}, \quad i = 1, \dots, m.$$

The elements of the Fisher information matrix are

$$F_{\varphi_i\varphi_j} = \frac{1}{2}\,\mathrm{tr}(\boldsymbol{P}\boldsymbol{G}_j\boldsymbol{P}\boldsymbol{G}_i), \quad i, j = 1, \dots, m.$$

For $\boldsymbol{\varphi} = (\varphi_1, \dots, \varphi_m)$ the updating equation of the Fisher-scoring algorithm is

$$\boldsymbol{\varphi}^{k+1} = \boldsymbol{\varphi}^{k} + \boldsymbol{F}^{-1}(\boldsymbol{\varphi}^{k})\boldsymbol{S}(\boldsymbol{\varphi}^{k}).$$

3.2.3 The PQL algorithm

The PQL algorithm is

Algorithm B. This algorithm calculates the PQL predictors of $\boldsymbol{u}$ and estimators of $\boldsymbol{\beta}$ and φ_i, $i = 1, \dots, m$.

(B.1) Do $\ell = 1$, where ℓ counts the external iterations. Set the values $\boldsymbol{\beta}^{(0)}$, $\boldsymbol{u}^{(0)}$ and $\varphi_i^{(1)}$, $i = 1, \dots, m$.

(B.2) Run *Algorithm A*. Use $\varphi_1^{(\ell)}, \dots, \varphi_m^{(\ell)}$ as know values and $\boldsymbol{\beta}^{(\ell-1)}$, $\boldsymbol{u}^{(\ell-1)}$ as algorithm seeds. Let $\boldsymbol{\beta}^{(\ell)}$ and $\boldsymbol{u}^{(\ell)}$ be the output of Algorithm A.

(B.3) Update φ_i by using the ML or the REML Fisher-scoring updating equation

$$\boldsymbol{\varphi}^{k+1} = \boldsymbol{\varphi}^{k} + \boldsymbol{F}^{-1}(\boldsymbol{\varphi}^{k})\boldsymbol{S}(\boldsymbol{\varphi}^{k}).$$

(B.4) Repeat the steps (B.2)-(B.3) until the convergence of $\boldsymbol{\beta}^{(\ell)}$, $\boldsymbol{u}_i^{(\ell)}$ and $\varphi_i^{(\ell)}$, $i = 1, \dots, m$.

3.3 Prediction and mean squared error

We are interested in predicting

$$\delta = A\boldsymbol{y}.$$

If we fit the model $\boldsymbol{\mu} = E[\boldsymbol{y}] = \boldsymbol{h}(\boldsymbol{\eta})$, where $\boldsymbol{\eta} = \boldsymbol{X\beta} + \boldsymbol{Zu}$, we get the prediction of $\boldsymbol{y}$, $\boldsymbol{h}(\hat{\boldsymbol{\eta}})$, where $\hat{\boldsymbol{\eta}} = \boldsymbol{X}\hat{\boldsymbol{\beta}} + \boldsymbol{Z}\hat{\boldsymbol{u}}$. The predictor of δ is

$$\hat{\delta} = A\boldsymbol{h}(\hat{\boldsymbol{\eta}}),$$

and the difference is $\hat{\delta} - \delta = A[\boldsymbol{h}(\hat{\boldsymbol{\eta}}) - \boldsymbol{y}]$. Therefore, the MSE of the predictor $\hat{\delta}$ is

$$MSE(\hat{\delta}) = E[(\hat{\delta} - \delta)(\hat{\delta} - \delta)'] = E\left[A\left\{\boldsymbol{h}(\hat{\boldsymbol{\eta}}) - \boldsymbol{y}\right\}\left\{\boldsymbol{h}(\hat{\boldsymbol{\eta}}) - \boldsymbol{y}\right\}' A'\right].$$

Let $\boldsymbol{V}_G = E[(\boldsymbol{y} - \boldsymbol{h}(\boldsymbol{\eta}))(\boldsymbol{y} - \boldsymbol{h}(\boldsymbol{\eta}))']$. By summing and substracting $\boldsymbol{h}(\boldsymbol{\eta})$, we get

$$\begin{aligned} MSE(\hat{\delta}) &= E\left[A\left\{\boldsymbol{h}(\hat{\boldsymbol{\eta}}) - \boldsymbol{h}(\boldsymbol{\eta})\right\}\left\{\boldsymbol{h}(\hat{\boldsymbol{\eta}}) - \boldsymbol{h}(\boldsymbol{\eta})\right\}' A'\right] + A\boldsymbol{V}_G A' \\ &+ AE\left[\left\{\boldsymbol{h}(\hat{\boldsymbol{\eta}}) - \boldsymbol{h}(\boldsymbol{\eta})\right\}\left\{\boldsymbol{h}(\boldsymbol{\eta}) - \boldsymbol{y}\right\}'\right] A' \\ &+ AE\left[\left\{\boldsymbol{h}(\boldsymbol{\eta}) - \boldsymbol{y}\right\}\left\{\boldsymbol{h}(\hat{\boldsymbol{\eta}}) - \boldsymbol{h}(\boldsymbol{\eta})\right\}'\right] A'. \end{aligned}$$

From the results in Kackar y Harville (1984), we have that the two last terms are null. It is possible to obtain an approximation to the first term on the right hand side of the above equality. Let us take an arbitrary element of $\boldsymbol{h}(\boldsymbol{\eta})$, for example $\boldsymbol{h}_j(\boldsymbol{\eta}_j)$. We recall that $\boldsymbol{h}_j(\boldsymbol{\eta}_j)$ is a vector of size q, i.e. $\boldsymbol{h}_j(\boldsymbol{\eta}_j) = (h_{j1}(\boldsymbol{\eta}_j), \ldots, h_{jq}(\boldsymbol{\eta}_j))'$. Let ζ_j be an admissible value of $\boldsymbol{\eta}_j$. We make a first-order Taylor series expansion of $h_{jk}(\zeta_j)$ around $\boldsymbol{\eta}_j = (\eta_{j1}, \ldots, \eta_{jq})'$, and we do the substitution $\zeta_j = \hat{\boldsymbol{\eta}}_j$. This to say,

$$h_{jk}(\hat{\boldsymbol{\eta}}_j) - h_{jk}(\boldsymbol{\eta}_j) \cong \sum_{\ell=1}^{q} \frac{\partial h_{jk}(\boldsymbol{\eta}_j)}{\partial \eta_{j\ell}}(\hat{\eta}_{j\ell} - \eta_{j\ell}).$$

If we do the same with $h_{j'k'}(\boldsymbol{\eta}_{j'})$ and we multiply both expressions, we get

$$[h_{jk}(\hat{\boldsymbol{\eta}}_j) - h_{jk}(\boldsymbol{\eta}_j)][h_{j'k'}(\hat{\boldsymbol{\eta}}_{j'}) - h_{j'k'}(\boldsymbol{\eta}_{j'})] \cong \boldsymbol{H}'_{jk}(\hat{\boldsymbol{\eta}}_j - \boldsymbol{\eta}_j)(\hat{\boldsymbol{\eta}}_{j'} - \boldsymbol{\eta}_{j'})'\boldsymbol{H}'_{j'k'},$$

where

$$\boldsymbol{H}'_{jk} = \left[\frac{\partial h_{jk}(\boldsymbol{\eta}_j)}{\partial \eta_{j1}}, \ldots, \frac{\partial h_{jk}(\boldsymbol{\eta}_j)}{\partial \eta_{jq}}\right].$$

By doing the same calculations on the pairs j, k and by defining the matrices

$$\boldsymbol{H}_j = \begin{bmatrix} \dfrac{\partial h_{j1}(\boldsymbol{\eta}_j)}{\partial \eta_{j1}} & \cdots & \dfrac{\partial h_{j1}(\boldsymbol{\eta}_j)}{\partial \eta_{jq}} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial h_{jq}(\boldsymbol{\eta}_j)}{\partial \eta_{j1}} & \cdots & \dfrac{\partial h_{jq}(\boldsymbol{\eta}_j)}{\partial \eta_{jq}} \end{bmatrix}, \ j = 1, \ldots, N, \quad \text{and} \quad \boldsymbol{H} = \operatorname{diag}\{\boldsymbol{H}_1, \ldots, \boldsymbol{H}_N\},$$

we obtain

$$[\boldsymbol{h}(\hat{\boldsymbol{\eta}}) - \boldsymbol{h}(\boldsymbol{\eta})][\boldsymbol{h}(\hat{\boldsymbol{\eta}}) - \boldsymbol{h}(\boldsymbol{\eta})]' \cong H(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta})(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta})'H'.$$

Therefore, we can write

$$\begin{aligned} E\left[A\left\{\boldsymbol{h}(\hat{\boldsymbol{\eta}}) - \boldsymbol{h}(\boldsymbol{\eta})\right\}\left\{\boldsymbol{h}(\hat{\boldsymbol{\eta}}) - \boldsymbol{h}(\boldsymbol{\eta})\right\}' A'\right] &\cong E\left[AH(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta})(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta})'H'A'\right] \\ &= E[(\hat{\tau}_{GE} - \tau_G)(\hat{\tau}_{GE} - \tau_G)'] = MSE(\hat{\tau}_{GE}), \end{aligned}$$

where

$$\tau_G = A_G\boldsymbol{\eta} = A_G(\boldsymbol{X\beta}+\boldsymbol{Zu}), \quad \hat{\tau}_{GE} = A_G\hat{\boldsymbol{\eta}} = A_G(\boldsymbol{X}\hat{\boldsymbol{\beta}}+\boldsymbol{Z}\hat{\boldsymbol{u}}), \quad A_G = AH.$$

In the case of using the maximum penalized likelihood approach for estimating $\boldsymbol{\beta}$ and φ_i, $i = 1,\ldots,m$, we do following:

1. For estimating $\boldsymbol{\beta}$ and predicting $\boldsymbol{u}$, maximize $l_1(\boldsymbol{y}|\boldsymbol{u}) + l_2(\boldsymbol{u})$.

2. For estimating $\varphi_1,\ldots,\varphi_m$, maximize the loglikelihood of the distribution $N(\boldsymbol{X\beta},\, \boldsymbol{V})$, where

$$\boldsymbol{V} = \boldsymbol{Z\Sigma}_u\boldsymbol{Z}' + \boldsymbol{W}^{-1}, \quad \boldsymbol{W} = -E\left[\frac{\partial^2 l_1(\boldsymbol{y}|\boldsymbol{u})}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}'}\right]\Bigg|_{\eta=\eta^\circ}, \quad \eta^\circ = \boldsymbol{X\beta}^\circ + \boldsymbol{Zu}^\circ.$$

   and $(\boldsymbol{\beta}^\circ, \boldsymbol{u}^\circ)$ are the values of $(\boldsymbol{\beta},\boldsymbol{u})$ maximizing $l_1(\boldsymbol{\beta},\boldsymbol{u}) = l_1(\boldsymbol{y}|\boldsymbol{u})$.

Observe that in step 2 we assume a normal mixed model with $\boldsymbol{\Sigma}_e^{-1} = \boldsymbol{W}$. In what follows, we show how to estimate the mean squared error $MSE(\hat{\tau}_{GE})$.

Let $\hat{\tau}_{GB} = A_G\hat{\boldsymbol{\eta}} = A_G(\boldsymbol{X}\hat{\boldsymbol{\beta}}+\boldsymbol{Z}\hat{\boldsymbol{u}})$ be the estimator of τ_G that is obtained from maximizing $\widetilde{l}_1(\widetilde{\boldsymbol{y}}|\boldsymbol{u}) + l_2(\boldsymbol{u})$ when the variance components φ_i, $i = 1,\ldots,m$, are known. For the predictor $\hat{\tau}_{GE}$, the MSE is

$$\begin{aligned} MSE(\hat{\tau}_{GE}) &= MSE(\hat{\tau}_{GB}) + E[(\hat{\tau}_{GE}-\hat{\tau}_{GB})(\hat{\tau}_{GE}-\hat{\tau}_{GB})'] \\ &+ E[(\hat{\tau}_{GE}-\hat{\tau}_{GB})(\hat{\tau}_{GB}-\tau_G)'] + E[(\hat{\tau}_{GB}-\tau_G)(\hat{\tau}_{GE}-\hat{\tau}_{GB})']. \end{aligned}$$

As the estimators $\hat{\varphi}_j$, that maximize $\widetilde{l}_1(\widetilde{\boldsymbol{y}}|\boldsymbol{u}) + l_2(\boldsymbol{u})$, are translation invariant, we have

$$MSE(\hat{\tau}_{GE}) = MSE(\hat{\tau}_{GB}) + E[(\hat{\tau}_{GE}-\hat{\tau}_{GB})(\hat{\tau}_{GE}-\hat{\tau}_{GB})'].$$

Furthermore, the MSE of the $\hat{\tau}_{GB}$ is

$$MSE(\hat{\tau}_{GB}) = \left[\begin{array}{cc} A_G\boldsymbol{X} & A_G\boldsymbol{Z} \end{array}\right] I_F^{-1}(\boldsymbol{\beta},\boldsymbol{u}) \left[\begin{array}{c} \boldsymbol{X}'A_G' \\ \boldsymbol{Z}'A_G' \end{array}\right],$$

where $I_F(\boldsymbol{\beta},\boldsymbol{u})$ is the Fisher information matrix derived from the joint p.d.f $f^\circ(\boldsymbol{y}^\circ,\boldsymbol{u}) = \exp\{l_1^\circ(\boldsymbol{y}^\circ|\boldsymbol{u}) + l_2(\boldsymbol{u})\}$. It holds that

$$I_F(\boldsymbol{\beta},\boldsymbol{u}) = \left[\begin{array}{cc} \boldsymbol{X}'\boldsymbol{W}\boldsymbol{X} & \boldsymbol{X}'\boldsymbol{W}\boldsymbol{Z} \\ \boldsymbol{Z}'\boldsymbol{W}\boldsymbol{X} & \boldsymbol{Z}'\boldsymbol{W}\boldsymbol{Z}+\boldsymbol{\Sigma}_u^{-1} \end{array}\right].$$

Define $\boldsymbol{T} = \left(\boldsymbol{\Sigma}_u^{-1} + \boldsymbol{Z}'\boldsymbol{W}\boldsymbol{Z}\right)^{-1}$ and $\boldsymbol{V} = \boldsymbol{Z\Sigma}_u\boldsymbol{Z}' + \boldsymbol{W}^{-1}$. The inverse of $I_F(\boldsymbol{\beta},\boldsymbol{u})$ is

$$I_F(\boldsymbol{\beta},\boldsymbol{u})^{-1} = \left[\begin{array}{cc} \boldsymbol{P} & \boldsymbol{Q} \\ \boldsymbol{Q}' & \boldsymbol{R} \end{array}\right],$$

where $\boldsymbol{R} = \boldsymbol{T} + \boldsymbol{TZ}'\boldsymbol{WXPX}'\boldsymbol{WZT}$, $\boldsymbol{Q} = -\boldsymbol{PX}'\boldsymbol{WZT}$ and $\boldsymbol{P} = (\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X})^{-1}$. Therefore

$$MSE(\hat{\tau}_{GB}) \cong \mathcal{G}_1(\boldsymbol{\varphi}) + \mathcal{G}_2(\boldsymbol{\varphi}) + \mathcal{G}_3(\boldsymbol{\varphi}),$$

where $\mathcal{G}_1(\boldsymbol{\varphi})$ and $\mathcal{G}_2(\boldsymbol{\varphi})$ are

$$\begin{aligned}
\mathcal{G}_1(\boldsymbol{\varphi}) &= \boldsymbol{AHZTZ'H'A'}, \\
\mathcal{G}_2(\boldsymbol{\varphi}) &= [\boldsymbol{AHX} - \boldsymbol{AHZTZ'WX}]\boldsymbol{Q}[\boldsymbol{X'H'A'} - \boldsymbol{X'WZTZ'H'A'}].
\end{aligned}$$

In a similar way, we can adapt the results of Prasad and Rao (1990) to obtain the approximation

$$\mathcal{G}_3(\boldsymbol{\varphi}) = E[(\hat{\tau}_{GE} - \hat{\tau}_{GB})(\hat{\tau}_{GE} - \hat{\tau}_{GB})'] \approx \sum_{k_1=1}^{m}\sum_{k_2=1}^{m} \operatorname{cov}(\hat{\varphi}_{k_1}, \hat{\varphi}_{k_2})\boldsymbol{AHL}^{(k_1)}\boldsymbol{VL}^{(k_2)\prime}\boldsymbol{H'A'},$$

where

$$\begin{aligned}
\boldsymbol{T} &= \boldsymbol{V}_u - \boldsymbol{V}_u\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{Z}\boldsymbol{V}_u, \quad \boldsymbol{Q} = (\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}, \\
\boldsymbol{L}^{(k)} &= (\boldsymbol{I} - \boldsymbol{R})\dot{\boldsymbol{V}}_k\boldsymbol{V}^{-1}, \quad \dot{\boldsymbol{V}}_k = \frac{\partial \boldsymbol{V}}{\partial \varphi_k}, \quad \boldsymbol{R} = \boldsymbol{V}_u\boldsymbol{V}^{-1}.
\end{aligned}$$

Chapter 4

Multinomial logit mixed model

This chapter introduces the multinomial logit mixed model that will be used for the estimation of totals of employed and unemployed people and of unemployment rates in the $D = 96$ domains defined by the combination of the 48 counties of Galicia and the two sexes. Section 4.1 describes the model, gives an algorithm to estimate the model parameters and proposes small area estimators of labor market indicators. It also shows how estimating the mean square errors. Section 4.2 presents two simulation experiments. The first simulation compares the small area estimators obtained under the introduced multinomial model with the corresponding ones obtained under the multinomial model of Molina et al. (2007) and under independent binomial logit mixed models. The second simulation studies the behavior of the three introduced MSE estimation methods. Finally, Section 4.3 applies the proposed methodology to data from the SLFS in Galicia.

4.1 The model

This section introduces an area-level multinomial logit mixed model with domain random effects associated to the categories of the response variable. We first give some notation and assumptions. Let use indexes $k = 1, \dots, q-1$ for the $q-1$ multinomial categories of the target variable and $d = 1, \dots, D$ for the D domains. In the real data presented in Section 4.3 there are $q = 3$ categories, i.e. employed, unemployed and inactive people. However, there are only $q-1 = 2$ categories in the multinomial model, i.e. employed and unemployed people. Let u_{dk} be the random effect associated to category k and domain d, and define $\boldsymbol{u}_d = \underset{1\leq k\leq q-1}{\text{col}}(u_{dk})$. We assume that $\boldsymbol{u}_d \sim N(\boldsymbol{0}, \boldsymbol{V}_{u_d})$ are independent with covariance matrices $\boldsymbol{V}_{u_d} = \underset{1\leq k\leq q-1}{\text{diag}}(\varphi_k)$. We further assume that the response vectors $\boldsymbol{y}_d = \underset{1\leq k\leq q-1}{\text{col}}(y_{dk})$, conditioned to $\boldsymbol{u}_d$, are independent with multinomial distributions

$$\boldsymbol{y}_d|_{\boldsymbol{u}_d} \sim \text{M}(\nu_d, p_{d1}, \dots, p_{dq-1}), \;\; d = 1, \dots, D, \tag{4.1.1}$$

where the ν_d's are known integer numbers. The covariance matrix of $\boldsymbol{y}_d$ conditioned to $\boldsymbol{u}_d$ is $\boldsymbol{W}_d = \text{var}(\boldsymbol{y}_d|\boldsymbol{u}_d) = \nu_d[\text{diag}(\boldsymbol{p}_d) - \boldsymbol{p}_d\boldsymbol{p}_d']$, where $\boldsymbol{p}_d = \underset{1\leq k\leq q-1}{\text{col}}(p_{dk})$ and $\text{diag}(\boldsymbol{p}_d) =$

$\operatorname*{diag}_{1\leq k\leq q-1}(p_{dk})$. For the natural parameters $\eta_{dk} = \log(\frac{p_{dk}}{p_{dq}})$, we assume the model

$$\eta_{dk} = \boldsymbol{x}_{dk}\boldsymbol{\beta}_k + u_{dk}, \;\; d = 1, \ldots, D, \; k = 1, \ldots, q-1, \tag{4.1.2}$$

where $\boldsymbol{x}_{dk} = \operatorname*{col}'_{1\leq r\leq l_k}(x_{dkr})$, $\boldsymbol{\beta}_k = \operatorname*{col}_{1\leq r\leq l_k}(\beta_{kr})$ and $l = \sum_{k=1}^{q-1} l_k$. The mean and the variance of y_{dk}, conditioned to $\boldsymbol{u}_d$, are $\mu_{dk} = \nu_d p_{dk}$ and $\omega_{dk} = \nu_d p_{dk}(1-p_{dk})$, respectively. The probability of the multinomial category k in the domain d is

$$p_{dk} = \frac{\exp\{\eta_{dk}\}}{1 + \sum_{\ell=1}^{q-1}\exp\{\eta_{d\ell}\}}, \;\; d = 1, \ldots, D, \; k = 1, \ldots, q-1.$$

In matrix notation, the model is

$$\boldsymbol{\eta}_d = \boldsymbol{X}_d\boldsymbol{\beta} + \boldsymbol{Z}_d\boldsymbol{u}_d, \;\; d = 1, \ldots, D, \tag{4.1.3}$$

where $\boldsymbol{\eta}_d = \operatorname*{col}_{1\leq k\leq q-1}(\eta_{dk})$, $\boldsymbol{X}_d = \operatorname*{diag}_{1\leq k\leq q-1}(\boldsymbol{x}_{dk})$, $\boldsymbol{\beta} = \operatorname*{col}_{1\leq k\leq q-1}(\boldsymbol{\beta}_k)$ and $\boldsymbol{Z}_d = \boldsymbol{I}_{q-1}$ with $\boldsymbol{I}_{q-1}$ being the $(q-1)\times(q-1)$ unit matrix. If we introduce the additional notation $\boldsymbol{\eta} = \operatorname*{col}_{1\leq d\leq D}(\boldsymbol{\eta}_d)$, $\boldsymbol{X} = \operatorname*{col}_{1\leq d\leq D}(\boldsymbol{X}_d)$, $\boldsymbol{u} = \operatorname*{col}_{1\leq d\leq D}(\boldsymbol{u}_d)$ and $\boldsymbol{Z} = \operatorname*{diag}_{1\leq d\leq D}(\boldsymbol{Z}_d)$, then (4.1.3) can be represented in the matrix form

$$\boldsymbol{\eta} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u}, \tag{4.1.4}$$

with $\boldsymbol{V}_u = \operatorname{var}(\boldsymbol{u}) = \operatorname*{diag}_{1\leq d\leq D}(\boldsymbol{V}_{u_d})$ and $\boldsymbol{W} = \operatorname{var}(\boldsymbol{y}|\boldsymbol{u}) = \operatorname*{diag}_{1\leq d\leq D}(\boldsymbol{W}_d)$.

Alternatively, (4.1.3) can be expressed in the form

$$\boldsymbol{\eta}_d = \boldsymbol{X}_d\boldsymbol{\beta} + \mathsf{Z}_d\mathsf{u}, \;\; d = 1, \ldots, D, \tag{4.1.5}$$

where $\mathsf{u} = \operatorname*{col}_{1\leq k\leq q-1}(\boldsymbol{u}_k)$, the $\boldsymbol{u}_k = \operatorname*{col}_{1\leq d\leq D}(u_{dk}) \sim N_D(\boldsymbol{0}, \varphi_k\boldsymbol{I}_D)$ are independent and $\mathsf{Z}_d = \operatorname*{diag}_{1\leq k\leq q-1}(\operatorname*{col}'_{1\leq\ell\leq D}(\delta_{\ell d}))$. We also can express (4.1.5) as

$$\boldsymbol{\eta} = \boldsymbol{X}\boldsymbol{\beta} + \mathsf{Z}\mathsf{u}, \tag{4.1.6}$$

where $\mathsf{Z} = \operatorname*{col}_{1\leq d\leq D}(\mathsf{Z}_d)$. We recall that (4.1.4) and (4.1.6) define the same model.

4.1.1 The PQL-REML fitting algorithm

To fit the introduced multinomial mixed model we combine the PQL method, described by Breslow and Clayton (1993) for estimating and predicting the $\boldsymbol{\beta}_k$'s and the $\boldsymbol{u}_d$'s, with the REML method for estimating the variance components φ_k's. The presented method is based on a normal approximation to the joint probability distribution of the vector $(\boldsymbol{y}, \boldsymbol{u})$. The combined algorithm was described in Section 3.2. It was first introduced by Schall (1991) and later used by Saei and Chambers (2003) and Molina et al. (2007) in applications of generalized linear mixed models to small area estimation problems. In this chapter, we adapt the combined algorithm to the multinomial logit mixed model defined by (4.1.1) and (4.1.2).

The log-likelihood of $\boldsymbol{y}$ conditioned to $\boldsymbol{u}$ is

$$\begin{aligned} l_1(\boldsymbol{y}|\boldsymbol{u}) &= \sum_{d=1}^{D}\left\{\sum_{k=1}^{q-1} y_{dk}\log\frac{p_{dk}}{p_{dq}} + \nu_d\log p_{dq} + \log\frac{\nu_d}{y_{d1}!\cdots y_{dq}!}\right\} \\ &= \sum_{d=1}^{D}\left\{\sum_{k=1}^{q-1} y_{dk}\eta_{dk} - \nu_d\log\big(1+\sum_{k=1}^{q-1}\exp\{\eta_{dk}\}\big) + \log\frac{\nu_d}{y_{d1}!\cdots y_{dq}!}\right\}. \end{aligned}$$

The partial derivatives of

$$\eta_{dk} = \sum_{r=1}^{p_k} x_{dkr}\beta_{kr} + u_{dk}, d=1,\ldots,D, \ , k=1,\ldots,q-1,$$

with respect to β_{kr} and u_{dk} are

$$\frac{\partial\eta_{dk}}{\partial\beta_{kr}} = x_{dkr}, \quad \frac{\partial\eta_{dk}}{\partial u_{dk}} = 1.$$

The first order partial derivatives of l_1 are

$$\begin{aligned} S_{1,\beta_{kr}} &= \frac{\partial l_1(\boldsymbol{y}|\boldsymbol{u})}{\partial\beta_{kr}} = \sum_{d=1}^{D}\left\{x_{dkr}y_{dk} - \frac{\nu_d x_{dkr}\exp\{\eta_{dk}\}}{1+\sum_{\ell=1}^{q-1}\exp\{\eta_{d\ell}\}}\right\} = \sum_{d=1}^{D} x_{dkr}(y_{dk}-\mu_{dk}), \\ S_{1,u_{dk}} &= \frac{\partial l_1(\boldsymbol{y}|\boldsymbol{u})}{\partial u_{dk}} = (y_{dk}-\mu_{dk}). \end{aligned}$$

The vector expressions of the first order partial derivatives of l_1 are

$$\begin{aligned} \boldsymbol{S}_{1,\beta_k} &= \frac{\partial l_1(\boldsymbol{y}|\boldsymbol{u})}{\partial\boldsymbol{\beta}_k} = \sum_{d=1}^{D}\boldsymbol{x}'_{dk}(y_{dk}-\mu_{dk}), \quad \boldsymbol{S}_{1,\beta} = \underset{1\le k\le q-1}{\text{col}}(\boldsymbol{S}_{1,\beta_k}), \\ \boldsymbol{S}_{1,u} &= \frac{\partial l_1(\boldsymbol{y}|\boldsymbol{u})}{\partial\boldsymbol{u}} = \underset{1\le d\le D}{\text{col}}\big(\underset{1\le k\le q-1}{\text{col}}(S_{1,u_{dk}})\big). \end{aligned}$$

The second order partial derivatives of l_1 are

$$\begin{aligned} H_{1,\beta_{kr}\beta_{ks}} &= \frac{\partial^2 l_1(\boldsymbol{y}|\boldsymbol{u})}{\partial\beta_{kr}\partial\beta_{ks}} = -\sum_{d=1}^{D}\nu_d x_{dkr}x_{dks}p_{dk}(1-p_{dk}), \\ H_{1,u_{dk}\beta_{kr}} &= \frac{\partial^2 l_1(\boldsymbol{y}|\boldsymbol{u})}{\partial u_{dk}\partial\beta} = -\nu_d x_{dkr}p_{dk}(1-p_{dk}), \\ H_{1,u_{dk},u_{dk}} &= \frac{\partial^2 l_1(\boldsymbol{y}|\boldsymbol{u})}{\partial u_{dk}\partial u_{dk}} = -\nu_d p_{dk}(1-p_{dk}), \end{aligned}$$

and for the case $k_1 \neq k_2$, we have

$$\begin{aligned} H_{1,\beta_{k_1r}\beta_{k_2s}} &= \frac{\partial^2 l_1(\boldsymbol{y}|\boldsymbol{u})}{\partial\beta_{k_1r}\partial\beta_{k_2s}} = \sum_{d=1}^{D}\nu_d x_{dk_1r}x_{dk_2s}p_{dk_1}p_{dk_2}, \\ H_{1,u_{dk_1}\beta_{k_2s}} &= \frac{\partial^2 l_1(\boldsymbol{y}|\boldsymbol{u})}{\partial u_{dk_1}\partial\beta_{k_2s}} = \nu_d x_{dk_2s}p_{dk_1}p_{dk_2}, \\ H_{1,u_{dk_1},u_{dk_2}} &= \frac{\partial^2 l_1(\boldsymbol{y}|\boldsymbol{u})}{\partial u_{dk_1}\partial u_{dk_2}} = \nu_d p_{dk_1}p_{dk_2}, \quad H_{1,u_{d_1k_1},u_{d_2k_2}} = \frac{\partial^2 l_1(\boldsymbol{y}|\boldsymbol{u})}{\partial u_{dk_1}\partial u_{dk_2}} = 0, \ d_1 \neq d_2. \end{aligned}$$

The vector expressions of the second order partial derivatives of l_1 are

$$\boldsymbol{H}_{1,\beta_{k_1}\beta_{k_2}} = \Big(H_{1,\beta_{k_1 r}\beta_{k_2 s}}\Big)_{\substack{r=1,\dots,p_{k_1}\\ s=1,\dots,p_{k_2}}}, \quad \boldsymbol{H}_{1,\beta\beta} = \Big(\boldsymbol{H}_{1,\beta_{k_1}\beta_{k_2}}\Big)_{\substack{k_1=1,\dots,q-1\\ k_2=1,\dots,q-1}},$$
$$\boldsymbol{H}_{1,u_d\beta_{k_2}} = \Big(H_{1,u_{dk_1}\beta_{k_2 s}}\Big)_{\substack{k_1=1,\dots,q-1\\ s=1,\dots,p_{k_2}}}, \quad \boldsymbol{H}_{1,u\beta} = \Big(\boldsymbol{H}_{1,u_d\beta_{k_2}}\Big)_{\substack{d=1,\dots,D\\ k_2=1,\dots,q-1}},$$
$$\boldsymbol{H}_{1,u_{d_1}u_{d_2}} = \Big(H_{1,u_{d_1k_1},u_{d_2k_2}}\Big)_{\substack{k_1=1,\dots,q-1\\ k_2=1,\dots,q-1}}, \quad \boldsymbol{H}_{1,uu} = \Big(\boldsymbol{H}_{1,u_{d_1}u_{d_2}}\Big)_{\substack{d_1=1,\dots,D\\ d_2=1,\dots,D}}.$$

It holds that $\boldsymbol{H}_{1,uu} = \underset{1\le d\le D}{\text{diag}}\,(\boldsymbol{H}_{1,u_du_d})$.

The likelihood of $\boldsymbol{u}$ is

$$f_2(\boldsymbol{u}) = \frac{1}{(2\pi)^{D(q-1)/2}|\boldsymbol{V}_u|^{1/2}} \exp\left\{-\frac{1}{2}\,\boldsymbol{u}'\boldsymbol{V}_u^{-1}\boldsymbol{u}\right\}.$$

The loglikelihood of $\boldsymbol{u}$ is

$$l_2(\boldsymbol{u}) = \kappa - \sum_{d=1}^{D}\sum_{k=1}^{q-1}\left\{\frac{D}{2}\log\varphi_k + \frac{1}{2}\frac{u_{dk}^2}{\varphi_k}\right\},$$

where $\kappa = -\frac{D(q-1)}{2}\log 2\pi$. The first order partial derivatives of l_2 are

$$S_{2,u_{dk}} = \frac{\partial l_2(\boldsymbol{u})}{\partial u_{dk}} = -\frac{1}{\varphi_k}\,u_{dk}.$$

The vector expression of partial derivatives of l_2 are

$$\boldsymbol{S}_{2,u} = \underset{1\le d\le D}{\text{col}}\big(\underset{1\le k\le q-1}{\text{col}}\big(S_{2,u_{dk}}\big)\big).$$

The second order partial derivatives of l_2 are

$$H_{2,u_{dk}u_{dk}} = -\frac{1}{\varphi_k}, \quad H_{2,u_{d_1k_1}u_{d_2k_2}} = 0,\ d_1\neq d_2 \text{ or } k_1\neq k_2.$$

The matrix of the second order partial derivatives of l_2 are

$$\boldsymbol{H}_{2,u_{d_1}u_{d_2}} = \Big(H_{2,u_{d_1k_1},u_{d_2k_2}}\Big)_{\substack{k_1=1,\dots,q-1\\ k_2=1,\dots,q-1}}, \quad \boldsymbol{H}_{2,uu} = \Big(\boldsymbol{H}_{2,u_{d_1}u_{d_2}}\Big)_{\substack{d_1=1,\dots,D\\ d_2=1,\dots,D}}.$$

The log-likelihood of $(\boldsymbol{y},\boldsymbol{u})$ is

$$l(\boldsymbol{y},\boldsymbol{u}) = l_1(\boldsymbol{y}|\boldsymbol{u}) + l_2(\boldsymbol{u}).$$

The first order partial derivatives l are

$$\boldsymbol{S}_\beta = \boldsymbol{S}_{1,\beta}, \quad \boldsymbol{S}_u = \boldsymbol{S}_{1,u} + \boldsymbol{S}_{2,u}.$$

The blocks of the Fisher information matrix associated to l are

$$\boldsymbol{F}_{\beta,\beta}=-\boldsymbol{H}_{1,\beta\beta},\ \boldsymbol{F}_{u,\beta}=-\boldsymbol{H}_{1,u\beta},\ \boldsymbol{F}_{\beta,u}=\boldsymbol{F}'_{u,\beta},\ \boldsymbol{F}_{u,u}=-\boldsymbol{H}_{1,uu}-\boldsymbol{H}_{2,uu}.$$

We define

$$\boldsymbol{S}=\boldsymbol{S}(\boldsymbol{\beta},\boldsymbol{u})=\begin{pmatrix}\boldsymbol{S}_\beta\\ \boldsymbol{S}_u\end{pmatrix},\ \boldsymbol{F}=\boldsymbol{F}(\boldsymbol{\beta},\boldsymbol{u})=\begin{pmatrix}\boldsymbol{F}_{\beta\beta}&\boldsymbol{F}_{\beta u}\\ \boldsymbol{F}_{u\beta}&\boldsymbol{F}_{uu}\end{pmatrix},\ \boldsymbol{F}^{-1}=\begin{pmatrix}\boldsymbol{F}^{\beta\beta}&\boldsymbol{F}^{\beta u}\\ \boldsymbol{F}^{u\beta}&\boldsymbol{F}^{uu}\end{pmatrix}.$$

It holds that

$$\begin{aligned}\boldsymbol{F}^{\beta\beta}&=(\boldsymbol{F}_{\beta\beta}-\boldsymbol{F}_{\beta u}\boldsymbol{F}_{uu}^{-1}\boldsymbol{F}_{u\beta})^{-1},\quad \boldsymbol{F}^{\beta u}=-\boldsymbol{F}^{\beta\beta}\boldsymbol{F}_{\beta u}\boldsymbol{F}_{uu}^{-1},\\ \boldsymbol{F}^{u\beta}&=(\boldsymbol{F}^{\beta u})',\quad \boldsymbol{F}^{uu}=\boldsymbol{F}_{uu}^{-1}+\boldsymbol{F}_{uu}^{-1}\boldsymbol{F}_{u\beta}\boldsymbol{F}^{\beta\beta}\boldsymbol{F}_{\beta u}\boldsymbol{F}_{uu}^{-1}.\end{aligned}$$

The combined algorithm has two parts: A and B. Algorithm A updates the values of $\boldsymbol{\beta}_k$ and $\boldsymbol{u}_d$, $k=1,\dots,q-1$, $d=1,\dots,D$. Algorithm B updates the variance components.

Algorithm A.

(A.1) *Beginning:* Assign the initial values $l=0$, $\boldsymbol{\beta}^{(0)}=\boldsymbol{\beta}^{initial}$ and $\boldsymbol{u}^{(0)}=\boldsymbol{u}^{initial}$.

(A.2) *Iteration $l+1$:* For $d=1,\dots,D$, $k=1,\dots,q-1$, calculate $\boldsymbol{F}(\boldsymbol{\beta}^{(l)},\boldsymbol{u}^{(l)})$ and $\boldsymbol{S}(\boldsymbol{\beta}^{(l)},\boldsymbol{u}^{(l)})$ and update $\boldsymbol{\beta}^{(l)}$ and $\boldsymbol{u}^{(l)}$ by using the equation

$$\begin{bmatrix}\boldsymbol{\beta}^{(l+1)}\\ \boldsymbol{u}^{(l+1)}\end{bmatrix}=\begin{bmatrix}\boldsymbol{\beta}^{(l)}\\ \boldsymbol{u}^{(l)}\end{bmatrix}+\boldsymbol{F}^{-1}(\boldsymbol{\beta}^{(l)},\boldsymbol{u}^{(l)})\boldsymbol{S}(\boldsymbol{\beta}^{(l)},\boldsymbol{u}^{(l)}),\tag{4.1.7}$$

where $\boldsymbol{S}$ and $\boldsymbol{F}$ are the vector of scores (first order partial derivatives) and the Fisher information matrix (minus expectation of second order partial derivatives) of the joint log-likelihood of $(\boldsymbol{y},\boldsymbol{u})$.

(A.3) *End:* Repeat the step (A.2) till convergence of $\boldsymbol{\beta}^{(l)}$ and $\boldsymbol{u}^{(l)}$ and obtain the final values $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{u}}$.

Algorithm A maximizes $l(\boldsymbol{y},\boldsymbol{u})$ in $\boldsymbol{\beta}$ and $\boldsymbol{u}$ for fixed values of $\varphi_1,\dots,\varphi_{q-1}$. To update the values of the variance components, we assume that $\boldsymbol{\beta}$ and $\boldsymbol{u}$ are known and we adapt the ideas of Schall (1991) to a multivariate setting. For this sake, we consider a Taylor expansion of

$$\xi_{dk}=g_k(\boldsymbol{y}_d)=\log\frac{y_{dk}}{\nu_d-\sum_{\ell=1}^{q-1}y_{d\ell}}$$

around the point $\boldsymbol{\mu}_d=\nu_d\boldsymbol{p}_d$. We obtain

$$\boldsymbol{\xi}_d\approx\boldsymbol{X}_d\boldsymbol{\beta}+\boldsymbol{Z}_d\boldsymbol{u}_d+\boldsymbol{e}_d,\tag{4.1.8}$$

where $\boldsymbol{\xi}_d=\underset{1\le k\le q-1}{\text{col}}(\xi_{dk})$, $\boldsymbol{e}_d=\boldsymbol{W}_d^{-1}(\boldsymbol{y}_d-\boldsymbol{\mu}_d)$. By assuming equality in (4.1.8), it holds that $E[\boldsymbol{\xi}]=\boldsymbol{X}\boldsymbol{\beta}$ and $\boldsymbol{V}=\text{var}(\boldsymbol{\xi})=\boldsymbol{Z}\boldsymbol{V}_u\boldsymbol{Z}'+\boldsymbol{W}^{-1}$. Schall (1991) proposed to update the variance components by maximizing the normal approximation to the distribution of $\boldsymbol{\xi}$, with $\boldsymbol{\beta}$ and $\boldsymbol{u}$ fixed. This proposal assumes the approximation of $l_1(\boldsymbol{y})$ by the log-likelihood $l_1(\boldsymbol{\xi})$ of the cited multivariate normal distribution. The basic idea is thus maximizing $l(\boldsymbol{\xi}, \boldsymbol{u})$ instead of $l(\boldsymbol{y}, \boldsymbol{u})$, where $\boldsymbol{\xi}$ is assumed to follow model (4.1.8) under normality. The approximating REML log-likelihood is

$$l_{reml}(\boldsymbol{\xi}) = -\frac{1}{2}(D(q-1)-t)\log 2\pi - \frac{1}{2}\log|\boldsymbol{K}^t\boldsymbol{V}\boldsymbol{K}| - \frac{1}{2}\boldsymbol{\xi}^t\boldsymbol{P}\boldsymbol{\xi},$$

where $\boldsymbol{P} = \boldsymbol{V}^{-1} - \boldsymbol{V}^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}^{-1}$ and $\boldsymbol{K} = \boldsymbol{W} - \boldsymbol{W}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{W}$.

The update of the variance components can be done by applying the Fisher-Scoring algorithm to the REML log-likelihood. The algorithm is described below.

Algorithm B.

(B.1) Do $\ell = 1$. Initiate the values $\boldsymbol{\beta}^{(0)}$, $\boldsymbol{u}^{(0)}$ and $\boldsymbol{\varphi}^{(0)}$.

(B.2) Run the *Algorithm A* by using $\boldsymbol{\varphi}^{(\ell)}$ as known value of the vector of variances and $\boldsymbol{\beta}^{(\ell-1)}$ and $\boldsymbol{u}^{(\ell-1)}$ as initial values. Let $\boldsymbol{\beta}^{(\ell)}$ and $\boldsymbol{u}^{(\ell)}$ be the obtained estimates and predictors.

(B.3) For $d = 1, \ldots, D$, calculate $\boldsymbol{\eta}_d^{(\ell)} = \boldsymbol{X}_d\boldsymbol{\beta}^{(\ell)} + \boldsymbol{Z}_d\boldsymbol{u}^{(\ell)}$ and apply the updating equation

$$\boldsymbol{\varphi}^{(\ell+1)} = \boldsymbol{\varphi}^{(\ell)} + \boldsymbol{F}^{-1}(\boldsymbol{\varphi}^{(\ell)})\boldsymbol{S}(\boldsymbol{\varphi}^{(\ell)}),$$

where $\boldsymbol{S}$ and $\boldsymbol{F}$ are the vector of scores and the Fisher information matrix of the log-likelihood $l_{reml}(\boldsymbol{\xi})$.

(B.4) Repeat the steps (B.2)-(B.3) until the convergence of $\boldsymbol{\beta}^{(\ell)}$, $\boldsymbol{u}^{(\ell)}$ and $\boldsymbol{\varphi}^{(\ell)}$.

The variance components φ_k can be also updated by applying the formula

$$\widehat{\varphi}_k = \frac{\widehat{\boldsymbol{u}}_k'\boldsymbol{\Sigma}_{u_k}^{-1}\widehat{\boldsymbol{u}}_k}{\dim(\boldsymbol{u}_k) - \tau_k} = \frac{\widehat{\boldsymbol{u}}_k'\widehat{\boldsymbol{u}}_k}{D - \tau_k},$$

where $\boldsymbol{\Sigma}_{u_k} = \boldsymbol{I}_D$, $\tau_k = \frac{1}{\widehat{\varphi}_k}\mathrm{tr}\big(\boldsymbol{\Sigma}_{u_k}^{-1}\widehat{\mathsf{T}}_{kk}^{rml}\big) = \frac{1}{\widehat{\varphi}_k}\mathrm{tr}\big(\widehat{\mathsf{T}}_{kk}^{rml}\big)$ and T_{kk}^{rml} is the block (k,k) of the matrix

$$\widehat{\mathsf{T}}^{rml} = \widehat{\mathsf{T}} + \widehat{\mathsf{T}}\mathsf{Z}'\boldsymbol{W}\boldsymbol{X}(\boldsymbol{X}'\widehat{\boldsymbol{V}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{W}\mathsf{Z}\widehat{\mathsf{T}},$$

with $\widehat{\mathsf{T}} = (\mathsf{Z}\boldsymbol{W}\mathsf{Z}' + \widehat{\boldsymbol{\Sigma}}_u^{-1})^{-1}$, $\widehat{\boldsymbol{\Sigma}}_u = \mathrm{diag}(\varphi_1\boldsymbol{I}_D, \ldots, \varphi_{q-1}\boldsymbol{I}_D)$ and $\widehat{\boldsymbol{V}} = \mathsf{Z}\widehat{\boldsymbol{\Sigma}}_u\mathsf{Z}' + \boldsymbol{W}^{-1}$.

Algorithm B (alternative).

(B.1) Do $\ell = 1$. Initiate the values $\boldsymbol{\beta}^{(0)}$, $\boldsymbol{u}^{(0)}$ and $\boldsymbol{\varphi}^{(0)}$.

(B.2) Run the *Algorithm A* by using $\boldsymbol{\varphi}^{(\ell)}$ as known value of the variance components, and $\boldsymbol{\beta}^{(\ell-1)}$ and $\mathsf{u}^{(\ell-1)}$ as initial values. Let $\boldsymbol{\beta}^{(\ell)}$ and $\mathsf{u}^{(\ell)}$ the obtained estimators and predictors.

(B.3) Calculate $\boldsymbol{\eta}_d^{(\ell)} = \boldsymbol{X}_d\boldsymbol{\beta}^{(\ell)} + \mathsf{Z}_d\mathsf{u}^{(\ell)}$, $d = 1, \ldots, D$.

$$\begin{aligned}
p_{dk}^{(\ell)} &= \frac{\exp(\eta_{dk}^{(\ell)})}{1 + \sum_{k=1}^{q-1} \exp(\eta_{dk}^{(\ell)})}, \ \boldsymbol{p}_d^{(\ell)} = \underset{1\le k\le q-1}{\text{col}} (p_{dk}^{(\ell)}), \ \boldsymbol{W}^{(\ell)} = \underset{1\le d\le D}{\text{diag}} \left(\nu_d[\text{diag}(\boldsymbol{p}_d^{(\ell)}) - \boldsymbol{p}_d^{(\ell)}\boldsymbol{p}_d^{(\ell)\prime}]\right) \\
\boldsymbol{\Sigma}_u^{(\ell)} &= \text{diag}\big(\varphi_1^{(\ell)}\boldsymbol{I}_D, \ldots, \varphi_{q-1}^{(\ell)}\boldsymbol{I}_D\big), \ \mathsf{T}^{(\ell)} = (\mathsf{Z}'\boldsymbol{W}\mathsf{Z} + \boldsymbol{\Sigma}_u^{(\ell)-1})^{-1}, \ \boldsymbol{V}^{(\ell)} = \mathsf{Z}\boldsymbol{\Sigma}_u^{(\ell)}\mathsf{Z}' + \boldsymbol{W}^{(\ell)-1}, \\
\mathsf{T}^{rml(\ell)} &= \mathsf{T}^{(\ell)} + \mathsf{T}^{(\ell)}\mathsf{Z}'\boldsymbol{W}^{(\ell)}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{V}^{(\ell)-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{W}^{(\ell)}\mathsf{Z}\mathsf{T}^{(\ell)}, \ \tau_k^{(\ell)} = \big(\varphi_k^{(\ell)}\big)^{-1}\text{tr}\big(\mathsf{T}_{kk}^{rml(\ell)}\big)
\end{aligned}$$

(B.4) Update the variance components by using the equations

$$\widehat{\varphi}_k^{(\ell+1)} = \frac{\widehat{\boldsymbol{u}}_k^{(\ell)\prime}\widehat{\boldsymbol{u}}_k^{(\ell)}}{D - \tau_k^{(\ell)}}, \quad k = 1, \ldots, q-1.$$

(B.5) Repeat the steps (B.2)-(B.4) until the convergence of $\boldsymbol{\beta}^{(\ell)}$, $\boldsymbol{u}^{(\ell)}$ (o $\mathsf{u}^{(\ell)}$) y $\boldsymbol{\sigma}^{(\ell)}$.

The difference between this algorithm and the previous is that the last is a fixed-point algorithm and the previous is an iterative Fisher-Scoring algorithm.

The above described algorithms require initial values for $\boldsymbol{\beta}$, $\boldsymbol{u}$ and $\boldsymbol{\varphi}$. We suggest employing some easy-to-calculate estimates. More concretely, we use $\boldsymbol{u}^{(0)} = \boldsymbol{0}$ and $\boldsymbol{\beta}^{(0)} = \tilde{\boldsymbol{\beta}}$, where $\tilde{\boldsymbol{\beta}}$ is obtained by fitting the non mixed variant of the model (4.1.1)-(4.1.2) without the random effect $\boldsymbol{u}$. The non mixed model is also used for calculating $\boldsymbol{\varphi}^{(0)}$ by means of the formula

$$\hat{\varphi}_k = \frac{1}{D-1}\sum_{d=1}^{D}(\tilde{\eta}_{dk}^{(dir)} - \tilde{\eta}_{dk})^2, \quad k = 1, \ldots, q-1, \tag{4.1.9}$$

where $\tilde{\eta}_{dk} = \tilde{\boldsymbol{\beta}}_k\boldsymbol{x}_{dk}$ and $\tilde{\eta}_{dk}^{(dir)} = \log\frac{y_{dk}}{y_{dq}}$, $k = 1, \ldots, q-1$, $d = 1, \ldots, D$.

Under regularity conditions the asymptotic distribution of the REML estimator $\hat{\boldsymbol{\beta}}$ is multivariate normal $N(\boldsymbol{\beta}, \boldsymbol{F}^{\beta\beta})$, where $\boldsymbol{F}^{\beta\beta} = (q_{rr})_{r=1,\ldots,t_k}$ is the block sub-matrix of the Fisher information matrix in the output of the fitting algorithm A. Therefore, an approximate $(1-\alpha)$-level confidence interval for β_{kr} is

$$\widehat{\beta}_{kr} \pm z_{\alpha/2}q_{rr}, \quad r = 1, \ldots, t_k,$$

where z_α is the α-quantile of the normal distribution $N(0,1)$. If we use $\hat{\beta}_{kr}$ to test $H_0: \beta_{kr} = 0$ and we observe the realization $\hat{\beta}_{kr} = \beta_0$, the approximate p-value is

$$p = 2P_{H_0}(\widehat{\beta}_{kr} > |\beta_0|) = 2P(Z > |\beta_0|\,/\sqrt{q_{rr}})$$

where Z follows a standard normal distribution.

4.1.2 Model-based small area estimation

In practice we are interested in estimating the domain totals

$$Y_{dk} = \sum_{j\in P_d} y_{dkj}, \quad d = 1, \ldots, D, \ k = 1, \ldots, q-1,$$

where P_d is the population domain with size N_d. A synthetic model-based estimator of Y_{dk} is $\hat{Y}_{dk} = \hat{m}_{dk} = \hat{N}_d \hat{p}_{dk}$. Estimates of rates can be obtained by plugging the corresponding estimators of totals. In this section we deal with the problem of estimating $\boldsymbol{m}_d = N_d \boldsymbol{p}_d$, $d = 1, \ldots, D$, where N_d is an estimated population size that can be obtained from the unit-level survey data or from administrative registers. As we are assuming the area-level model approach, N_d is treated as a known constant. In the application to real data, we take $N_d = \hat{N}_d^{dir}$ because $\hat{N}_d^{dir}$ is the official best estimate of the domain size. We further estimate $\boldsymbol{m}_d$ by means of $\hat{\boldsymbol{m}}_d = N_d \hat{\boldsymbol{p}}_d$, where

$$\hat{\boldsymbol{p}}_d = \underset{1 \leq k \leq q-1}{\text{col}} (\hat{p}_{dk}), \quad \hat{p}_{dk} = \frac{\exp\{\hat{\eta}_{dk}\}}{1 + \sum_{\ell=1}^{q-1} \exp\{\hat{\eta}_{d\ell}\}}, \quad \hat{\eta}_{dk} = \boldsymbol{x}_{dk} \hat{\boldsymbol{\beta}}_k + \hat{u}_{dk}.$$

Note that $\boldsymbol{m}_d$ can be written in the form $\boldsymbol{m}_d = h_d(\boldsymbol{\eta}_d)$, where $h_d = \underset{1 \leq k \leq q-1}{\text{col}} (h_{dk})$ and

$$h_{dk}(\boldsymbol{\eta}_d) = \hat{N}_d p_{dk} = \hat{N}_d \frac{\exp\{\eta_{dk}\}}{1 + \sum_{\ell=1}^{q-1} \exp\{\eta_{d\ell}\}}, \quad k = 1, \ldots, q-1.$$

The partial derivatives of h_{dk} are

$$\frac{\partial h_{dk}}{\partial \eta_{dk}} = \hat{N}_d p_{dk}(1 - p_{dk}), \quad \frac{\partial h_{dk_1}}{\partial \eta_{dk_2}} = -\hat{N}_d p_{dk_1} p_{dk_2}, \; k_1 \neq k_2,$$

and the matrix of derivatives is

$$\boldsymbol{H}_d = \left(\frac{\partial h_{dk_1}}{\partial \eta_{dk_2}} \right)_{k_1, k_2 = 1, \ldots, q-1} = \hat{N}_d [\text{diag}(\boldsymbol{p}_d) - \boldsymbol{p}_d \boldsymbol{p}_d'].$$

A Taylor expansion of $h_{dk}(\hat{\boldsymbol{\eta}}_d)$ around $\boldsymbol{\eta}_d$ yield to

$$h_{dk}(\hat{\boldsymbol{\eta}}_d) - h_{dk}(\boldsymbol{\eta}_d) \approx \sum_{\ell=1}^{q-1} \frac{\partial h_{dk}}{\partial \eta_{d\ell}} (\hat{\boldsymbol{\eta}}_{d\ell} - \boldsymbol{\eta}_{d\ell}),$$

or equivalently to

$$h_d(\hat{\boldsymbol{\eta}}_d) - h_d(\boldsymbol{\eta}_d) \approx \boldsymbol{H}_d (\hat{\boldsymbol{\eta}}_d - \boldsymbol{\eta}_d), \quad h(\hat{\boldsymbol{\eta}}) - h(\boldsymbol{\eta}) \approx \boldsymbol{H} (\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}).$$

where $\boldsymbol{H} = \underset{1 \leq d \leq D}{\text{diag}} (\boldsymbol{H}_d)$ and $h = \underset{1 \leq d \leq D}{\text{col}} (h_d)$. As $\hat{\boldsymbol{m}}_d = \boldsymbol{A} \boldsymbol{H} \hat{\boldsymbol{\eta}}$, with $\boldsymbol{A} = \underset{1 \leq d_1 \leq D}{\text{col}'} (\delta_{dd_1} \boldsymbol{I}_{q-1})$, then $\hat{\boldsymbol{\eta}}$ can be viewed as a vector of EBLUPs in the linear mixed model (4.1.8). Therefore, we propose to apply the methodology of Prasad and Rao (1990) as explained in Chapter 3 (Section 3.3). This is to say, we approximate the MSE of $\hat{\boldsymbol{m}}_d$ by means of

$$MSE(\hat{\boldsymbol{m}}_d) \approx \mathcal{G}_1(\boldsymbol{\varphi}) + \mathcal{G}_2(\boldsymbol{\varphi}) + \mathcal{G}_3(\boldsymbol{\varphi}). \tag{4.1.10}$$

where

$$\begin{aligned}
\mathcal{G}_1(\boldsymbol{\varphi}) &= \boldsymbol{A} \boldsymbol{H} \boldsymbol{Z} \boldsymbol{T} \boldsymbol{Z}' \boldsymbol{H}' \boldsymbol{A}', \\
\mathcal{G}_2(\boldsymbol{\varphi}) &= [\boldsymbol{A} \boldsymbol{H} \boldsymbol{X} - \boldsymbol{A} \boldsymbol{H} \boldsymbol{Z} \boldsymbol{T} \boldsymbol{Z}' \boldsymbol{W} \boldsymbol{X}] \boldsymbol{Q} [\boldsymbol{X}' \boldsymbol{H}' \boldsymbol{A}' - \boldsymbol{X}' \boldsymbol{W} \boldsymbol{Z} \boldsymbol{T} \boldsymbol{Z}' \boldsymbol{H}' \boldsymbol{A}'], \\
\mathcal{G}_3(\boldsymbol{\varphi}) &\approx \sum_{k_1}^{q-1} \sum_{k_2}^{q-1} \text{cov}(\hat{\varphi}_{k_1}, \hat{\varphi}_{k_2}) \boldsymbol{A} \boldsymbol{H} \boldsymbol{L}^{(k_1)} \boldsymbol{V} \boldsymbol{L}^{(k_2)\prime} \boldsymbol{H}' \boldsymbol{A}',
\end{aligned}$$

$$
\begin{aligned}
\boldsymbol{T} &= \boldsymbol{V}_u - \boldsymbol{V}_u\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{Z}\boldsymbol{V}_u, \quad \boldsymbol{Q} = (\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1},\\
\boldsymbol{V}_u &= \underset{1\le\ell\le D}{\text{diag}}\,(\boldsymbol{V}_{u\ell}), \quad \boldsymbol{V}_{u\ell} = \underset{1\le k\le q-1}{\text{diag}}\,(\varphi_k), \quad \boldsymbol{V} = \boldsymbol{Z}\boldsymbol{V}_u\boldsymbol{Z}' + \boldsymbol{W}^{-1} = \underset{1\le\ell\le D}{\text{diag}}\,(\boldsymbol{V}_\ell),\\
\boldsymbol{Z} &= \underset{1\le\ell\le D}{\text{diag}}\,(\boldsymbol{Z}_\ell), \quad \boldsymbol{Z}_\ell = \boldsymbol{I}_{q-1}, \quad \boldsymbol{T} = \underset{1\le\ell\le D}{\text{diag}}\,(\boldsymbol{T}_\ell), \quad \boldsymbol{T}_\ell = \boldsymbol{V}_{u\ell} - \boldsymbol{V}_{u\ell}\boldsymbol{V}_\ell^{-1}\boldsymbol{V}_{u\ell},\\
\boldsymbol{L}^{(k)} &= (\boldsymbol{I}-\boldsymbol{R})\boldsymbol{V}_k\boldsymbol{V}^{-1}, \quad \boldsymbol{V}_k = \frac{\partial \boldsymbol{V}}{\partial\varphi_k} = \underset{1\le d\le D}{\text{diag}}\,(\underset{1\le i\le q-1}{\text{diag}}\,(\delta_{ik})), \quad \boldsymbol{R} = \boldsymbol{V}_u\boldsymbol{V}^{-1}.
\end{aligned}
$$

It holds that

$$
\mathcal{G}_1(\boldsymbol{\varphi}) = \boldsymbol{I}_{q-1}\boldsymbol{H}_d\boldsymbol{I}_{q-1}\boldsymbol{T}_d\boldsymbol{I}_{q-1}\boldsymbol{H}_d\boldsymbol{I}_{q-1} = \boldsymbol{H}_d\boldsymbol{T}_d\boldsymbol{H}_d.
$$

The expression of $\mathcal{G}_2(\boldsymbol{\varphi})$ is

$$
\mathcal{G}_2(\boldsymbol{\varphi}) = [\boldsymbol{A}_{21} - \boldsymbol{A}_{22}]\boldsymbol{Q}[\boldsymbol{A}_{21}' - \boldsymbol{A}_{22}'],
$$

where

$$
\boldsymbol{A}_{21} = \boldsymbol{I}_{q-1}\boldsymbol{H}_d\boldsymbol{X}_d = \boldsymbol{H}_d\boldsymbol{X}_d, \quad \boldsymbol{A}_{22} = \boldsymbol{I}_{q-1}\boldsymbol{H}_d\boldsymbol{I}_{q-1}\boldsymbol{T}_d\boldsymbol{I}_{q-1}\boldsymbol{W}_d\boldsymbol{X}_d = \boldsymbol{H}_d\boldsymbol{T}_d\boldsymbol{H}_d\boldsymbol{X}_d.
$$

In the calculation of $\mathcal{G}_3(\boldsymbol{\varphi})$, we observe that $\boldsymbol{L}^{(k)} = \underset{1\le d\le D}{\text{diag}}\,(\boldsymbol{L}_d^{(k)})$, where

$$
\boldsymbol{L}_d^{(k)} = (\boldsymbol{I}_{q-1} - \underset{1\le i\le q-1}{\text{diag}}\,(\varphi_i)\boldsymbol{V}_d^{-1})\underset{1\le i\le q-1}{\text{diag}}\,(\delta_{ik})\boldsymbol{V}_d^{-1}.
$$

and $\text{cov}(\hat{\varphi}_{k_1}, \hat{\varphi}_{k_2})$ is obtained from the inverse of the Fisher information matrix $\boldsymbol{F}$ in the algorithm B. The proposed analytic MSE estimator is

$$
mse(\hat{\boldsymbol{m}}_d) = \mathcal{G}_1(\hat{\boldsymbol{\varphi}}) + \mathcal{G}_2(\hat{\boldsymbol{\varphi}}) + 2\mathcal{G}_3(\hat{\boldsymbol{\varphi}}), \tag{4.1.11}
$$

where $\hat{\boldsymbol{\varphi}}$ is the estimator of $\boldsymbol{\varphi}$ obtained from algorithm B.

Concerning the estimation of the MSE of $\hat{m}_{dk}$, we can also use the approach of González-Manteiga et al.(2008a) by introducing the following parametric bootstrap method.

1. Fit the model (4.1.1)-(4.1.2) and calculate $\hat{\varphi}_k$ and $\hat{\boldsymbol{\beta}}_k$, $k = 1, \ldots, q-1$.
2. Generate $\boldsymbol{u}_d^* \sim N(\boldsymbol{0}, \underset{1\le k\le q-1}{\text{diag}}\,(\hat{\varphi}_k))$ and $\boldsymbol{y}_d^* \sim \text{M}(\nu_d, p_{d1}^*, \ldots, p_{dq-1}^*)$, where

$$
p_{dk}^* = \frac{\exp\{\eta_{dk}^*\}}{1+\sum_{\ell=1}^{q-1}\exp\{\eta_{d\ell}^*\}}, \eta_{dk}^* = \hat{\boldsymbol{\beta}}_k x_{dk} + u_{dk}^*, m_{dk}^* = \hat{N}_d p_{dk}^*, d = 1, \ldots, D, k = 1, \ldots, q-1.
$$

3. Calculate $\hat{\varphi}_k^*$, $\hat{\boldsymbol{\beta}}_k^*$ and

$$
\hat{p}_{dk}^* = \frac{\exp\{\hat{\eta}_{dk}^*\}}{1+\sum_{\ell=1}^{q-1}\exp\{\hat{\eta}_{d\ell}^*\}}, \hat{\eta}_{dk}^* = \hat{\boldsymbol{\beta}}_k^* x_{dk} + \hat{u}_{dk}^*, \ \hat{m}_{dk}^* = \hat{N}_d \hat{p}_{dk}^*, d = 1, \ldots, D, k = 1, \ldots, q-1.
$$

4. Repeat B times steps 2-3 and calculate the bootstrap mean square error estimator

$$mse_{dk}^{*1} = \frac{1}{B}\sum_{b=1}^{B}(\hat{m}_{dk}^{*} - m_{dk}^{*})^2, \quad d = 1, \dots, D, \; k = 1, \dots, q-1.$$

In addition to mse_{dk}^{*1}, we propose another estimator. The second bootstrap MSE estimator is based on the analytic one (4.1.11) and follows the ideas of bagging from Breiman (1996). The bootstrap approximation of $MSE(\hat{\boldsymbol{m}}_d)$ is

$$mse^{*2}(\hat{\boldsymbol{m}}_d) = E_*[\mathcal{G}_1(\hat{\boldsymbol{\varphi}}^*) + \mathcal{G}_2(\hat{\boldsymbol{\varphi}}^*) + 2\mathcal{G}_3(\hat{\boldsymbol{\varphi}}^*)],$$

with the Monte Carlo approximation

$$mse^{*2}(\hat{\boldsymbol{m}}_d) = \frac{1}{B}\sum_{b=1}^{B}(\mathcal{G}_1^{*b}(\hat{\boldsymbol{\varphi}}^*) + \mathcal{G}_2^{*b}(\hat{\boldsymbol{\varphi}}^*) + 2\mathcal{G}_3^{*b}(\hat{\boldsymbol{\varphi}}^*)).$$

4.2 Simulation study

In this section we present two simulation experiments designed to analyze the behavior of the estimators $\hat{\beta}_k$, $\hat{\varphi}_k$, $\hat{p}_{dk}$ and the proposed estimators of the MSE. We consider a multinomial logit mixed model with two model categories ($q - 1 = 2$).

4.2.1 Sample simulation

For $d = 1, \dots, D$ and $k = 1, 2$, we generate the explanatory variables

$$x_{d1} = \mu_1 + \sigma_{x11}^{1/2} U_{d1}, \; x_{d2} = \mu_2 + \sigma_{x22}^{1/2}\big[\rho_x U_{d1} + \sqrt{1-\rho_x^2}U_{d2}\big], \; U_{dk} = \frac{d-D}{2D} + \frac{k}{6},$$

where $\mu_1 = \mu_2 = 1$, $\sigma_{x11} = \sigma_{x22} = 1$ and $\rho_x = 0.75$.

We generate random effects $u_{dk} \sim N(0, \varphi_k)$, with $\varphi_1 = 1$, $\varphi_2 = 2$. We also generate the linear parameter

$$\eta_{dk} = \beta_{0k} + \beta_{1k}x_{dk} + u_{dk}, \quad \text{with} \;\; \beta_{01} = 1.3, \; \beta_{02} = -1.2, \beta_{11} = -1.3, \; \beta_{12} = 1, \qquad (4.2.1)$$

and the cell probabilities

$$p_{dk} = \frac{\exp\{\eta_{dk}\}}{1 + \exp\{\eta_{d1}\} + \exp\{\eta_{d2}\}}.$$

We finally generate the response variable

$$\boldsymbol{y}_d = (y_{d1}, y_{d2})' \sim \mathrm{M}(\nu_d, p_{d1}, p_{d2}), \quad \boldsymbol{y}_d^r = (y_{d1}^r, y_{d2}^r)' \sim \mathrm{M}(\nu_d^r, p_{d1}, p_{d2}), \;\; d = 1, \dots, D,$$

with $\nu_d = 100$, $N_d = 1000$, $\nu_d^r = N_d - \nu_d$.

In matrix notation, model (4.2.1) can be written in the form

$$\begin{pmatrix} \eta_{11} \\ \eta_{12} \\ \hline \eta_{21} \\ \eta_{22} \\ \hline \vdots \\ \hline \eta_{D1} \\ \eta_{D2} \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & 0 & 0 \\ 0 & 0 & 1 & x_{12} \\ \hline 1 & x_{21} & 0 & 0 \\ 0 & 0 & 1 & x_{22} \\ \hline \vdots & & & \\ \hline 1 & x_{D1} & 0 & 0 \\ 0 & 0 & 1 & x_{D2} \end{pmatrix} \begin{pmatrix} \beta_{01} \\ \beta_{11} \\ \beta_{02} \\ \beta_{12} \end{pmatrix} + \begin{pmatrix} u_{11} \\ u_{12} \\ \hline u_{21} \\ u_{22} \\ \hline \vdots \\ \hline u_{D1} \\ u_{D2} \end{pmatrix} + \begin{pmatrix} e_{11} \\ e_{12} \\ \hline e_{21} \\ e_{22} \\ \hline \vdots \\ \hline e_{D1} \\ e_{D2} \end{pmatrix},$$

or equivalently, in the more concise notation

$$\begin{pmatrix} \boldsymbol{\eta}_1 \\ \boldsymbol{\eta}_2 \\ \vdots \\ \boldsymbol{\eta}_D \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \\ \vdots \\ \boldsymbol{X}_D \end{pmatrix} \boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u},$$

where

$$\boldsymbol{\eta}_d = \begin{pmatrix} \eta_{d1} \\ \eta_{d2} \end{pmatrix}, \boldsymbol{X}_d = \begin{pmatrix} 1 & x_{d1} & 0 & 0 \\ 0 & 0 & 1 & x_{d2} \end{pmatrix}, \boldsymbol{Z} = \boldsymbol{I}_{2D} = \begin{pmatrix} \boldsymbol{I}_2 & & 0 \\ & \ddots & \\ 0 & & \boldsymbol{I}_2 \end{pmatrix}, \boldsymbol{I}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

$$\boldsymbol{u} \sim N(0, \boldsymbol{\varphi}\boldsymbol{I}_{2D}), \, \boldsymbol{\varphi} = (\varphi_1, \varphi_2).$$

4.2.2 Simulation experiment 1: model fit

We consider three area-level models. Model A is the multinomial mixed model described in Section 4.1. Model B is the product of two independent binomial mixed models with the same explanatory variables and model parameters as model A with this expresion:

$$\eta_{d1} = \beta_{10} + \beta_{11} x_{d1} + u_{d1}, \; d = 1, \dots, D$$

$$\eta_{d2} = \beta_{20} + \beta_{21} x_{d2} + u_{d2}, \; d = 1, \dots, D$$

Model C is Model A with a common random effect for all the categories, i.e. with $u_{dk} = u_d$ in (4.1.2) and have this expresion.

$$\eta_{dk} = \beta_{k0} + \beta_{k1} x_{dk} + u_d, \; d = 1, \dots, D, \; k = 1, 2$$

Model C is the one considered by Molina et al. (2007).

Under Model A, the random effects are $u_{dk} \sim N(0, \varphi_k)$, with $\varphi_1 = 1$ and $\varphi_2 = 2$. The target variable is $\boldsymbol{y}_d = (y_{d1}, y_{d2})' \sim \mathrm{M}(\nu_d, p_{d1}, p_{d2})$, where

$$p_{dk} = \frac{\exp\{\eta_{dk}\}}{1 + \exp\{\eta_{d1}\} + \exp\{\eta_{d2}\}}, \; \eta_{dk} = \beta_{k0} + \beta_{k1} x_{dk} + u_{dk}, \; d = 1, \dots, D, \; k = 1, 2,$$

$\nu_d = 100$, $\beta_{10} = 1.3$, $\beta_{20} = -1.2$, $\beta_{11} = -1.3$ and $\beta_{21} = 1$. Under model B, data is similar to Model A, but $\boldsymbol{y}_d \sim \mathrm{M}(\nu_d, p_{d1}, p_{d2})$ is substituted by independent $y_{dk} \sim B(\nu_d, p_{dk})$, $k = 1, 2, 3$, $p_{d3} = 1 - p_{d1} - p_{d2}$ and $\hat{\nu}_d = y_{d1} + y_{d2} + y_{d3}$ are taken as multinomial size parameters for obtaining estimators based on models A and C. Under Model C, data is similar to Model A, but u_{dk} is substituted by $u_d \sim N(0, \phi)$ with $\phi = (\varphi_1 + \varphi_2)/2$.

The target of the first simulation is to analyze the behavior of the estimators $\hat{\boldsymbol{\beta}}_k$, $\hat{\varphi}_k$ and $\hat{p}_{dk}$. As efficiency measures, we consider the relative empirical bias (RBIAS) and relative mean squared error (RMSE). The simulation is described below.

1. Repeat $I = 1000$ times ($i = 1, \ldots, 1000$)

   1.1. Generate a sample (y_{dk}, x_{dk}), $d = 1, \ldots, D$, $k = 1, 2$, under Model A, B or C.

   1.2 Calculate $\hat{\beta}_{kr}^{(i)}$, $\hat{\varphi}_k^{(i)}$, $\hat{\phi}^{(i)}$ and $\hat{p}_{dk}^{(i)}$ with $r = 0, 1$, $d = 1, \ldots, D$, $k = 1, 2$.

2. For $r = 0, 1$, $d = 1, \ldots, D$ and $k = 1, 2$, calculate

$$BIAS(\hat{\beta}_{kr}) = \frac{1}{I}\sum_{i=1}^{I}(\hat{\beta}_{kr}^{(i)} - \beta_{kr}),\ BIAS(\hat{\varphi}_k) = \frac{1}{I}\sum_{i=1}^{I}(\hat{\varphi}_k^{(i)} - \varphi_k),$$

$$MSE(\hat{\beta}_{kr}) = \frac{1}{I}\sum_{i=1}^{I}(\hat{\beta}_{kr}^{(i)} - \beta_k)^2,\ MSE(\hat{\varphi}_k) = \frac{1}{I}\sum_{i=1}^{I}(\hat{\varphi}_k^{(i)} - \varphi_k)^2,$$

$$RBIAS(\hat{\beta}_{kr}) = \frac{BIAS(\hat{\beta}_{kr})}{|\beta_{kr}|},\ RBIAS(\hat{\varphi}_k) = \frac{BIAS(\hat{\varphi}_k)}{\varphi_k},$$

$$RMSE(\hat{\beta}_{kr}) = \frac{\sqrt{MSE(\hat{\beta}_{kr})}}{|\beta_{kr}|},\ RMSE(\hat{\varphi}_k) = \frac{\sqrt{MSE(\hat{\varphi}_k)}}{\varphi_k},$$

$$MEAN_{dk} = \frac{1}{I}\sum_{i=1}^{I} m_{dk}^{(i)},\ MSE_{dk} = \frac{1}{I}\sum_{i=1}^{I}(\hat{m}_{dk}^{(i)} - m_{dk}^{(i)})^2,\ RMSE_d = \frac{\sqrt{MSE_d}}{MEAN_d}.$$

Table 4.2.1 presents the RMSE-values for the parameters of models A and B under Model A. We observe that the RSME is higher in model B than the rest of the models, so that the use of a multivariate model produce a significant gain of efficiency. We also observe that as D increases from 50 to 300 we obtain a great reduction in RMSE (around 60% in model A). Table 4.2.2 presents the RBIAS-values for the parameters of models A and B under Model A. Relative bias is greater for model B, so the use of the multinomial model is again recommended.

	Model A					Model B				
D	50	100	150	200	300	50	100	150	200	300
$RMSE(\hat{\beta}_{01})$	0.73	0.53	0.42	0.35	0.28	0.90	0.62	0.50	0.47	0.39
$RMSE(\hat{\beta}_{02})$	0.85	0.60	0.50	0.41	0.32	1.05	0.81	0.70	0.63	0.59
$RMSE(\hat{\beta}_{11})$	0.78	0.56	0.45	0.37	0.30	1.37	1.24	1.18	1.14	1.11
$RMSE(\hat{\beta}_{12})$	0.99	0.70	0.59	0.48	0.38	1.14	0.87	0.75	0.67	0.60
$RMSE(\hat{\varphi}_1)$	0.27	0.18	0.14	0.13	0.10	0.62	0.52	0.48	0.48	0.46
$RMSE(\hat{\varphi}_2)$	0.26	0.18	0.14	0.12	0.10	0.48	0.30	0.25	0.22	0.20

Table 4.2.1: RMSE for the parameters of models A and B under model A.

	Model A					Model B				
D	50	100	150	200	300	50	100	150	200	300
$RBIAS(\hat{\beta}_{01})$	-0.02	-0.02	0.001	-0.01	-0.02	0.18	0.16	0.17	0.21	0.19
$RBIAS(\hat{\beta}_{02})$	-0.05	-0.05	-0.05	-0.04	-0.03	0.46	0.48	0.48	0.43	0.45
$RBIAS(\hat{\beta}_{11})$	-0.03	-0.02	-0.002	-0.01	-0.02	1.05	1.06	1.07	1.06	1.04
$RBIAS(\hat{\beta}_{12})$	-0.04	-0.05	-0.05	-0.02	-0.02	0.43	0.44	0.44	0.43	0.43
$RBIAS(\hat{\varphi}_1)$	0.01	-0.004	-0.01	-0.01	-0.02	0.49	0.46	0.44	0.45	0.43
$RBIAS(\hat{\varphi}_2)$	0.01	-0.002	-0.01	-0.01	-0.01	0.29	0.20	0.17	0.16	0.16

Table 4.2.2: RBIAS for the parameters of models A and B under model A.

In the previous simulations we do not present Model C, because this model only has one variance component, then this model is not comparable with the others. To investigate the behavior of the three candidate models, we have repeated the simulation experiment generating data from models A, B and C. Further, to analyze the importance of including a random effect per category (model A), we also include the estimators based on model C. Table 4.2.3 presents the RMSE values of the estimators of the cell totals based on models A, B and C (columns), when data is generated under models A, B and C (rows). The number of domains is $D = 100$, which is the value of D from the first simulation that is closest to the our real data case.

			Estimator		
Scenario		D	Model A	Model B	Model C
Model A	$RMSE_{d1}$	$d = 1$	0.09	0.15	0.38
		$d = \frac{1}{2}D$	0.11	0.16	0.48
		$d = D$	0.14	0.13	0.57
	$RMSE_{d2}$	$d = 1$	0.14	0.2	0.69
		$D/2$	0.12	0.17	0.56
		$d = D$	0.1	0.15	0.43
Model B	$RMSE_{d1}$	$d = 1$	0.15	0.08	0.5
		$d = \frac{1}{2}D$	0.17	0.1	0.66
		$d = D$	0.19	0.12	0.81
	$RMSE_{d2}$	$d = 1$	0.2	0.13	0.94
		$D/2$	0.17	0.1	0.78
		$d = D$	0.16	0.09	0.61
Model C	$RMSE_{d1}$	$d = 1$	0.09	0.14	0.07
		$d = \frac{1}{2}D$	0.11	0.16	0.07
		$d = D$	0.14	0.18	0.09
	$RMSE_{d2}$	$d = 1$	0.15	0.2	0.1
		$d = \frac{1}{2}D$	0.12	0.17	0.08
		$d = D$	0.1	0.16	0.07

Table 4.2.3: $RMSE_{dk}$ for $D = 100$ and models A-C.

If data is generated under Model A, Table 4.2.3 shows that the loss of efficiency when passing from model A to model B is not very high; however the loss is dramatic when passing from model A to model C. This is somehow expected because substituting u_{dk} by u_d in (4.1.2) is a rather extreme simplification. If data is generated under model B, we observe that the loss of efficiency when passing from model B to model A is negligible; however there is a significant loss when passing from model B to model C. If data is generated under model C, we observe that the loss of efficiency when passing from model C to model A is also negligible; however the loss is larger when passing from model C to model B. The final conclusion is that estimators based on model A behaves much better than the corresponding ones based on models B and C when data is generated under Model A. Furthermore, estimators based on model A still behaves well when data is generated under model B or C. This last conclusion does not hold for estimators based on models B or C.

4.2.3 Simulation experiment 2

The second simulation experiment is designed to study the behavior of the three mean square error estimators (analytic and bootstrap). In this case we consider $D = 100$, which is the closest to our real data case. The simulation has the following steps.

1. Repeat $I = 500$ times ($i = 1, \ldots, 500$)

   1.1. Generate a sample $(y_{dk}^{(i)}, x_{dk}^{(i)})$, $d = 1, \ldots, D$, $k = 1, 2$ in the same way as in simulation 1.

   1.2. For $d = 1, \ldots, D$, $k = 1, 2$, $r = 0, 1$, calculate $\hat{\varphi}_1^{(i)}$, $\hat{\varphi}_2^{(i)}$, $\hat{\beta}_{kr}^{(i)}$, $\hat{p}_{dk}$ and
$$mse_{dk}^{(i)} = \mathcal{G}_{1dk}^{(i)}(\hat{\varphi}_1^{(i)}, \hat{\varphi}_2^{(i)}) + \mathcal{G}_{2dk}^{(i)}(\hat{\varphi}_1^{(i)}, \hat{\varphi}_2^{(i)}) + 2\mathcal{G}_{3dk}^{(i)}(\hat{\varphi}_1^{(i)}, \hat{\varphi}_2^{(i)}).$$

   1.3. Repeat $B = 500$ times ($b = 1, \ldots, B$)

   1.3.1. For $d = 1, \ldots, D$, $k = 1, 2$, generate
$$\begin{aligned} u_{dk}^{*(ib)} &\sim N(0, \varphi_k^{(i)}), \ \ \eta_{dk}^{*(ib)} = \hat{\boldsymbol{\beta}}_k^{(i)} x_{dk}^{(i)} + u_{dk}^{*(ib)}, \\ \boldsymbol{y}_d^{*(ib)} &= (y_{d1}^{*(ib)}, y_{d2}^{*(ib)})' \sim \mathrm{M}(\nu_d, p_{d1}^{*(ib)}, p_{d2}^{*(ib)}), \ p_{dk}^{*(ib)} = \frac{\exp\{\eta_{dk}^{*(ib)}\}}{1 + \exp\{\eta_{d1}^{*(ib)}\} + \exp\{\eta_{d2}^{*(ib)}\}}. \end{aligned}$$

   1.3.2. Calculate $\hat{\varphi}_k^{*(ib)}$, $\hat{\beta}_{kr}^{*(ib)}$, $\hat{p}_{dk}^{*(ib)}$, $\tilde{p}_{dk}^{*(ib)}$, $\hat{m}_{dk}^{*(ib)}$, $k = 1, 2$, $r = 0, 1$, $d = 1, \ldots, D$.

   1.4 For $d = 1, \ldots, D$, $k = 1, 2$, calculate
$$\begin{aligned} mse_{dk}^{*1(i)} &= \frac{1}{B}\sum_{b=1}^{B}(\hat{m}_{dk}^{*(ib)} - m_{dk}^{*(ib)})^2, \\ mse_{dk}^{*2(i)} &= \frac{1}{B}\sum_{b=1}^{B}\left\{\mathcal{G}_{1dk}^{(*ib)}(\hat{\varphi}_1^{*(ib)}, \hat{\varphi}_2^{*(ib)}) + \mathcal{G}_{2dk}^{*(ib)}(\hat{\varphi}_1^{*(ib)}, \hat{\varphi}_2^{*(ib)}) + 2\mathcal{G}_{3dk}^{*(ib)}(\hat{\varphi}_1^{*(ib)}, \hat{\varphi}_2^{*(ib)})\right\}. \end{aligned}$$

2. Calculate
$$B_{dk}^0 = \frac{1}{I}\sum_{i=1}^{I}(mse_{dk}^{(i)} - MSE_{dk}), \ B_{dk}^{\ell} = \frac{1}{I}\sum_{i=1}^{I}(mse_{dk}^{*\ell(i)} - MSE_{dk}), \quad k = 1, 2, \ \ell = 1, 2,$$
$$E_{dk}^0 = \frac{1}{I}\sum_{i=1}^{I}(mse_{dk}^{(i)} - MSE_{dk})^2, \ E_{dk}^{\ell} = \frac{1}{I}\sum_{i=1}^{I}(mse_{dk}^{*\ell(i)} - MSE_{dk})^2, \quad k = 1, 2, \ \ell = 1, 2,$$
$$RB_{dk}^{\ell} = \frac{B_{dk}^{*\ell}}{MSE_{dk}}, \ RE_{dk}^{\ell} = \frac{\sqrt{E_{dk}^{*\ell}}}{MSE_{dk}}, \quad k = 1, 2, \ \ell = 0, 1, 2,$$
where the MSE_{dk}'s were obtained in the simulation 1.

Table 4.2.4 shows that the bootstrap-based estimators give better results than his analytic competitor.

	d	mse	mse^{*1}	mse^{*2}		d	mse	mse^{*1}	mse^{*2}
RB_{d1}^{ℓ}	1	0.13	-0.11	-0.04	RE_{d1}^{ℓ}	1	0.33	0.14	0.07
	50	0.07	-0.07	-0.01		50	0.35	0.10	0.05
	100	0.12	-0.04	0.05		100	0.49	0.11	0.09
RB_{d2}^{ℓ}	1	0.08	0.10	0.18	RE_{d2}^{ℓ}	1	0.67	0.15	0.21
	50	0.05	-0.03	0.04		50	0.52	0.09	0.07
	100	0.06	-0.12	-0.04		100	0.42	0.14	0.08

Table 4.2.4: RB_{dk}^{ℓ}, RE_{dk}^{ℓ} for $D = 100$, $k = 1, 2$, $\ell = 1, 2$, under model A.

Figure 4.2.1 presents the box-plots of the simulated $mse_{dk}^{(i)}, mse_{dk}^{*\ell(i)}$, $k=1,2$, $\ell=1,2$, $i=1,\dots,500$, for the three estimators in three particular domains $d=1$, 50, 100. The true MSE is plotted in a horizontal line. We observe that mse has the largest variability and the best estimators are mse^{*1} and mse^{*2}.

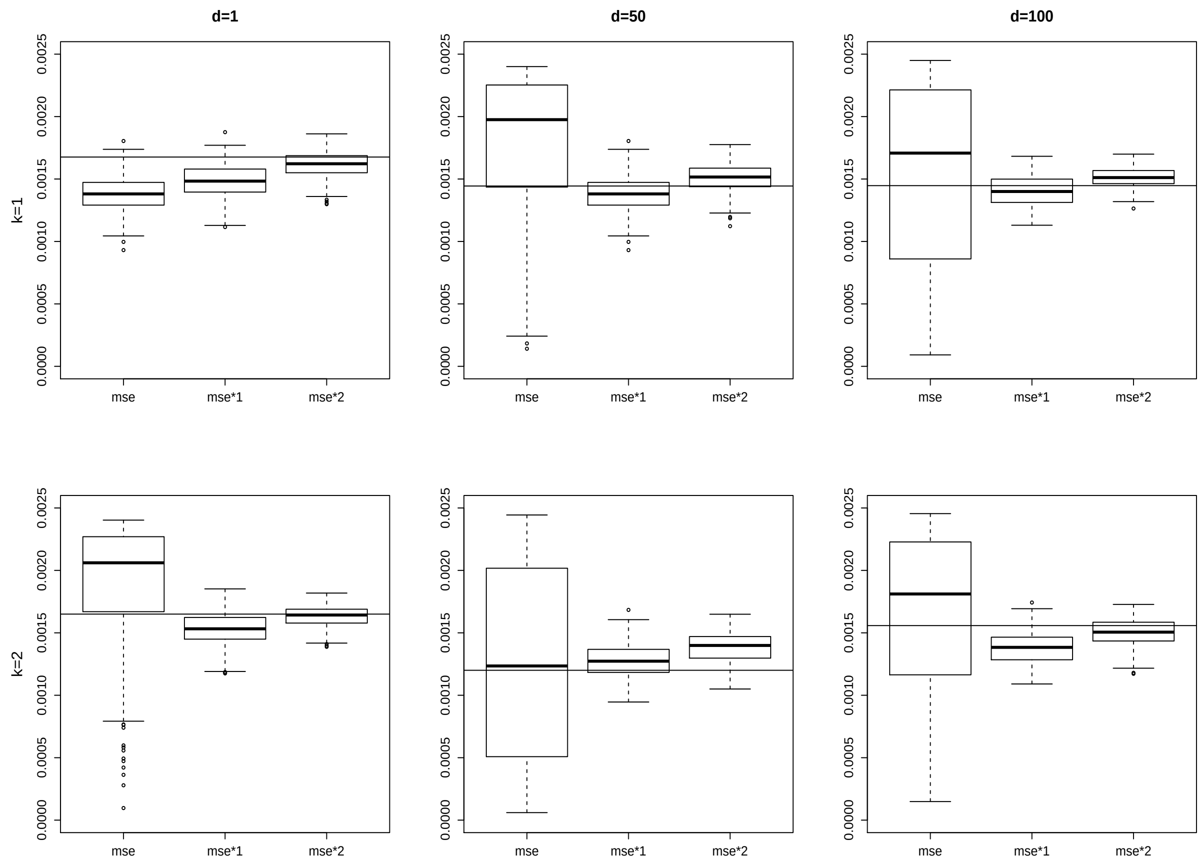


Figure 4.2.1: Boxplots of Simulation 2 for $D=100$ and $d=1,50,100$.

4.3 Application to real data

4.3.1 Data Description

We deal with data from the SLFS of Galicia in the fourth quarter of 2008. In this section the domains of interest are the counties crossed with sex. As there are 48 counties in the SLFS of Galicia for the fourth quarter of 2008, we have $D=96$ domains P_d partitioned in the subsets P_{dk}, $k=1,2,3$, of employed ($k=1$), unemployed ($k=2$) and inactive ($k=3$) people. Our target population parameters are the totals of employed and unemployed people and the unemployment rate, this is to say

$$Y_{dk}=\sum_{j\in P_d} y_{dkj}, \quad R_d=\frac{Y_{d2}}{Y_{d_1}+Y_{d2}}, \quad d=1,\dots,96, \quad k=1,2,$$

where $y_{dkj} = 1$ if the individual j of domain d is in P_{dk} and $y_{dkj} = 0$ otherwise.

The SLFS does not produce official estimates at the domain level, but the analogous direct estimators of the total Y_{dk}, the mean $\bar{Y}_{dk} = Y_{dk}/N_d$, the size N_d and the rate R_d are

$$\hat{Y}_{dk}^{dir} = \sum_{j \in S_d} w_{dj}\, y_{dkj},\ \hat{\bar{Y}}_{dk}^{dir} = \hat{Y}_{dk}^{dir}/\hat{N}_d^{dir},\ \hat{N}_d^{dir} = \sum_{j \in S_d} w_{dj},\ \hat{R}_d^{dir} = \frac{\hat{Y}_{d2}^{dir}}{\hat{Y}_{d1}^{dir} + \hat{Y}_{d2}^{dir}},\ k = 1, 2, \tag{4.3.1}$$

where S_d is the domain sample and the w_{dj}'s are the official calibrated sampling weights which take into account for non response. The design-based covariance $\text{cov}_\pi(\hat{Y}_{dk_1}^{dir}, \hat{Y}_{dk_2}^{dir})$, $k_1, k_2 = 1, 2$, can be estimated by

$$\hat{\text{cov}}_\pi(\hat{Y}_{dk_1}^{dir}, \hat{Y}_{dk_2}^{dir}) = \sum_{j \in S_d} w_{dj}(w_{dj} - 1)(y_{dk_1j} - \hat{\bar{Y}}_{dk_1}^{dir})(y_{dk_2j} - \hat{\bar{Y}}_{dk_2}^{dir}), \tag{4.3.2}$$

where the case $k_1 = k_2 = k$ denotes estimated variance, i.e. $\hat{V}_\pi(\hat{Y}_{dk}^{dir}) = \hat{\text{cov}}_\pi(\hat{Y}_{dk}^{dir}, \hat{Y}_{dk}^{dir})$. The last formulas are obtained from Särndal et al. (1992), pp. 43, 185 and 391, with the simplifications $w_{dj} = 1/\pi_{dj}$, $\pi_{dj,dj} = \pi_{dj}$ and $\pi_{di,dj} = \pi_{di}\pi_{dj}$, $i \neq j$ in the second order inclusion probabilities. The design-based variance of $\hat{R}_d^{dir}$ can be approximated by Taylor linearization, i.e.

$$\hat{V}_\pi(\hat{R}_d^{dir}) = \frac{(\hat{Y}_{d1}^{dir})^2}{(\hat{Y}_{d1}^{dir} + \hat{Y}_{d2}^{dir})^4}\hat{V}_\pi(\hat{Y}_{d2}^{dir}) + \frac{\hat{Y}_{d2}^2}{(\hat{Y}_{d2}^{dir} + \hat{Y}_{d1}^{dir})^4}\hat{V}_\pi(\hat{Y}_{d1}^{dir}) - \frac{2\hat{Y}_{d1}^{dir}\hat{Y}_{d2}^{dir}}{(\hat{Y}_{d1}^{dir} + \hat{Y}_{d2}^{dir})^4}\hat{\text{cov}}_\pi(\hat{Y}_{d1}^{dir}, \hat{Y}_{d2}^{dir}).$$

In the fourth quarter of 2008 the distribution of the sample sizes per domains in the SLFS of Galicia has the quantiles $q_{min} = 8$, $q_1 = 24$, $q_2 = 47$, $q_3 = 90$ and $q_{max} = 734$. This means that direct estimates in (4.3.1) and (4.3.2) are not reliable. Therefore, small area estimation methods are needed.

Small area estimators based on unit level models (models stated for the individual units) are likely to achieve high precision when the model is correctly specified. However, the estimators derived from these models usually need auxiliary data for each unit of the sample along with aggregated data for each domain, something not always available. Area level models need only the domain totals of the auxiliary variables, information that can be usually found in administrative registers. This is why we consider an area-level model using as auxiliary variables the domain proportions of individuals within the categories of the following grouping variables.

- SEX: This variable is coded 1 for men and 2 for women.
- AGE: This variable is categorized in 3 groups with codes 1 for 16-24, 2 for 25-54 and 3 for $\geq$55
- SS: This variable indicates if an individual is registered or not in the social security system.
- REG: This variable indicates if an individual is registered or not as unemployed in the administrative register of employment claimants.

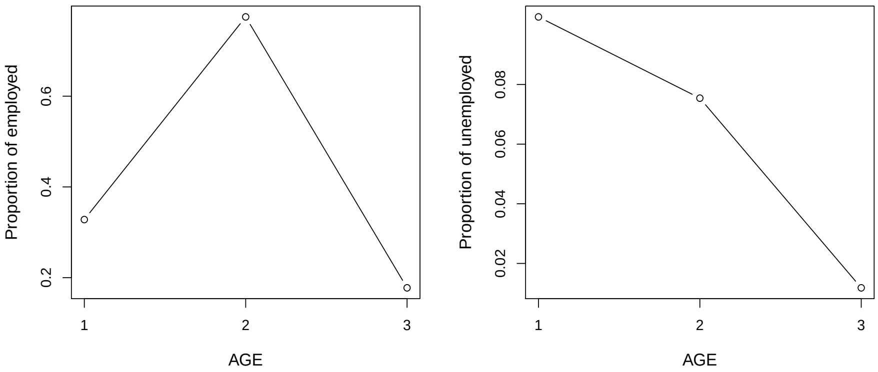


Figure 4.3.1: Proportions of employed (left) and unemployed (right) people versus AGE categories.

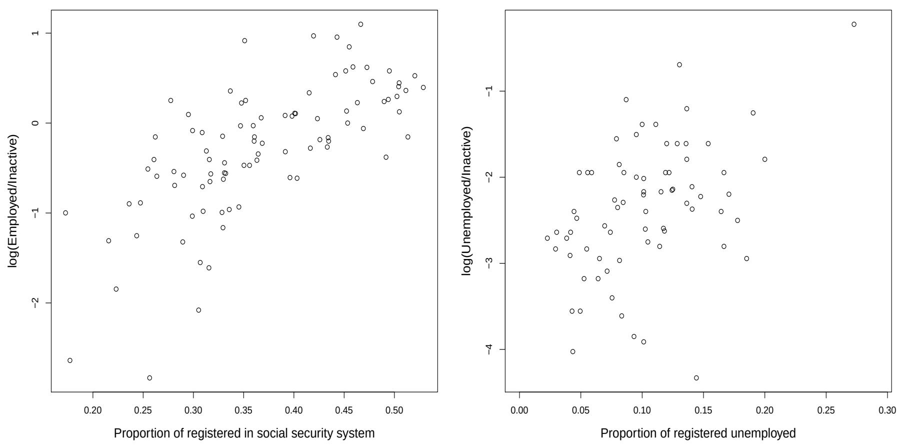


Figure 4.3.2: Log-rates of employed and unemployed over inactive people versus proportions of people in the social security system (left) and registered as unemployed (right) respectively.

We first analyze the potential predictive power of these auxiliary variables through an exploratory data analysis. Figure 4.3.1 plots the SLFS estimated proportions of employed (left) and unemployed (right) for each AGE category. Observe that both estimated proportions vary considerably across the AGE categories. Moreover, comparing the left and right plots, we can see that the lines are far from being parallel. This suggests that the three estimated proportions vary differently across categories. Thus, these figures suggest that the variable AGE can be a good predictor of the probabilities of being employed and unemployed.

Figure 4.3.2 shows the scatterplots of the log-rates of employed over inactive people against the proportions of people in social security system (left) and the log-rates of unemployed over inactive people against the proportions of people registered as unemployed (right). We observe that, despite the large variability observed in both plots, the log-rates of the two considered proportions seem to increase linearly with the proportions of people in the social security system and registered as unemployed respectively. Then, the proportions of people in the social security system and registered as unemployed could probably be good covariates for modeling the two probabilities. Indeed, after fitting the model described in Section 4.1, tests of significance for the regression parameters and diagnosis of residuals confirmed the explanatory power of the auxiliary variables that were selected for each model.

4.3.2 Model estimation

We are interested in estimating the totals of employed and unemployed people, and the unemployment rates per sex in the counties of Galicia. We consider the multinomial mixed model (4.1.1)-(4.1.2) with $q = 3$ categories (employed, unemployed and inactive people) and we choose inactive people as reference (third) category. The multinomial size is $\nu_d = n_d$, where n_d is the size of the domain sample S_d. The target variable is $\boldsymbol{y}_d = (y_{d1}, y_{d2})'$, where y_{dk} is the sample total

$$y_{dk} = \sum_{j \in S_d} y_{dkj},$$

$y_{dkj} = 1$ if individual j is in category k $(k = 1, 2)$ and $y_{dkj} = 0$ otherwise.

The explanatory variables are the domain means of the indicators of the categories of SEX, AGE, SS and REG. Their values are taken from the SLFS and from administrative registers. The model is firstly fitted to the complete data set. An analysis of residuals is then carried out and two counties are marked as outliers. These two counties correspond to the largest cities in Galicia (A Coruña and Vigo) where the relationships between the auxiliary variables SS and REG with the employment and unemployment status are typically weaker than in less populated counties. The model is finally fitted to the reduced (excluding the two marked counties) data set. The rest of the statistical analysis is carried out under the model fitted to the reduced data set. On the other hand, the sample sizes of the excluded counties are large enough to produce reliable direct estimates. Therefore, no model estimates are given for the excluded counties. Instead, direct estimates are used for them.

Concerning Figure 4.3.2, we would like to say that by taking out few dots from the right and left plots we would obtain much more linear clouds. The linearity of these clouds of dots can also be measure via the estimated Pearson correlations of variables. They are 0.71 (left) and 0.41 (right) respectively. Further, the corresponding 95% confidence intervals are (0.58, 0.79) and (0.21,0.59) respectively. Therefore, the proportions of people in the social security system and registered as unemployed could probably be good covariates for modeling the two probabilities. Indeed, after fitting the multinomial mixed model, tests of significance for the regression parameters and diagnosis of residuals confirmed the explanatory power of the auxiliary variables that were selected for each model.

For each category and data set, Table 4.3.1 presents the estimates of the regression parameters and their standard deviations. It also presents the p-values for testing the hypothesis $H_0 : \beta_{kr} = 0$. The estimates of the model variances and their standard deviations (in parentheses) are $\hat{\varphi}_1 = 0.17$ (0.03) and $\hat{\varphi}_2 = 0.50$ (0.13) for the model fitted to the complete data set and they are $\hat{\varphi}_1 = 0.17$ (0.04) and $\hat{\varphi}_2 = 0.59$ (0.10) for the model fitted to the reduced data set.

Employed people				Unemployed people			
Variable	$\hat{\beta}_{kr}$	Std.Dev.	p-value	Variable	$\hat{\beta}_{kr}$	Std.Dev.	p-value
Constant	-1.979	0.196	0.000	CONSTANT	-5.450	0.408	0.000
SS	1.146	0.669	0.087	REG	13.105	3.830	0.001
AGE=1	2.750	0.369	0.000	AGE=1	3.315	1.601	0.038
AGE=2	-0.317	0.068	0.000	AGE=2	4.356	0.754	0.000
SEX=1	1.684	0.510	0.001	SEX=1	-0.394	0.138	0.004
Constant	-1.972	0.215	0.000	CONSTANT	-5.254	0.429	0.000
SS	1.217	0.679	0.073	REG	11.907	4.102	0.004
AGE=1	2.741	0.378	0.000	AGE=1	3.959	1.640	0.016
AGE=2	-0.356	0.073	0.000	AGE=2	3.832	0.806	0.000
SEX=1	1.696	0.537	0.002	SEX=1	-0.370	0.151	0.014

Table 4.3.1: Estimates $\hat{\beta}_{kr}$ for the full (top) and reduced (bottom) data set.

Figure 4.3.3 presents the boxplots of the domain standardized residuals of models fitted to the full (left) and to the reduced (right) data set. The four dots outside the interval (-3,3) correspond to the men and women counts in marked counties (A Coruña and Vigo). The remaining model-based statistical analysis is carried out for the reduced data set.

Figure 4.3.4 plots the domain standardized residuals of employment (left) and unemployment (right) categories versus the proportions of people registered in the social security system (SS) and registered as unemployed (REG). The residuals are randomly distributed above and below zero and no rare pattern is observed. Therefore no diagnostics problems are found for the two main explanatory variables: SS and REG.

4.3.3 Model diagnostics

Figure 4.3.5 plots the domain residuals

$$r_{dk} = \frac{y_{dk} - \hat{y}_{dk}}{\hat{y}_{dk}}, \quad d = 1, \ldots, 96, \quad k = 1, 2,$$

versus the predicted sample totals of employed and unemployed people. We observe that the relative difference of the predicted values ($\hat{y}_{dk} = n_d \hat{p}_{dk}$) with respect to the observed ones (y_{dk}) is below 2% in the case of employed people and below 4% in the case of unemployed people.

Figure 4.3.6 plots the direct ($\hat{Y}_{dk}^{dir}$) versus the model-based estimates ($\hat{N}_d^{dir} \hat{p}_{dk}$) of the totals of employed and unemployed people. We observe that both type of estimates behave quite similarly for employed people. This is because there are enough observations for these estimates. However, their behavior presents slightly greater differences for unemployed people, which are due to the lower number of observations. We also observe that the model-based estimates are in general smother than the direct ones.

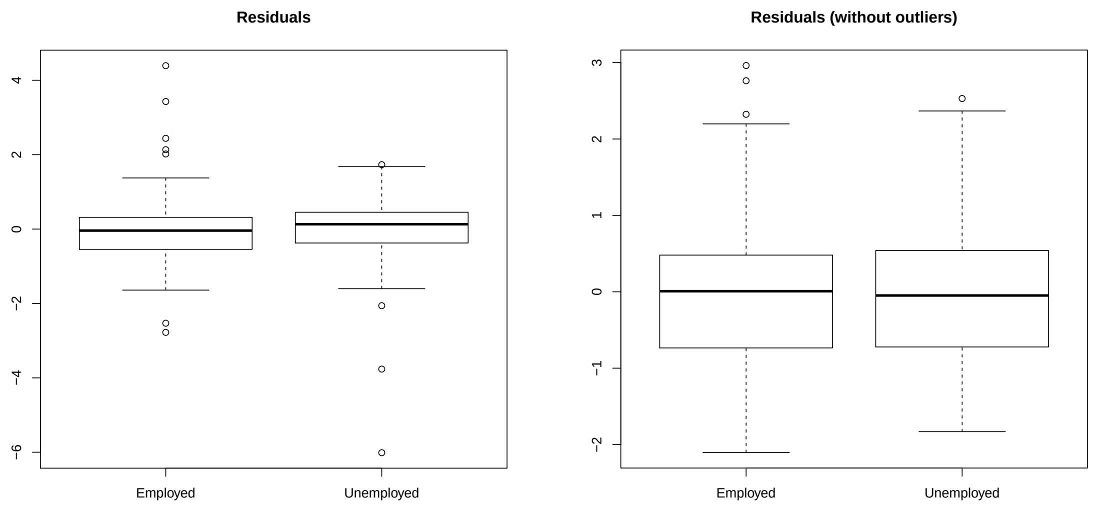


Figure 4.3.3: Boxplots of standardized residuals of models fitted to the full (left) and reduced (right) data set.

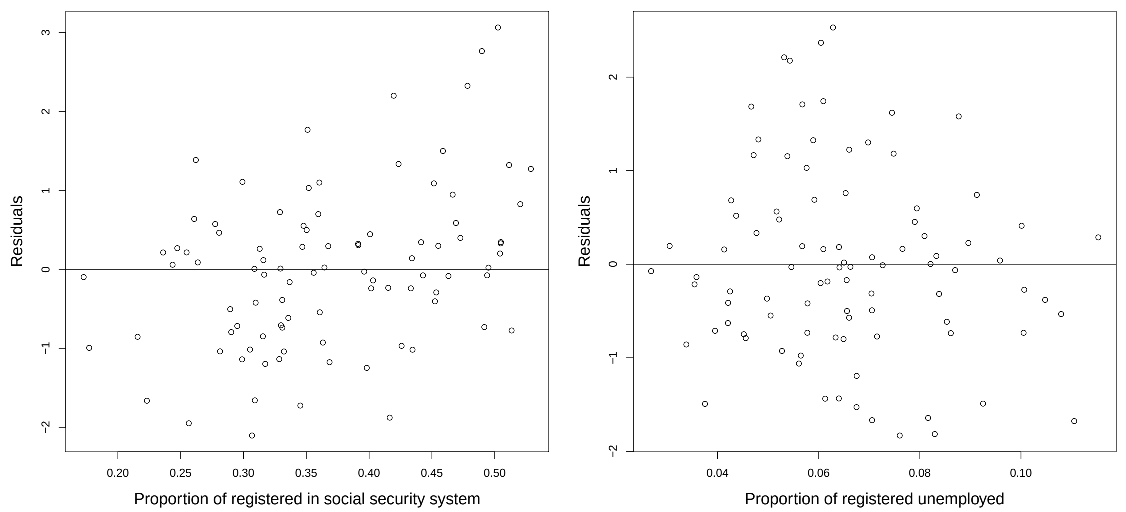


Figure 4.3.4: Domain standardized residuals of employment (left) and unemployment (right) categories versus proportions of people registered in the social security system (SS) and registered as unemployed (REG).

4.3.4 Small area estimates and RMSE

Figures 4.3.7 and 4.3.8 plot the estimated employment totals and unemployment rates respectively for men (left) and women (right), with the counties sorted by sample size. We observe that both estimates (direct and model-based) tend to be closer as soon as the sample size increases.

Tables 4.3.2 and 4.3.3 present some condensed numerical results for men and women

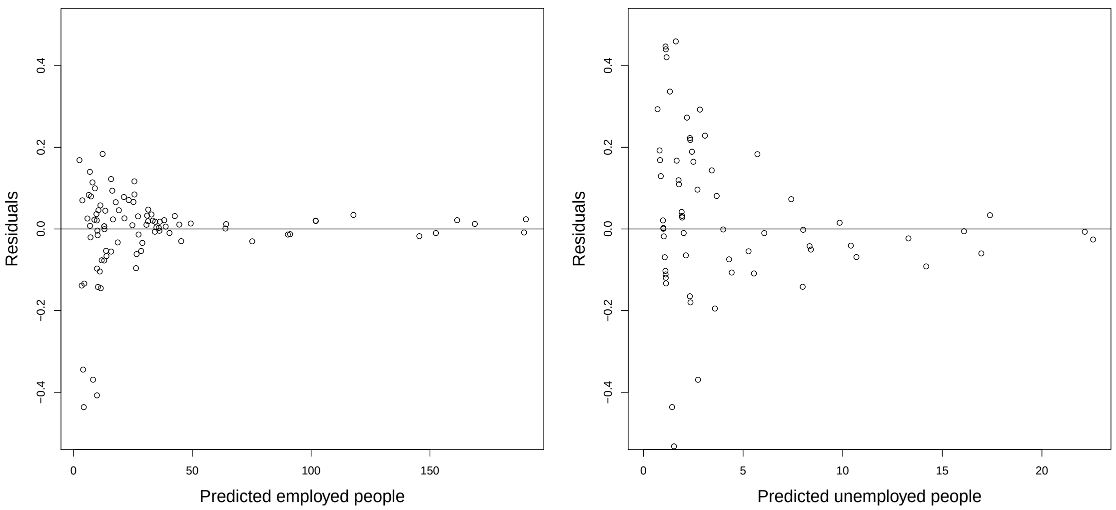


Figure 4.3.5: Domain residuals versus predicted values of employed people (left) and unemployed people (right).

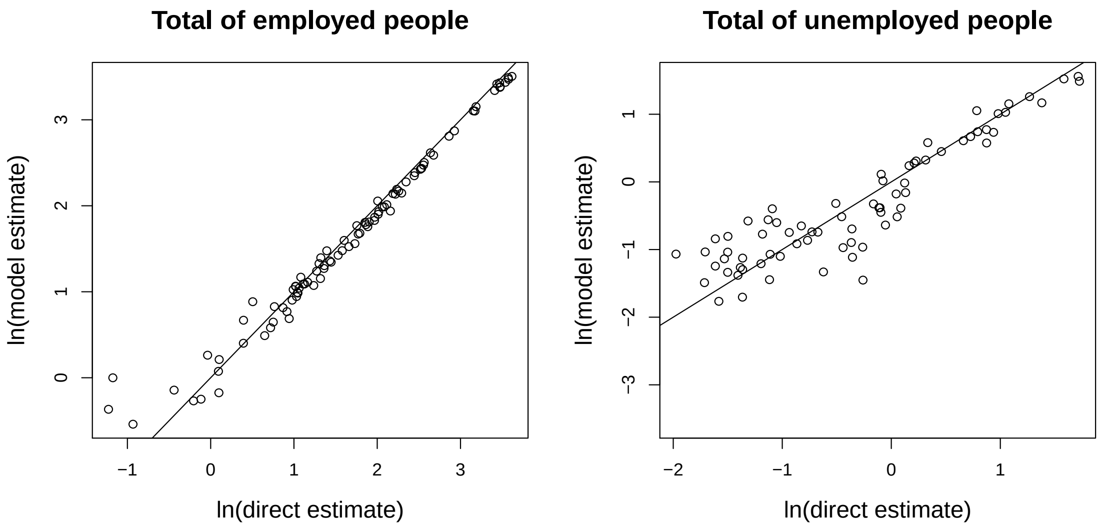


Figure 4.3.6: Direct versus model-based estimates of the total of employed people (left) and unemployed people (right).

respectively. The tables has been constructed in two steps. We sort the domains by province and after that, in each province, we sort the domains by sample size, starting by the domain with smallest sample size. We present the results of the direct and the model-based estimates (labeled by "dir" and "mod" respectively) and the corresponding RMSE estimates for five domains in each province. The chosen domains correspond to the quintiles. The provinces are labeled by p and the sample sizes by n. Table 4.3.2 and 4.3.3 are partitioned in three vertical sections dealing with the estimation of totals of employed and unemployed people and with unemployment rates. The RMSEs of the model-based

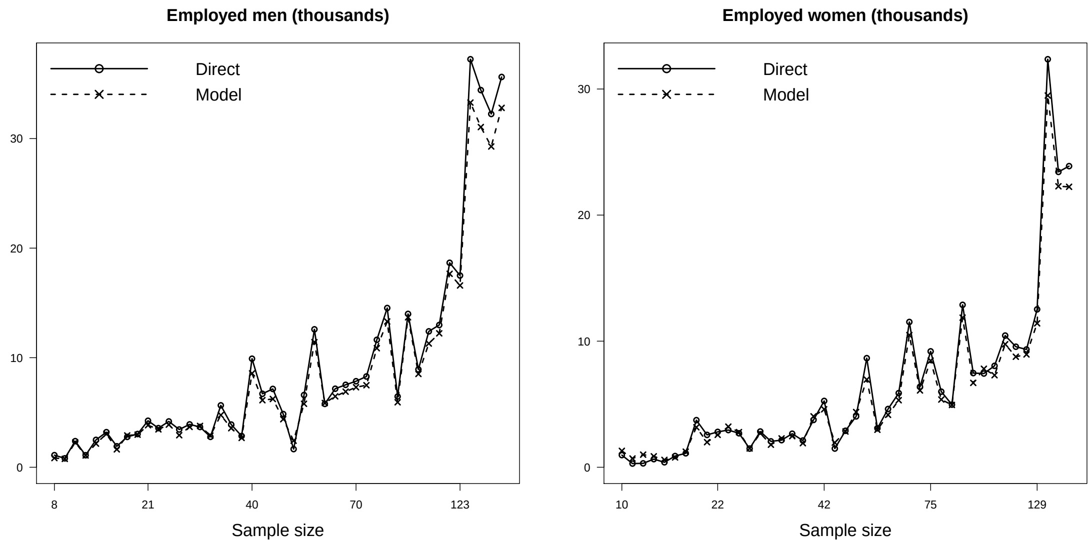


Figure 4.3.7: Men (left) and women (right) employment totals.

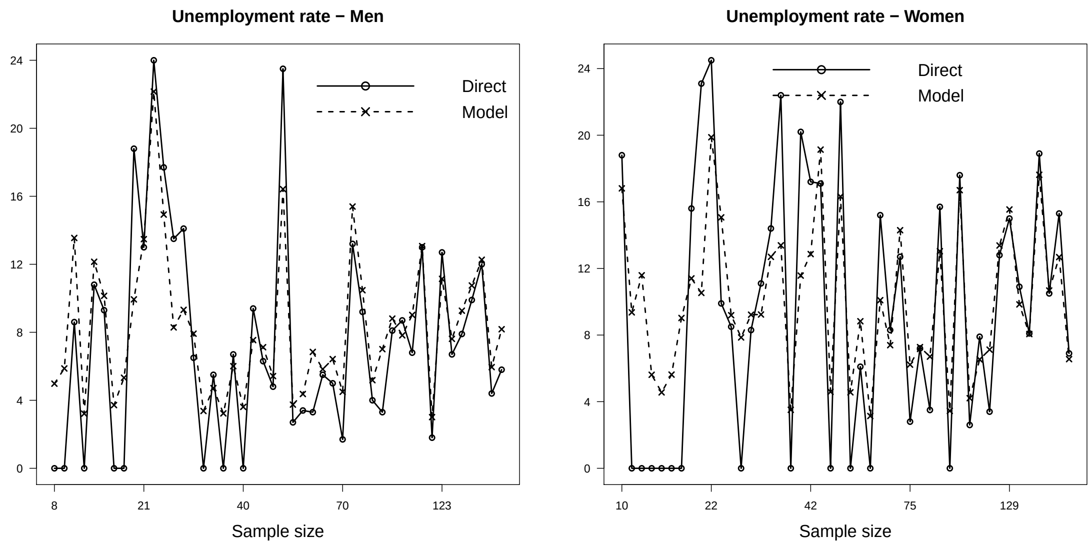


Figure 4.3.8: Men (left) and women (right) unemployment rates.

estimators are calculated by using the parametric bootstrap method 1 (mse^{*1}). The RMSEs of the corresponding direct estimates are estimated by using (4.3.2) and they are much greater than their counterparts. By observing the columns of RMSEs we conclude that model-based estimators are preferred to the direct estimates.

In domains d with $n_d = 0$, direct estimates cannot be calculated. In those cases, model-based estimates are calculated by using the synthetic part of the linear predictor, i.e. $\hat{\eta}_{dk}^{synth} = \boldsymbol{x}_{dk}\hat{\boldsymbol{\beta}}_k$. Tables 4.3.2 and 4.3.3 present blank spaces in domains where $y_{dk} = 0, k = 1, 2$.

		Employed people				Unemployed people				Unemployment rates			
		estimate		rmse		estimate		rmse		estimate		rmse	
p	n	dir	mod	dir	mod	dir	mod	dir	mod	dir	mod	dir	mod
1	13	2383	2259	24.2	17.0	223	354	100.3	43.2	8.6	13.5	239.0	38.6
1	21	4246	3836	18.0	16.2	634	598	68.1	42.3	13.0	13.5	108.6	41.3
1	27	3676	3771	20.8	15.6	256	324	98.5	31.4	6.5	7.9	281.1	32.2
1	79	13998	13687	9.9	8.7	1233	1321	48.8	20.6	8.1	8.8	104.6	19.6
1	213	34425	31052	6.7	6.4	2939	3172	31.5	14.4	7.9	9.3	72.9	14.4
2	11	814	764	47.1	28.8		48		54.3		5.9		46.2
2	32	2776	2890	21.3	15.7		101		43.2		3.4		40.5
2	41	6707	6121	11.8	10.9	697	499	72.5	32.3	9.4	7.5	106.4	32.4
2	70	7525	6901	11.3	10.5	392	474	69.7	23.8	5.0	6.4	205.9	22.9
2	300	30266	28235	4.8	4.6	1397	1787	31.6	13.7	4.4	6.0	99.4	12.8
3	8	1104	840	43.9	33.4		44		67.1		5.0		62.7
3	33	5649	4759	15.0	16.0	327	236	98.1	37.4	5.5	4.7	245.8	38.1
3	42	1657	2422	37.9	17.6	510	476	69.7	31.4	23.5	16.4	97.1	30.2
3	70	7852	7304	12.7	11.6	139	344	100.3	26.8	1.7	4.5	704.7	30.0
3	356	35758	32181	4.7	4.4	2190	2867	25.1	11.7	5.8	8.2	71.9	11.6
4	17	2500	2158	22.6	18.1	303	298	96.3	44.6	10.8	12.1	170.0	42.2
4	41	7150	6231	11.4	13.1	483	478	69.3	33.0	6.3	7.1	159.0	29.9
4	90	12984	12223	9.1	8.2	1937	1838	34.8	19.7	13.0	13.1	54.8	19.6
4	123	17499	16593	8.3	7.5	2555	2081	34.4	18.5	12.7	11.1	51.7	17.8
4	291	35624	32801	5.0	4.6	4871	4585	21.7	10.9	12.0	12.3	33.6	11.1

Table 4.3.2: Estimated men totals and rates with their estimated RMSE's.

The Spanish Statistics Institute (INE) publishes LFS estimates of employed and unemployed totals at province level. In the case of extending these publications to the more disaggregated levels, Ugarte et al. (2009b) point out that Statistical Offices might be interested in publishing data with the property that the sum of the estimated totals in all the domains within a province coincide with the official province total estimate. In order to fullfil this consistency criterion, we propose a modification of all the considered small area estimators. Let $\widehat{Y}_p^{dir}$ be the SLFS estimator of the total Y_p in province p. Assume that province p is partitioned in D_p domains. Let $\widehat{Y}_1, \dots, \widehat{Y}_{D_p}$ be some given estimators of totals $Y_1, \dots, Y_{D_p}$. In general, the consistency property

$$\widehat{Y}_p^{dir} = \sum_{d=1}^{D_p} \widehat{Y}_d$$

does not hold. In such cases, $\widehat{Y}_1, \dots, \widehat{Y}_{D_p}$ can be transformed into consistent estimators by

$$\widehat{Y}_d^c = \lambda_p \widehat{Y}_d, \quad \lambda_p = \frac{\widehat{Y}_p^{dir}}{\sum_{d=1}^{D_p} \widehat{Y}_d}.$$

Table 4.3.4 presents the direct and model-based estimators for employed and unemployed people at the province level for men (top) and women (bottom) in the SLFS. Table 4.3.4 also gives the consistency factors λ_p. We observe that the deviations from the SLFS estimation at province level are at most of 10% for employed people. This is something expected as the amount of people in the category of employed is large. However, the deviations from the SLFS estimation at the province level goes up over 20% for unemployed people in the two provinces with lower sample sizes.

		Employed people				Unemployed people				Unemployment rates			
		estimate		rmse		estimate		rmse		estimate		rmse	
p	n	dir	mod	dir	mod	dir	mod	dir	mod	dir	mod	dir	mod
1	10	964	1300	47.5	27.2	223	263	99.4	48.2	18.8	16.8	171.0	41.1
1	22	2804	2571	25.0	19.8	909	638	53.6	40.0	24.5	19.9	64.4	35.6
1	48	8653	6954	13.7	15.5		332		30.5		4.6		28.0
1	90	7435	7810	17.5	12.2	1583	1566	43.8	22.0	17.6	16.7	69.6	22.2
1	237	23884	22243	9.7	7.9	5553	4759	24.5	12.9	18.9	17.6	35.1	13.3
2	18	645	866	52.7	27.7		52		41.5		5.6		39.6
2	42	2120	1910	27.4	20.3		69		33.6		3.5		30.5
2	47	2894	2836	23.2	17.1		136		37.3		4.6		38.1
2	78	5996	5372	14.7	14.4	464	422	69.7	24.3	7.2	7.3	176.7	23.4
2	350	24190	23401	6.4	6.0	2851	2797	23.3	11.0	10.5	10.7	49.3	11.4
3	15	293	693	96.9	32.6		72		51.0		9.4		42.2
3	41	2661	2466	25.8	23.1	769	381	55.8	32.0	22.4	13.4	72.7	30.9
3	46	1483	1952	41.1	20.2	307	462	72.6	28.7	17.1	19.1	166.1	24.7
3	79	4967	4941	16.1	11.6	181	355	99.1	25.6	3.5	6.7	425.1	24.6
3	375	32114	30875	5.2	5.5	2394	2166	29.6	11.7	6.9	6.6	65.4	12.5
4	22	2951	3221	22.0	18.2	323	571	96.9	28.5	9.9	15.1	184.0	27.4
4	47	4033	4379	22.1	16.6	1140	854	48.0	29.2	22.0	16.3	63.7	26.6
4	97	10445	9754	12.1	11.9	893	682	56.9	27.0	7.9	6.5	131.1	23.6
4	129	12524	11414	11.8	10.5	2208	2100	34.7	17.7	15.0	15.5	58.2	17.2
4	362	31100	30473	6.5	5.7	5615	4426	22.2	9.8	15.3	12.7	31.5	9.4

Table 4.3.3: Estimated women totals and rates with their estimated RMSE's.

			Employed people			Unemployed people		
Sex	Province	n	dir	mod	λ_p	dir	mod	λ_p
Men	1	1503	274405	262422	1.05	25593	26581	0.96
	2	756	80538	76662	1.05	4142	4801	0.86
	3	700	72563	69664	1.04	4370	5635	0.78
	4	1641	231585	220953	1.05	29228	28493	1.03
	Total	4600	659091	629701	1.05	63333	65510	0.97
Women	1	1700	233245	225775	1.03	27379	25374	1.08
	2	860	61575	60658	1.02	5896	5560	1.06
	3	757	58816	59231	0.99	4942	4931	1.00
	4	1834	182339	178435	1.02	27442	25808	1.06
	Total	5151	535975	524100	1.02	65658	61674	1.06

Table 4.3.4: Estimated men and women province totals.

Figures 4.3.9 plots the RMSEs for employed and unemployed men under the proposed multinomial mixed model (Model A), under the two separate independent logit mixed models (Model B) and under model proposed by Molina et al. (2007) (Model C). The counties are sorted by sample size. We observe that the Model A estimates are generally better (lower RMSEs) for unemployed men totals than the corresponding ones based on the models B and C. This is the case where the direct estimates are the worst. For employed men totals all the models give good results. This is because there are more employed than unemployed men and therefore it is easier to obtain estimates with RMSEs bellow 20%. We also observe that the employment category dominates the unique random effect in Model C, so that model produce good estimates of employment totals and bad estimates of unemployment totals. These results are also coherent with the simulation experiment. As similar conclusion are obtained here for women, we skip the corresponding figures.

The three candidate models (models A, B and C) can also be compared with the usual measures for model comparison such as the loglikelihood or the AIC. The resulting values of these measures are listed in Table 4.3.5. We can see that model A has the best Loglikelihood and AIC. Therefore we recommend model A.

	Model A	Model B	Model C
Loglikelihood	-469.4	-490.35	-493.65
AIC	962.81	1004.7	1011.3

Table 4.3.5: Measures for model comparison.

The key point when comparing the multinomial models A and C is that model C has only one common random effect for the two modeled multinomial categories. In the case of labour data, it is hard to find examples where the categories of employed and unemployed people could be modeled with the same random effect. As the range of values of totals of employed and unemployed people and their variabilities are in general quite different, it is more intuitive to use one random effect for each category.

Further simulations analysis could be done by bootstrapping the sample registers to create population data files, or by using census data files, and later by carrying out design-based simulation studies. As can be seen in Salvati et al. (2010), this approach is quite

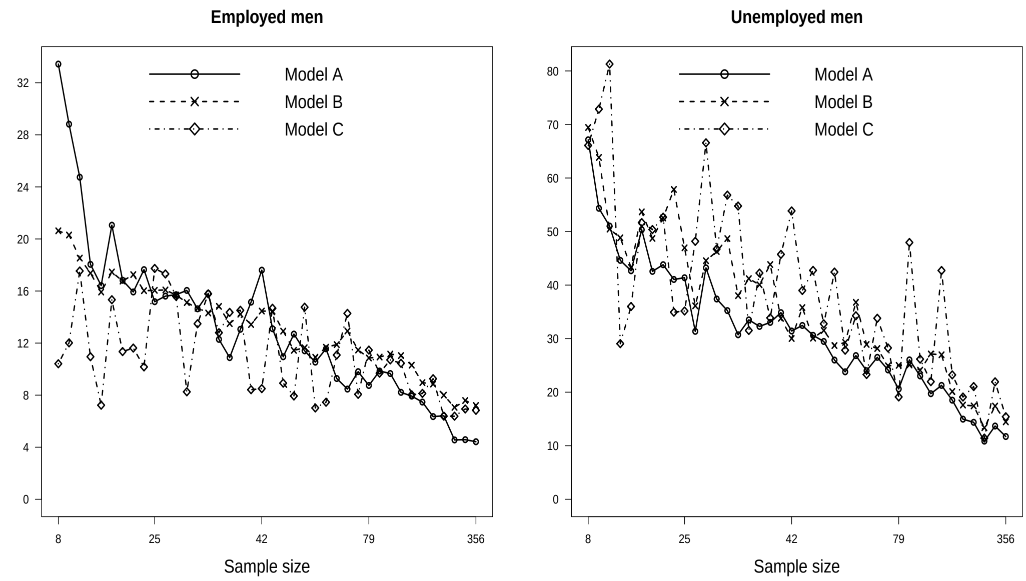


Figure 4.3.9: RMSE's for employed and unemployed men under model A, B and C.

useful when implementing model-based methodologies in official surveys by Statistical Offices.

Chapter 5

Multinomial mixed model with independent time and area effects

Like in the model of Chapter 4 traditional small area estimators borrow strength either from similar small areas or from the same area across time. Nevertheless some authors have developed contributions borrowing strength simultaneously across sections and time. Estimators provided by Ghosh et al. (1996), Datta et al. (1999), Datta et al. (2002) and You et al. (2001) exploit the two dimensions to produce estimates with good properties for small areas. It is possible to take advantage from the availability of survey data from different time periods by introducing in the model independent time effects. This approach has been implemented in other works that use linear models, see e.g. Pfeffermann and Burck (1990), Rao and Yu (1994), Saei and Chambers (2003), Ugarte et al. (2009a) and Esteban et al. (2012).

Unlike the work by Molina et al. (2007), we employ a multinomial model with two random effects, one associated with the category of employed people and the other associated with the category of unemployed people. This is due to the different modeling requirements for each labour category in the Galician data. Indeed, to take advantage from the availability of survey data from different time periods we propose a multinomial model with independent time and domain random effects.

This chapter is organized as follows. Section 5.1 introduces the multinomial logit mixed model with time and domain random effects. Section 5.2 introduces the fitting algorithm. Section 5.3 develops the proposed model-based estimators and the corresponding MSE estimation procedures. Section 5.4 presents two simulation experiments. The first simulation studies the behavior of the estimates of the regression parameters, the variance components and the target population indicators. The second simulation studies the performance of the three introduced MSE estimation methods. Finally, Section 5.5 applies the proposed methodology to data from the SLFS in Galicia.

5.1 The model

In this section we introduce the multinomial mixed model with independent random effects that will be used on the estimation of domain totals of employed, unemployed and inactive people and of unemployment rates.

Let us first observe that at the unit level y_{dktj}, $k = 1, 2, 3$, are binary 0-1 variables such that $y_{d1tj} + y_{d2tj} + y_{d3tj} = 1$. Therefore an adequate probability distribution for $\boldsymbol{y}_{dtj} = (y_{d1tj}, y_{d2tj})$ is multi-Bernoulli with unknown probability vector parameter $\boldsymbol{p}_{dtj} = (p_{d1tj}, p_{d2tj})$, such that $p_{d1tj} > 0$, $p_{d2tj} > 0$ and $p_{d1tj} + p_{d2tj} < 1$. If we assume that the random vectors $\boldsymbol{y}_{dtj}$ are independent and that the corresponding probability vectors $\boldsymbol{p}_{dtj}$ are constant within domain d at time t (i.e. $\boldsymbol{p}_{dtj} = \boldsymbol{p}_{dt}$ for all $j \in P_{dt}$), then $\boldsymbol{y}_{dt} = \sum_{j \in S_{dt}} \boldsymbol{y}_{dtj}$ is multinomially distributed with size parameter $n_{dt} = \#(S_{dt})$ and probability vector $\boldsymbol{p}_{dt}$. This is because we introduce a multinomial logit mixed model for estimating the domain totals of employed, unemployed and inactive people and the unemployment rates. The multinomial models guarantee that the sum of estimated totals of employed and unemployed people is lower than the total of people. This is a desirable property that allows estimating the total of inactive people in a coherent way.

Let us start by giving some notation and assumptions for the more general problem with q categories. Let us use indexes $d = 1, \ldots, D$, $k = 1, \ldots, q-1$ and $t = 1, \ldots, T$ for the D domains, for the $q-1$ model categories of the target variable and for the T time periods, respectively. We recall that there are $q = 3$ categories in the real data presented in Section 5.2, i.e. employed, unemployed and inactive people. However, there are only $q - 1 = 2$ categories in the multinomial model, i.e. employed and unemployed people and we considered the third category as the reference. Let $u_{1,dk}$ and $u_{2,dkt}$ be the random effects associated to category k, domain d and time t. We suppose that the random effects $\boldsymbol{u}_{1,d} = \underset{1 \le k \le q-1}{\text{col}}(u_{1,dk}) \sim N(\boldsymbol{0}, \underset{1 \le k \le q-1}{\text{diag}}(\varphi_{1k}))$ and $\boldsymbol{u}_{2,dt} = \underset{1 \le k \le q-1}{\text{col}}(u_{2,dkt}) \sim N(\boldsymbol{0}, \underset{1 \le k \le q-1}{\text{diag}}(\varphi_{2k}))$, $d = 1, \ldots, D$, $t = 1, \ldots, T$, are independent. We also assume that the response vectors $\boldsymbol{y}_{dt} = \underset{1 \le k \le q-1}{\text{col}}(y_{dkt})$, conditioned to $\boldsymbol{u}_{1,d}$ and $\boldsymbol{u}_{2,dt}$, are independent with multinomial distributions

$$\boldsymbol{y}_{dt}|_{\boldsymbol{u}_{1,d}, \boldsymbol{u}_{2,dt}} \sim \text{M}(\nu_{dt}, p_{d1t}, \ldots, p_{dqt-1}), \ d = 1, \ldots, D, \ t = 1, \ldots, T, \tag{5.1.1}$$

where the ν_{dt}'s are known integer numbers which are equal to n_{dt} in the considered real data case. The covariance matrix of $\boldsymbol{y}_{dt}$ conditioned to $\boldsymbol{u}_{1,d}$ and $\boldsymbol{u}_{2,dt}$ is $\text{var}(\boldsymbol{y}_{dt}|\boldsymbol{u}_{1,d}, \boldsymbol{u}_{2,dt}) = \boldsymbol{W}_{dt} = \nu_{dt}[\text{diag}(\boldsymbol{p}_{dt}) - \boldsymbol{p}_{dt}\boldsymbol{p}_{dt}']$, where $\boldsymbol{p}_{dt} = \underset{1 \le k \le q-1}{\text{col}}(p_{dkt})$ and $\text{diag}(\boldsymbol{p}_{dt}) = \underset{1 \le k \le q-1}{\text{diag}}(p_{dkt})$. For the natural parameters $\eta_{dkt} = \log \frac{p_{dkt}}{p_{dqt}}$, $k = 1, \ldots, q-1$, we assume the model

$$\eta_{dkt} = \boldsymbol{x}_{dkt}\boldsymbol{\beta}_k + u_{1,dk} + u_{2,dkt}, \ \ d = 1, \ldots, D, \ k = 1, \ldots, q-1, \ t = 1, \ldots, T, \tag{5.1.2}$$

where $\boldsymbol{x}_{dkt} = \underset{1 \le r \le l_k}{\text{col}'}(x_{dktr})$, $\boldsymbol{\beta}_k = \underset{1 \le r \le l_k}{\text{col}}(\beta_{kr})$ and $l = \sum_{k=1}^{q-1} l_k$. The mean and the variance of y_{dkt}, conditioned to $\boldsymbol{u}_{1,d}$ and $\boldsymbol{u}_{2,dt}$, are $\mu_{dkt} = \nu_{dt}p_{dkt}$ and $w_{dkt} = \nu_{dt}p_{dkt}(1 - p_{dkt})$, respectively. The probability of the multinomial category k at the domain d and the time instant t is

$$p_{dkt} = \frac{\exp\{\eta_{dkt}\}}{1 + \sum_{\ell=1}^{q-1} \exp\{\eta_{d\ell t}\}}, \quad d = 1, \ldots, D, \; k = 1, \ldots, q-1, \; t = 1, \ldots, T.$$

In matrix notation, the model is

$$\boldsymbol{\eta}_{dt} = \boldsymbol{X}_{dt}\boldsymbol{\beta} + \boldsymbol{Z}_{1dt}\boldsymbol{u}_{1,d} + \boldsymbol{Z}_{2dt}\boldsymbol{u}_{2,dt}, \quad d = 1, \ldots, D, \; t = 1, \ldots, T, \tag{5.1.3}$$

where $\boldsymbol{\eta}_{dt} = \underset{1\leq k\leq q-1}{\text{col}}(\eta_{dkt})$, $\boldsymbol{X}_{dt} = \underset{1\leq k\leq q-1}{\text{col}}(\boldsymbol{x}_{dkt})$, $\boldsymbol{\beta} = \underset{1\leq k\leq q-1}{\text{col}}(\boldsymbol{\beta}_k)$, $\boldsymbol{Z}_{1dt} = \boldsymbol{Z}_{2dt} = \boldsymbol{I}_{q-1}$ with $\boldsymbol{I}_{q-1}$ being the $(q-1)\times(q-1)$ unit matrix. If we introduce the additional notation $\boldsymbol{\eta} = \underset{1\leq d\leq D}{\text{col}}(\boldsymbol{\eta}_d)$, $\boldsymbol{\eta}_d = \underset{1\leq t\leq T}{\text{col}}(\boldsymbol{\eta}_{dt})$, $\boldsymbol{X} = \underset{1\leq d\leq D}{\text{col}}(\boldsymbol{X}_d)$, $\boldsymbol{X}_d = \underset{1\leq t\leq T}{\text{col}}(\boldsymbol{X}_{dt})$, $\boldsymbol{u}_1 = \underset{1\leq d\leq D}{\text{col}}(\boldsymbol{u}_{1,d})$, $\boldsymbol{u}_2 = \underset{1\leq d\leq D}{\text{col}}(\underset{1\leq t\leq T}{\text{col}}(\boldsymbol{u}_{2,dt}))$, $\boldsymbol{Z} = (\boldsymbol{Z}_1, \boldsymbol{Z}_2)$, $\boldsymbol{u} = (\boldsymbol{u}_1', \boldsymbol{u}_2')'$, $\boldsymbol{Z}_1 = \underset{1\leq d\leq D}{\text{diag}}(\boldsymbol{Z}_{1d})$, $\boldsymbol{Z}_{1d} = \underset{1\leq t\leq T}{\text{col}}(\boldsymbol{Z}_{1dt})$, $\boldsymbol{Z}_2 = \underset{1\leq d\leq D}{\text{diag}}(\boldsymbol{Z}_{2d})$, $\boldsymbol{Z}_{2d} = \underset{1\leq t\leq T}{\text{diag}}(\boldsymbol{Z}_{2dt})$, then (5.1.3) can be represented in the matrix form

$$\boldsymbol{\eta} = \boldsymbol{X\beta} + \boldsymbol{Z}_1\boldsymbol{u}_1 + \boldsymbol{Z}_2\boldsymbol{u}_2 = \boldsymbol{X\beta} + \boldsymbol{Zu}. \tag{5.1.4}$$

Alternatively the model (5.1.3) can be expressed as

$$\boldsymbol{\eta}_{dt} = \boldsymbol{X}_{dt}\boldsymbol{\beta} + \mathsf{Z}_{1,dt}\mathsf{u}_1 + \mathsf{Z}_{2,dt}\mathsf{u}_2 = \boldsymbol{X}_{dt}\boldsymbol{\beta} + \mathsf{Z}_{dt}\mathsf{u}, \quad d = 1, \ldots, D, \; t = 1, \ldots, T, \tag{5.1.5}$$

where $\mathsf{u} = (\mathsf{u}_1', \mathsf{u}_2')'$, $\mathsf{u}_1 = \underset{1\leq k\leq q-1}{\text{col}}(\boldsymbol{u}_{1,k})$, $\mathsf{u}_2 = \underset{1\leq k\leq q-1}{\text{col}}(\boldsymbol{u}_{2,k})$, $\mathsf{Z}_{dt} = (\mathsf{Z}_{1,dt}, \mathsf{Z}_{2,dt})$, $\mathsf{Z}_{1,dt} = \underset{1\leq k\leq q-1}{\text{diag}}(\underset{1\leq \ell\leq D}{\text{col}'}(\delta_{\ell d}))$ and $\mathsf{Z}_{2,dt} = \underset{1\leq k\leq q-1}{\text{diag}}(\underset{1\leq \ell\leq D}{\text{col}'}(\underset{1\leq t\leq T}{\text{col}'}(\delta_{\ell d}\delta_{st})))$. We further assume that $\boldsymbol{u}_{1,k} = \underset{1\leq d\leq D}{\text{col}}(u_{1,dk}) \sim N(\boldsymbol{0}, \varphi_{1k}\boldsymbol{I}_D)$ and $\boldsymbol{u}_{2,k} = \underset{1\leq d\leq D}{\text{col}}(\underset{1\leq t\leq T}{\text{col}}(u_{2,dk})) \sim N(\boldsymbol{0}, \varphi_{2k}\boldsymbol{I}_{DT})$ are independent. In matrix notation (5.1.5) can be expressed as

$$\boldsymbol{\eta} = \boldsymbol{X\beta} + \mathsf{Z}_1\mathsf{u}_1 + \mathsf{Z}_2\mathsf{u}_2 = \boldsymbol{X\beta} + \mathsf{Z}\mathsf{u}, \tag{5.1.6}$$

where $\mathsf{Z}_r = \underset{1\leq d\leq D}{\text{col}}(\underset{1\leq t\leq T}{\text{col}}(\mathsf{Z}_{r,dt}))$, $r = 1, 2$, and $\mathsf{Z} = (\mathsf{Z}_1, \mathsf{Z}_2)$. It should be remembered that (5.1.4) and (5.1.6) are the same model.

5.2 The PQL-REML fitting algorithm

To fit the model we combine the PQL method, described by Breslow and Clayton (1996) for estimating and predicting the $\boldsymbol{\beta}_k$'s, the $\boldsymbol{u}_{1d}$'s and the $\boldsymbol{u}_{2d}$'s, with the REML method for estimating the variance components $\boldsymbol{\varphi}_1$ and $\boldsymbol{\varphi}_2$. The presented method was described in Section 3.2. It is based on a normal approximation to the joint probability distribution of the vector $(\boldsymbol{y}, \boldsymbol{u})$. The combined algorithm was first introduced by Schall (1991) and later used by Saei and Chambers (2003) and Molina et al. (2007) in applications of generalized linear mixed models to small area estimation problems. In this chapter, we adapt the combined algorithm to the multinomial logit mixed model defined by (5.1.1) and (5.1.2). The log-likelihood of $\boldsymbol{y}$ conditioned to $\boldsymbol{u}$ is

$$\begin{aligned}
l_1(\boldsymbol{y}|\boldsymbol{u}) &= \sum_{d=1}^{D}\sum_{t=1}^{T}\left\{\sum_{k=1}^{q-1} y_{dkt}\log\frac{p_{dkt}}{p_{dqt}} + \nu_{dt}\log p_{dqt} + \log\frac{\nu_{dt}}{y_{d1t}!\cdots y_{dqt}!}\right\}\\
&= \sum_{d=1}^{D}\sum_{t=1}^{T}\left\{\sum_{k=1}^{q-1} y_{dkt}\eta_{dkt} - \nu_{dt}\log\big(1+\sum_{k=1}^{q-1}\exp\{\eta_{dkt}\}\big) + \log\frac{\nu_{dt}}{y_{d1t}!\cdots y_{dqt}!}\right\}.
\end{aligned}$$

The partial derivatives of

$$\eta_{dkt} = \sum_{r=1}^{p} x_{dktr}\beta_r + u_{1,dk} + u_{2,dkt}, d=1,\ldots,D,\ t=1,\ldots,T,\ k=1,\ldots,q-1,$$

with respect to β_r, $u_{1,dk}$ and $u_{2,dkt}$ are

$$\frac{\partial\eta_{dkt}}{\partial\beta_r} = x_{dktr},\quad \frac{\partial\eta_{dkt}}{\partial u_{1,dk}} = 1,\quad \frac{\partial\eta_{dkt}}{\partial u_{2,dkt}} = 1.$$

The first order partial derivatives of l_1 are

$$\begin{aligned}
S_{1,\beta_r} &= \frac{\partial l_1(\boldsymbol{y}|\boldsymbol{u})}{\partial\beta_r} = \sum_{d=1}^{D}\sum_{t=1}^{T}\left\{\sum_{k=1}^{q-1} x_{dktr}y_{dkt} - \frac{\nu_{dt}\sum_{k=1}^{q-1} x_{dktr}\exp\{\eta_{dkt}\}}{1+\sum_{k=1}^{q-1}\exp\{\eta_{dkt}\}}\right\}\\
&= \sum_{d=1}^{D}\sum_{t=1}^{T}\sum_{k=1}^{q-1} x_{dktr}(y_{dkt}-\mu_{dkt}),\\
S_{1,u_{1,dk}} &= \frac{\partial l_1(\boldsymbol{y}|\boldsymbol{u})}{\partial u_{1,dk}} = \sum_{t=1}^{T}(y_{dkt}-\mu_{dkt}),\quad S_{1,u_{2,dkt}} = \frac{\partial l_1(\boldsymbol{y}|\boldsymbol{u})}{\partial u_{2,dkt}} = (y_{dkt}-\mu_{dkt}).
\end{aligned}$$

The vector expressions of the first order partial derivatives of l_1 are

$$\begin{aligned}
\boldsymbol{S}_{1,\beta} &= \frac{\partial l_1(\boldsymbol{y}|\boldsymbol{u})}{\partial\boldsymbol{\beta}} = \sum_{d=1}^{D}\sum_{t=1}^{T}\sum_{k=1}^{q-1}\boldsymbol{x}'_{dkt}(y_{dkt}-\mu_{dkt}) = \underset{1\le r\le p}{\operatorname{col}}(S_{1,\beta_r}),\\
\boldsymbol{S}_{1,u_1} &= \frac{\partial l_1(\boldsymbol{y}|\boldsymbol{u})}{\partial\boldsymbol{u}_1} = \underset{1\le d\le D}{\operatorname{col}}(\underset{1\le k\le q-1}{\operatorname{col}}(S_{1,u_{1,dk}})),\\
\boldsymbol{S}_{1,u_2} &= \frac{\partial l_1(\boldsymbol{y}|\boldsymbol{u})}{\partial\boldsymbol{u}_2} = \underset{1\le d\le D}{\operatorname{col}}(\underset{1\le k\le q-1}{\operatorname{col}}(\underset{1\le t\le T}{\operatorname{col}}(S_{1,u_{2,dkt}}))).
\end{aligned}$$

The second order partial derivatives of l_1 are

$$\begin{aligned}
H_{1,\beta_{kr}\beta_{ks}} &= \frac{\partial^2 l_1(\boldsymbol{y}|\boldsymbol{u})}{\partial\beta_{kr}\partial\beta_{ks}} = -\sum_{d=1}^{D}\sum_{t=1}^{T}\nu_{dt}x_{dktr}x_{dkts}p_{dkt}(1-p_{dkt}),\\
H_{1,u_{1,dk}\beta_{kr}} &= \frac{\partial^2 l_1(\boldsymbol{y}|\boldsymbol{u})}{\partial u_{1,dk}\partial\beta_{kr}} = -\sum_{t=1}^{T}\nu_{dt}x_{dktr}p_{dkt}(1-p_{dkt}),\\
H_{1,u_{2,dkt}\beta_{kr}} &= \frac{\partial^2 l_1(\boldsymbol{y}|\boldsymbol{u})}{\partial u_{2,dkt}\partial\beta_{kr}} = -\nu_{dt}x_{dktr}p_{dkt}(1-p_{dkt}),\\
H_{1,u_{1,dk},u_{1,dk}} &= \frac{\partial^2 l_1(\boldsymbol{y}|\boldsymbol{u})}{\partial u_{1,dk}\partial u_{1,dk}} = -\sum_{t=1}^{T}\nu_{dt}p_{dkt}(1-p_{dkt})\\
H_{1,u_{2,dkt},u_{2,dkt}} &= \frac{\partial^2 l_1(\boldsymbol{y}|\boldsymbol{u})}{\partial u_{2,dkt}\partial u_{2,dkt}} = -\nu_{dt}p_{dkt}(1-p_{dkt}),\\
H_{1,u_{1,dk},u_{2,dkt}} &= \frac{\partial^2 l_1(\boldsymbol{y}|\boldsymbol{u})}{\partial u_{1,dk}\partial u_{2,dkt}} = -\nu_{dt}p_{dkt}(1-p_{dkt}),
\end{aligned}$$

and for the case $k_1 \neq k_2$, we have

$$\begin{aligned}
H_{1,\beta_{k_1r}\beta_{k_2s}} &= \frac{\partial^2 l_1(\boldsymbol{y}|\boldsymbol{u})}{\partial\beta_{k_1r}\partial\beta_{k_2s}} = \sum_{d=1}^{D}\sum_{t=1}^{T}\nu_{dt}x_{dk_1tr}x_{dk_2s}p_{dk_1t}p_{dk_2t},\\
H_{1,u_{1,dk_1}\beta_{k_2s}} &= \frac{\partial^2 l_1(\boldsymbol{y}|\boldsymbol{u})}{\partial u_{1,dk_1}\partial\beta_{k_2s}} = \sum_{t=1}^{T}\nu_{dt}x_{dk_2ts}p_{dk_1t}p_{dk_2t},\\
H_{1,u_{2,dkt_1}\beta_{k_2s}} &= \frac{\partial^2 l_1(\boldsymbol{y}|\boldsymbol{u})}{\partial u_{2,dk_1t}\partial\beta_{k_2s}} = \nu_{dt}x_{dk_2ts}p_{d_1kt}p_{dk_2t},\\
H_{1,u_{1,dk_1},u_{1,dk_2}} &= \frac{\partial^2 l_1(\boldsymbol{y}|\boldsymbol{u})}{\partial u_{1,dk_1}\partial u_{1,dk_2}} = \sum_{t=1}^{T}\nu_{dt}p_{dk_1t}p_{dk_2t},\\
H_{1,u_{2,dk_1t},u_{2,dk_2t}} &= \frac{\partial^2 l_1(\boldsymbol{y}|\boldsymbol{u})}{\partial u_{2,dk_1t}\partial u_{2dk_2t}} = \nu_{dt}p_{dkt_1}p_{dkt_2},\\
H_{1,u_{1,dk_1}u_{2,dkt_2}} &= \frac{\partial^2 l_1(\boldsymbol{y}|\boldsymbol{u})}{\partial u_{1,dk_1}\partial u_{2,dk_2t}} = \nu_{dt}p_{dk_1t}p_{dk_2t}.
\end{aligned}$$

The second order partial derivatives of l_1 are

$$\begin{aligned}
H_{1,u_{1,d_1k_1},u_{1,d_2,k_2}} &= \frac{\partial^2 l_1(\boldsymbol{y}|\boldsymbol{u})}{\partial u_{1,d_1k_1}\partial u_{1,d_2,k_2}} = 0,\ d_1 \neq d_2,\\
H_{1,u_{2,d_1k_1t_1},u_{2,d_2k_2t_2}} &= \frac{\partial^2 l_1(\boldsymbol{y}|\boldsymbol{u})}{\partial u_{2,d_1k_1t_1}\partial u_{2,d_2k_2t_2}} = 0,\ d_1 \neq d_2 \text{ or } t_1 \neq t_2,\\
H_{1,u_{1,d_1k_1}u_{2k_2,d_2t}} &= \frac{\partial^2 l_1(\boldsymbol{y}|\boldsymbol{u})}{\partial u_{1,d_1k_1}\partial u_{2,d_2k_2t}} = 0,\ d_1 \neq d_2.
\end{aligned}$$

The matrix expressions of the second order partial derivatives of l_1 are

$$\boldsymbol{H}_{1,\beta_{k_1}\beta_{k_2}} = \Big(H_{1,\beta_{k_1 r}\beta_{k_2 s}}\Big)_{\substack{r=1,\dots,p_{k_1}\\ s=1,\dots,p_{k_2}}}, \quad \boldsymbol{H}_{1,\beta\beta} = \Big(\boldsymbol{H}_{1,\beta_{k_1}\beta_{k_2}}\Big)_{\substack{k_1=1,\dots,q-1\\ k_2=1,\dots,q-1}},$$

$$\boldsymbol{H}_{1,u_{1,d}\beta_{k_2}} = \Big(H_{1,u_{1,dk_1}\beta_{k_2 s}}\Big)_{\substack{k_1=1,\dots,q-1\\ s=1,\dots,p_{k_2}}}, \quad \boldsymbol{H}_{1,u\beta} = \Big(\boldsymbol{H}_{1,u_d\beta_{k_2}}\Big)_{\substack{d=1,\dots,D\\ k_2=1,\dots,q-1}},$$

$$\boldsymbol{H}_{1,u_{2,dt}\beta_{k_2}} = \Big(H_{1,u_{2,dk_1 t}\beta_{k_2 s}}\Big)_{\substack{k_1=1,\dots,q-1\\ s=1,\dots,p_{k_2}}}, \quad \boldsymbol{H}_{1,u_{2,d}\beta} = \Big(\boldsymbol{H}_{1,u_{2,dt}\beta_{k_2}}\Big)_{\substack{t=1,\dots,T\\ k_2=1,\dots,q-1}},$$

$$\boldsymbol{H}_{1,u_2\beta} = \underset{1\le d\le D}{\mathrm{col}}\big(\boldsymbol{H}_{1,u_{2,d}\beta}\big), \quad \boldsymbol{H}_{1,u_{1,d_1}u_{2,d_2t}} = \Big(H_{1,u_{1,d_1k_1},u_{2,d_2k_2t}}\Big)_{\substack{k_1=1,\dots,q-1\\ k_2=1,\dots,q-1}},$$

$$\boldsymbol{H}_{1,u_{1,d_1},u_{2,d_2}} = \underset{1\le t\le T}{\mathrm{col}'}\big(\boldsymbol{H}_{1,u_{1,d_1}u_{2,d_2t}}\big), \quad \boldsymbol{H}_{1,u_1u_2} = \Big(\boldsymbol{H}_{1,u_{1,d_1},u_{2,d_2}}\Big)_{\substack{d_1=1,\dots,D\\ d_2=1,\dots,D}}.$$

Equivalently, we can write

$$\boldsymbol{H}_{1,\beta\beta} = \Big(H_{1,\beta_{k_1 r}\beta_{k_2 s}}\Big)_{\substack{r=1,\dots,p_{k_1};k_1=1,\dots,q-1\\ s=1,\dots,p_{k_2};k_2=1,\dots,q-1}},$$

$$\boldsymbol{H}_{1,\beta u_1} = \Big(H_{1,\beta_{k_1 r}u_{1,dk_2}}\Big)_{\substack{k_1=1,\dots,q-1;r=1,\dots,p_{k_1}\\ d=1,\dots,D;k_2=1,\dots,q-1}},$$

$$\boldsymbol{H}_{1,\beta u_2} = \Big(H_{1,\beta_{k_1 r}u_{2,dkt_2}}\Big)_{\substack{k_1=1,\dots,q-1;r=1,\dots,p_{k_1}\\ d=1,\dots,D;k_2=1,\dots,q-1;t=1,\dots,T}},$$

$$\boldsymbol{H}_{1,u_1u_1} = \Big(H_{1,u_{1,d_1k_1},u_{1,d_2k_2}}\Big)_{\substack{d_1=1,\dots,D;k_1=1,\dots,q-1\\ d_2=1,\dots,D;k_2=1,\dots,q-1}},$$

$$\boldsymbol{H}_{1,u_2u_2} = \Big(H_{1,u_{2,d_1k_1t_1},u_{2,d_2k_2t_2}}\Big)_{\substack{d_1=1,\dots,D;k_1=1,\dots,q-1;t_1=1,\dots,T\\ d_2=1,\dots,D;k_2=1,\dots,q-1;t_2=1,\dots,T}},$$

$$\boldsymbol{H}_{1,u_1u_2} = \Big(H_{1,u_{1d_1k_1},u_{2,d_2k_2t}}\Big)_{\substack{d_1=1,\dots,D;k_1=1,\dots,q-1\\ d_2=1,\dots,D;k_2=1,\dots,q-1;t=1,\dots,T}}.$$

It holds that

$$\boldsymbol{H}_{1,\beta\beta} = -\sum_{d=1}^{D}\sum_{k=1}^{q-1}\sum_{t=1}^{T}\nu_{dt}p_{dkt}(1-p_{dkt})\boldsymbol{x}'_{dkt}\boldsymbol{x}_{dkt},$$

$$\boldsymbol{H}_{1,u_1\beta} = \underset{1\le d\le D}{\mathrm{col}}\Big(\underset{1\le k\le q-1}{\mathrm{col}}\Big(-\sum_{t=1}^{T}\nu_{dt}p_{dkt}(1-p_{dkt})\boldsymbol{x}_{dkt}\Big)\Big),$$

$$\boldsymbol{H}_{1,u_2\beta} = \underset{1\le d\le D}{\mathrm{col}}\Big(\underset{1\le k\le q-1}{\mathrm{col}}\Big(\underset{1\le t\le T}{\mathrm{col}}\Big(-\nu_{dt}p_{dkt}(1-p_{dkt})\boldsymbol{x}_{dkt}\Big)\Big)\Big).$$

The log-likelihood of $\boldsymbol{u}$ is

$$f_2(\boldsymbol{u}) = \frac{1}{(\sqrt{2\pi})^{(M+D)(q-1)}|\boldsymbol{V}_u|^{1/2}}\exp\left\{-\frac{1}{2}\boldsymbol{u}'\boldsymbol{V}_u^{-1}\boldsymbol{u}\right\},$$

where $M = DT$. The log-likelihood of $\boldsymbol{u}$ is

$$l_2(\boldsymbol{u}) = \kappa - \sum_{k=1}^{q-1}\left\{\frac{D}{2}\log\varphi_{1k} + \frac{M}{2}\log\varphi_{2k} + \frac{1}{2}\sum_{d=1}^{D}\log T + \frac{1}{2}\frac{\boldsymbol{u}'_{1k}\boldsymbol{u}_{1k}}{\varphi_{1k}} + \frac{1}{2}\sum_{d=1}^{D}\frac{\boldsymbol{u}'_{2,dk}\boldsymbol{u}_{2,dk}}{\varphi_{2k}}\right\},$$

where $\kappa = -\frac{(M+D)(q-1)}{2}\log 2\pi$. The first order partial derivatives of l_2 are

$$S_{2,u_{1,dk}} = \frac{\partial l_2(\boldsymbol{u})}{\partial u_{1,dk}} = -\frac{1}{\varphi_{1k}}\, u_{1,dk}, \quad S_{2,u_{2,dkt}} = \frac{\partial l_2(\boldsymbol{u})}{\partial u_{2,dkt}} = -\frac{1}{\varphi_{2k}}\, u_{2,dkt}.$$

The vector expressions of the partial derivatives of l_2 are

$$\boldsymbol{S}_{2,u_1} = \underset{1\le d\le D}{\text{col}}\Big(\underset{1\le k\le q-1}{\text{col}}\big(S_{2,u_{1,dk}}\big)\Big), \quad \boldsymbol{S}_{2,u_2} = \underset{1\le d\le D}{\text{col}}\Big(\underset{1\le k\le q-1}{\text{col}}\Big(\underset{1\le t\le T}{\text{col}}\big(S_{2,u_{2,dkt}}\big)\Big)\Big).$$

The second order partial derivatives of l_2 are

$$\begin{aligned} H_{2,u_{1,dk}u_{1,dk}} &= -\frac{1}{\varphi_{1k}}, \quad H_{2,u_{2,dkt}u_{2,dkt}} = -\frac{1}{\varphi_{2k}}, \\ H_{2,u_{1,d_1k_1}u_{1,d_2k_2}} &= 0, \ H_{2,u_{2,d_1k_1t_1}u_{2,d_2k_2t_2}} = 0, \ k_1 \neq k_2 \text{ or } d_1 \neq d_2. \end{aligned}$$

The matrix expressions of the second order partial derivatives of l_2 are $\boldsymbol{H}_{2,u_1u_2} = \boldsymbol{0}$

$$\begin{aligned} \boldsymbol{H}_{2,u_1u_1} &= \Big(H_{2,u_{1,d_1k_1},u_{1,d_2k_1}}\Big)_{\substack{d_1=1,\dots,D;k_1=1,\dots,q-1\\ d_2=1,\dots,D;k_2=1,\dots,q-1}}, \\ \boldsymbol{H}_{2,u_2u_2} &= \Big(H_{2,u_{2,d_1k_1t_1},u_{2,d_2t_2k_1}}\Big)_{\substack{d_1=1,\dots,D;k_1=1,\dots,q-1;t_1=1,\dots,T\\ d_2=1,\dots,D;k_2=1,\dots,q-1;t_2=1,\dots,T}}, \\ \boldsymbol{H}_{2,u_1u_2} &= \Big(H_{2,u_{1,d_1k_1},u_{2,d_2t_2k_1}}\Big)_{\substack{d_1=1,\dots,D;k_1=1,\dots,q-1\\ d_2=1,\dots,D;k_2=1,\dots,q-1;t_2=1,\dots,T}}. \end{aligned}$$

The log-likelihood of $(\boldsymbol{y}, \boldsymbol{u})$ is

$$l(\boldsymbol{y}, \boldsymbol{u}) = l_1(\boldsymbol{y}|\boldsymbol{u}) + l_2(\boldsymbol{u}).$$

The first order partial derivatives of l are

$$\begin{aligned} \boldsymbol{S}_\beta &= \boldsymbol{S}_{1,\beta} = \sum_{d=1}^{D}\sum_{k=1}^{q-1}\sum_{t=1}^{T} \boldsymbol{x}'_{dkt}(y_{dkt} - \mu_{dkt}), \\ \boldsymbol{S}_{u_1} &= \boldsymbol{S}_{1,u_1} + \boldsymbol{S}_{2,u_1} = \underset{1\le d\le D}{\text{col}}\Big(\underset{1\le k\le q-1}{\text{col}}\Big(\sum_{t=1}^{T}(y_{dkt} - \mu_{dkt}) - \frac{1}{\varphi_{1k}}\, u_{1,dk}\Big)\Big), \\ \boldsymbol{S}_{u_2} &= \boldsymbol{S}_{1,u_2} + \boldsymbol{S}_{2,u_2} = \underset{1\le d\le D}{\text{col}}\Big(\underset{1\le k\le q-1}{\text{col}}\Big(\underset{1\le t\le T}{\text{col}}\Big((y_{dkt} - \mu_{dkt}) - \frac{1}{\varphi_{2k}}\, u_{2,dkt}\Big)\Big)\Big). \end{aligned}$$

The blocks of the Fisher information matrix associated to l are

$$\begin{aligned} \boldsymbol{F}_{\beta,\beta} &= -\boldsymbol{H}_{1,\beta\beta},\ \boldsymbol{F}_{u_1,\beta} = -\boldsymbol{H}_{1,u_1\beta},\ \boldsymbol{F}_{\beta,u_1} = \boldsymbol{F}'_{u_1,\beta},\ \boldsymbol{F}_{u_2,\beta} = -\boldsymbol{H}_{1,u_2\beta},\ \boldsymbol{F}_{\beta,u_2} = \boldsymbol{F}'_{u_2,\beta}, \\ \boldsymbol{F}_{u_1,u_1} &= -\boldsymbol{H}_{1,u_1u_1} - \boldsymbol{H}_{2,u_1u_1}, \boldsymbol{F}_{u_2,u_2} = -\boldsymbol{H}_{1,u_2u_2} - \boldsymbol{H}_{2,u_2u_2}, \boldsymbol{F}_{u_1,u_2} = -\boldsymbol{H}_{1,u_1u_2}, \boldsymbol{F}_{u_2,u_1} = \boldsymbol{F}'_{u_1,u_2}. \end{aligned}$$

We define

$$\boldsymbol{S} = \begin{pmatrix} \boldsymbol{S}_\beta \\ \boldsymbol{S}_u \end{pmatrix}, \boldsymbol{S}_u = \begin{pmatrix} \boldsymbol{S}_{u_1} \\ \boldsymbol{S}_{u_2} \end{pmatrix}, \boldsymbol{F}_{uu} = \begin{pmatrix} \boldsymbol{F}_{u_1u_1} & \boldsymbol{F}_{u_1u_2} \\ \boldsymbol{F}_{u_2u_1} & \boldsymbol{F}_{u_2u_2} \end{pmatrix}, \boldsymbol{F}_{u\beta} = \begin{pmatrix} \boldsymbol{F}_{u_1\beta} \\ \boldsymbol{F}_{u_2\beta} \end{pmatrix}, \boldsymbol{F}_{\beta u} = \boldsymbol{F}'_{u\beta},$$

$$\boldsymbol{F}=\left(\begin{array}{cc}\boldsymbol{F}_{\beta\beta} & \boldsymbol{F}_{\beta u}\\ \boldsymbol{F}_{u\beta} & \boldsymbol{F}_{uu}\end{array}\right),\ \boldsymbol{F}^{-1}=\left(\begin{array}{cc}\boldsymbol{F}^{\beta\beta} & \boldsymbol{F}^{\beta u}\\ \boldsymbol{F}^{u\beta} & \boldsymbol{F}^{uu}\end{array}\right),\ \boldsymbol{F}_{uu}^{-1}=\left(\begin{array}{cc}\boldsymbol{F}^{u_1u_1} & \boldsymbol{F}^{u_1u_2}\\ \boldsymbol{F}^{u_2u_1} & \boldsymbol{F}^{u_2u_2}\end{array}\right).$$

It holds that

$$\begin{aligned}
\boldsymbol{F}^{\beta\beta} &= (\boldsymbol{F}_{\beta\beta}-\boldsymbol{F}_{\beta u}\boldsymbol{F}_{uu}^{-1}\boldsymbol{F}_{u\beta})^{-1}, \quad \boldsymbol{F}^{\beta u}=-\boldsymbol{F}^{\beta\beta}\boldsymbol{F}_{\beta u}\boldsymbol{F}_{uu}^{-1},\\
\boldsymbol{F}^{u\beta} &= (\boldsymbol{F}^{\beta u})', \quad \boldsymbol{F}^{uu}=\boldsymbol{F}_{uu}^{-1}+\boldsymbol{F}_{uu}^{-1}\boldsymbol{F}_{u\beta}\boldsymbol{F}^{\beta\beta}\boldsymbol{F}_{\beta u}\boldsymbol{F}_{uu}^{-1},\\
\boldsymbol{F}^{u_1u_1} &= (\boldsymbol{F}_{u_1u_1}-\boldsymbol{F}_{u_1u_2}\boldsymbol{F}_{u_2u_2}^{-1}\boldsymbol{F}_{u_2u_1})^{-1}, \quad \boldsymbol{F}^{u_1u_2}=-\boldsymbol{F}^{u_1u_1}\boldsymbol{F}_{u_1u_2}\boldsymbol{F}_{u_2u_2}^{-1},\\
\boldsymbol{F}^{u_2u_1} &= (\boldsymbol{F}^{u_1u_2})', \quad \boldsymbol{F}^{u_2u_2}=\boldsymbol{F}_{u_2u_2}^{-1}+\boldsymbol{F}_{u_2u_2}^{-1}\boldsymbol{F}_{u_2u_1}\boldsymbol{F}^{u_1u_1}\boldsymbol{F}_{u_1u_2}\boldsymbol{F}_{u_2u_2}^{-1}.
\end{aligned}$$

The algorithm has two parts: A and B. In the first part the algorithm updates the values of $\boldsymbol{\beta}_k$, $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$. In the second part it updates the variance components. Algorithm A maximizes the joint log-likelihood $l(\boldsymbol{y},\boldsymbol{u})=l_1(\boldsymbol{y}|\boldsymbol{u})+l_2(\boldsymbol{u})$, with known vector of variances $\boldsymbol{\varphi}$.

Algorithm A.

(A.1) *Beginning:* Assign the initial values $l=0$, $\boldsymbol{\beta}^{(0)}=\boldsymbol{\beta}^{initial}$ and $\boldsymbol{u}^{(0)}=\boldsymbol{u}^{initial}$.

(A.2) *Iteration $l+1$:* For $d=1,\dots,D$, $k=1,\dots,q-1$, $t=1,\dots,T$ calculate $\boldsymbol{F}(\boldsymbol{\beta}^{(l)},\boldsymbol{u}^{(l)})$ and $\boldsymbol{S}(\boldsymbol{\beta}^{(l)},\boldsymbol{u}^{(l)})$ and update $\boldsymbol{\beta}^{(l)}$ and $\boldsymbol{u}^{(l)}$ by using the equation

$$\left[\begin{array}{c}\boldsymbol{\beta}^{(l+1)}\\ \boldsymbol{u}^{(l+1)}\end{array}\right]=\left[\begin{array}{c}\boldsymbol{\beta}^{(l)}\\ \boldsymbol{u}^{(l)}\end{array}\right]+\boldsymbol{F}^{-1}(\boldsymbol{\beta}^{(l)},\boldsymbol{u}^{(l)})\boldsymbol{S}(\boldsymbol{\beta}^{(l)},\boldsymbol{u}^{(l)}), \tag{5.2.1}$$

where $\boldsymbol{S}$ and $\boldsymbol{F}$ are the vector of scores (first order partial derivatives) and the Fisher information matrix (minus expectation of second order partial derivatives) of the joint log-likelihood of $(\boldsymbol{y},\boldsymbol{u})$.

(A.3) *End:* Repeat the step (A.2) until convergence of $\boldsymbol{\beta}^{(l)}$ and $\boldsymbol{u}^{(l)}$ and obtain the final values $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{u}}$.

Algorithm A maximizes $l(\boldsymbol{y},\boldsymbol{u})$ in $\boldsymbol{\beta}$ and $\boldsymbol{u}$ for fixed values of $\boldsymbol{\varphi}$. To update the values of the variance components, we assume that $\boldsymbol{\beta}$ and $\boldsymbol{u}$ are known and we adapt the ideas of Schall (1991) to a multivariate setting. We consider a Taylor expansion of

$$\xi_{dkt}=g_k(\boldsymbol{y}_{dt})=\log\frac{y_{dkt}}{\nu_{dt}-\sum_{\ell=1}^{q-1}y_{d\ell t}}$$

around the point $\boldsymbol{\mu}_{dt}=\nu_{dt}\boldsymbol{p}_{dt}$. We obtain

$$\boldsymbol{\xi}_{dt}\approx\boldsymbol{X}_{dt}\boldsymbol{\beta}+\boldsymbol{Z}_{1dt}\boldsymbol{u}_{1d}+\boldsymbol{Z}_{2dt}\boldsymbol{u}_{2d}+\boldsymbol{e}_{dt}, \tag{5.2.2}$$

where $\boldsymbol{\xi}_{dt}=\underset{1\leq k\leq q-1}{\operatorname{col}}(\xi_{dkt})$, $\boldsymbol{e}_{dt}=\boldsymbol{W}_d^{-1}(\boldsymbol{y}_d-\boldsymbol{\mu}_d)$. By assuming equality in (5.2.2), it holds that $E[\boldsymbol{\xi}]=\boldsymbol{X}\boldsymbol{\beta}$ and $\boldsymbol{V}=\operatorname{var}(\boldsymbol{\xi})=\boldsymbol{Z}\boldsymbol{V}_u\boldsymbol{Z}'+\boldsymbol{W}^{-1}$. Schall (1991) proposed to update the variance components by maximizing the normal approximation to the distribution of $\boldsymbol{\xi}$, with $\boldsymbol{\beta}$ and $\boldsymbol{u}$ fixed. This proposal assumes the approximation of $l_1(\boldsymbol{y})$ by the log-likelihood $l_1(\boldsymbol{\xi})$ of the cited multivariate normal distribution. The basic idea is thus maximizing $l(\boldsymbol{\xi}, \boldsymbol{u})$ instead of $l(\boldsymbol{y}, \boldsymbol{u})$, where $\boldsymbol{\xi}$ is assumed to follow model (5.2.2) under normality. The approximating REML log-likelihood is

$$l_{reml}(\boldsymbol{\xi}) = -\frac{1}{2}(DT(q-1)-l)\log 2\pi - \frac{1}{2}\log|\boldsymbol{K}^t\boldsymbol{V}\boldsymbol{K}| - \frac{1}{2}\boldsymbol{\xi}^t\boldsymbol{P}\boldsymbol{\xi},$$

where $\boldsymbol{P} = \boldsymbol{V}^{-1} - \boldsymbol{V}^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}^{-1}$ and $\boldsymbol{K} = \boldsymbol{W} - \boldsymbol{W}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{W}$.

Algorithm B update the variance components. This can be done by applying the Fisher-Scoring algorithm to the REML log-likelihood. The algorithm is described below.

Algorithm B.

(B.1) Do $\ell = 1$. Initiate the values $\boldsymbol{\beta}^{(0)}$, $\boldsymbol{u}^{(0)}$ and $\boldsymbol{\varphi}^{(0)}$.

(B.2) Run the *Algorithm A* by using $\boldsymbol{\varphi}^{(\ell)}$ as known value of the vector of variances and $\boldsymbol{\beta}^{(\ell-1)}$ and $\boldsymbol{u}^{(\ell-1)}$ as initial values. Let $\boldsymbol{\beta}^{(\ell)}$ and $\boldsymbol{u}^{(\ell)}$ be the obtained estimates and predictors.

(B.3) For $d = 1, \ldots, D$, $t = 1, \ldots, T$ calculate $\boldsymbol{\eta}_{dt}^{(\ell)} = \boldsymbol{X}_{dt}\boldsymbol{\beta}^{(\ell)} + \boldsymbol{Z}_{1dt}\boldsymbol{u}_{1d}^{(\ell)} + \boldsymbol{Z}_{2dt}\boldsymbol{u}_{2dt}^{(\ell)}$ and apply the updating equation

$$\boldsymbol{\varphi}^{(\ell+1)} = \boldsymbol{\varphi}^{(\ell)} + \boldsymbol{F}^{-1}(\boldsymbol{\varphi}^{(\ell)})\boldsymbol{S}(\boldsymbol{\varphi}^{(\ell)}),$$

where $\boldsymbol{S}$ and $\boldsymbol{F}$ are the vector of scores and the Fisher information matrix of the log-likelihood $l_{reml}(\boldsymbol{\xi})$.

(B.4) Repeat the steps (B.2)-(B.3) until the convergence of $\boldsymbol{\beta}^{(\ell)}$, $\boldsymbol{u}^{(\ell)}$ and $\boldsymbol{\varphi}^{(\ell)}$.

The variance components φ_{rk}, $r = 1, 2$, $k = 1, \ldots, q-1$, can also be updated by using the formula

$$\widehat{\varphi}_{rk} = \frac{\widehat{\boldsymbol{u}}_{rk}'\boldsymbol{\Sigma}_{u_{rk}}^{-1}\widehat{\boldsymbol{u}}_{rk}}{\dim(\boldsymbol{u}_{rk}) - \tau_{rk}} = \frac{\widehat{\boldsymbol{u}}_{rk}'\widehat{\boldsymbol{u}}_{rk}}{A_r - \tau_{rk}},$$

where $\boldsymbol{\Sigma}_{u_{1k}} = \boldsymbol{I}_D$, $\boldsymbol{\Sigma}_{u_{2k}} = \boldsymbol{I}_{DT}$, $A_1 = D$, $A_2 = DT$, $\tau_{rk} = \frac{1}{\widehat{\varphi}_{rk}}\mathrm{tr}\big(\boldsymbol{\Sigma}_{u_{rk}}^{-1}\widehat{\mathsf{T}}_{rkk}^{rml}\big) = \frac{1}{\widehat{\varphi}_{rk}}\mathrm{tr}\big(\widehat{\mathsf{T}}_{rkk}^{rml}\big)$ and T_{rkk}^{rml} is the block (k,k) of dimension A_r of the matrix

$$\widehat{\mathsf{T}}_r^{rml} = \widehat{\mathsf{T}}_r + \widehat{\mathsf{T}}_r\mathsf{Z}_r'\boldsymbol{W}\boldsymbol{X}(\boldsymbol{X}'\widehat{\boldsymbol{V}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{W}\mathsf{Z}_r\widehat{\mathsf{T}}_r,$$

where $\widehat{\mathsf{T}}_r = (\mathsf{Z}_r\boldsymbol{W}\mathsf{Z}_r' + \widehat{\boldsymbol{\Sigma}}_{u_r}^{-1})^{-1}$, $\widehat{\boldsymbol{\Sigma}}_{u_r} = \mathrm{diag}(\varphi_{r1}\boldsymbol{I}_{A_r}, \ldots, \varphi_{rq-1}\boldsymbol{I}_{A_r})$, $\widehat{\boldsymbol{V}} = \mathsf{Z}\widehat{\boldsymbol{\Sigma}}_u\mathsf{Z}' + \boldsymbol{W}^{-1}$ and $\boldsymbol{\Sigma}_u = \mathrm{diag}(\boldsymbol{\Sigma}_{u_1}, \boldsymbol{\Sigma}_{u_2})$. Then, we have an alternative Algorithm B.

Algorithm B (alternative).

(B.1) Do $\ell = 1$. Initiate the values $\boldsymbol{\beta}^{(0)}$, $\boldsymbol{u}^{(0)}$ and $\boldsymbol{\varphi}^{(0)}$.

(B.2) Run the *Algorithm A.* by using $\boldsymbol{\varphi}^{(\ell)}$ as known value of the vector of variance components, and $\boldsymbol{\beta}^{(\ell-1)}$ and $\mathsf{u}^{(\ell-1)}$ as initial values. Let $\boldsymbol{\beta}^{(\ell)}$ and $\mathsf{u}^{(\ell)}$ be the obtained estimates and predictors.

(B.3) Calculate $\boldsymbol{\eta}_{dt}^{(\ell)} = \boldsymbol{X}_{dt}\boldsymbol{\beta}^{(\ell)} + \mathsf{Z}_{1dt}\mathsf{u}_1^{(\ell)} + \mathsf{Z}_{2dt}\mathsf{u}_2^{(\ell)}$, $d = 1, \ldots, D$, $t = 1, \ldots, T$. Calculate

$$
\begin{aligned}
p_{dkt}^{(\ell)} &= \frac{\exp(\eta_{dkt}^{(\ell)})}{1 + \sum_{k=1}^{q-1} \exp(\eta_{dkt}^{(\ell)})}, \ \boldsymbol{p}_{dt}^{(\ell)} = \underset{1\le k\le q-1}{\text{col}} (p_{dkt}^{(\ell)}), \ \boldsymbol{W}^{(\ell)} = \underset{1\le d\le D}{\text{diag}} (\boldsymbol{W}_d), \\
\boldsymbol{W}_d &= \underset{1\le t\le T}{\text{diag}} (\nu_{dt}[\text{diag}(\boldsymbol{p}_{dt}^{(\ell)}) - \boldsymbol{p}_{dt}^{(\ell)}\boldsymbol{p}_{dt}^{(\ell)\prime}])), \ \boldsymbol{\Sigma}_{u_r}^{(\ell)} = \text{diag}\big(\varphi_{r1}^{(\ell)}\boldsymbol{I}_{A_r}, \ldots, \varphi_{rq-1}^{(\ell)}\boldsymbol{I}_{A_r}\big), \\
\mathsf{T}_r^{(\ell)} &= (\mathsf{Z}_r'\boldsymbol{W}\mathsf{Z}_r + \boldsymbol{\Sigma}_{u_r}^{(\ell)-1})^{-1}, \ \boldsymbol{V}^{(\ell)} = \mathsf{Z}\boldsymbol{\Sigma}_u^{(\ell)}\mathsf{Z}' + \boldsymbol{W}^{(\ell)-1}, \\
\mathsf{T}_r^{rml(\ell)} &= \mathsf{T}_r^{(\ell)} + \mathsf{T}_r^{(\ell)}\mathsf{Z}_r'\boldsymbol{W}^{(\ell)}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{V}^{(\ell)-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{W}^{(\ell)}\mathsf{Z}_r\mathsf{T}_r^{(\ell)}, \ \tau_{rk}^{(\ell)} = \big(\varphi_{rk}^{(\ell)}\big)^{-1}\text{tr}\big(\mathsf{T}_{rkk}^{rml(\ell)}\big)
\end{aligned}
$$

(B.4) Update φ_{rk} using the equations

$$
\widehat{\varphi}_{rk}^{(\ell+1)} = \frac{\widehat{\boldsymbol{u}}_{rk}^{(\ell)\prime}\widehat{\boldsymbol{u}}_{rk}^{(\ell)}}{A_r - \tau_{rk}^{(\ell)}}, \quad r = 1, 2, \ k = 1, \ldots, q-1.
$$

(B.5) Repeat the steps (B.2)-(B.4) until the convergence of $\boldsymbol{\beta}^{(\ell)}$, $\boldsymbol{u}^{(\ell)}$ (or $\mathsf{u}^{(\ell)}$) and $\boldsymbol{\varphi}^{(\ell)}$.

Like in the previous chapter, the difference between this algoritms is that Algorithm B (alternative) is a fixed-point algorithm and the Algorithm B is an iterative Fisher-Scoring algorithm.

The above described algorithms requires initial values for $\boldsymbol{\beta}$, $\boldsymbol{u}$ and $\boldsymbol{\varphi}$. A simple possibility is $\boldsymbol{u}^{(0)} = \boldsymbol{0}$ and $\boldsymbol{\beta}^{(0)} = \tilde{\boldsymbol{\beta}}$, where $\tilde{\boldsymbol{\beta}}$ is obtained by fitting the non mixed variant of the model (5.1.1)-(5.1.2) without the random effect $\boldsymbol{u}$. The non mixed model is also used for calculating $\boldsymbol{\varphi}^{(0)}$ by means of the formula

$$
\hat{\varphi}_{rk} = \frac{1}{2(D-1)} \sum_{d=1}^{D} \sum_{t=1}^{T} (\tilde{\eta}_{dkt}^{(dir)} - \tilde{\eta}_{dkt})^2, \quad k = 1, \ldots, q-1, r = 1, 2, \tag{5.2.3}
$$

where

$$
\tilde{\eta}_{dkt} = \tilde{\beta}_k x_{dkt}, \quad \tilde{\eta}_{dkt}^{(dir)} = \log \frac{y_{dkt}}{y_{dqt}}, \quad d = 1, \ldots, D, \ k = 1, \ldots, q-1, \ t = 1, \ldots, T.
$$

Under regularity conditions the asymptotic distribution of the REML-PQL estimator $\hat{\boldsymbol{\beta}}$ is multivariate normal $N(\boldsymbol{\beta}, \boldsymbol{F}^{\beta\beta})$, where $\boldsymbol{F}^{\beta\beta} = (q_{rr})_{r=1,\ldots,l_k}$ is the block sub-matrix of the Fisher information matrix in the output of Algorithm A. Therefore, an approximate $(1-\alpha)$-level confidence interval for β_{kr} is

$$
\hat{\beta}_{kr} \pm z_{\alpha/2} q_{rr}, \quad r = 1, \ldots, l_k,
$$

where z_α is the α-quantile of the normal distribution $N(0,1)$. If we use $\hat{\beta}_{kr}$ to test $H_0: \beta_{kr} = 0$ and we observe the realization $\hat{\beta}_{kr} = \beta_0$, the approximate p-value is

$$
p = 2P_{H_0}(\hat{\beta}_{kr} > |\beta_0|) = 2P(Z > |\beta_0| / \sqrt{q_{rr}})
$$

where Z follows a standard normal distribution.

5.3 Model-based small area estimation

Let us consider the population quantity $\boldsymbol{m}_{0dt} = N_{dt}p_{dt}$, where $N_{dt} = \#(P_{dt})$. In practice N_{dt} is unknown and it is estimated by combining administrative registers and population projections models, so we are more rigorously interested in estimating $\boldsymbol{m}_{dt} = N_{dt}p_{dt}$. In the case of the SLFS, we use $N_{dt} = \hat{N}_{dt}^{dir}$ because the sample weights are calibrated to population projections by sex and age groups at the province level. We estimate $\boldsymbol{m}_{dt}$ by means of $\hat{\boldsymbol{m}}_{dt} = N_{dt}\hat{\boldsymbol{p}}_{dt}$, where

$$\hat{\boldsymbol{p}}_{dt} = \underset{1\leq k\leq q-1}{\text{col}}(\hat{p}_{dkt}), \quad \hat{p}_{dkt} = \frac{\exp\{\hat{\eta}_{dkt}\}}{1+\sum_{\ell=1}^{q-1}\exp\{\hat{\eta}_{d\ell t}\}}, \quad \hat{\eta}_{dkt} = \boldsymbol{x}_{dkt}\hat{\boldsymbol{\beta}}_k + \hat{u}_{1,dk} + \hat{u}_{2,dkt},$$

and $\hat{\boldsymbol{\beta}}_k$, $\hat{u}_{1,dk}$ and $\hat{u}_{2,dkt}$ are obtained from the output of the fitting algorithm A. We are further interested in estimating the domain totals

$$Y_{dkt} = \sum_{j\in P_{dt}} y_{dktj}, \quad d = 1,\dots,D,\ k = 1,\dots,q-1,\ t = 1,\dots,T.$$

A synthetic model-based estimator of Y_{dkt} is $\hat{Y}_{dkt} = \hat{m}_{dkt} = \hat{N}_{dt}\hat{p}_{dkt}$. Estimates of rates can be obtained by plugging the corresponding estimators of totals.

For deriving an approximation to the mean squared error of $\hat{m}_{dkt}$, we treat $\hat{N}_{dt}$ as a known constant. Let us write

$$m_{dkt} = h_{dkt}(\boldsymbol{\eta}_{dt}) = \hat{N}_{dt}p_{dkt} = \hat{N}_{dt}\frac{\exp\{\eta_{dkt}\}}{1+\sum_{\ell=1}^{q-1}\exp\{\eta_{d\ell t}\}}.$$

The partial derivatives of h_{dkt} are

$$\frac{\partial h_{dkt}}{\partial \eta_{dkt}} = \hat{N}_{dt}p_{dkt}(1-p_{dkt}), \qquad \frac{\partial h_{dk_1t}}{\partial \eta_{dk_2t}} = -\hat{N}_{dt}p_{dk_1t}p_{dk_2t},\ k_1 \neq k_2.$$

We define

$$h(\boldsymbol{\eta}) = (h_1(\boldsymbol{\eta}),\dots,h_{q-1}(\boldsymbol{\eta}))', \quad h_{dt}(\boldsymbol{\eta}_{dt}) = (h_{d1t}(\boldsymbol{\eta}_{dt}),\dots,h_{dqt-1}(\boldsymbol{\eta}_{dt}))',$$

$$\boldsymbol{H}_{dt} = \left(\frac{\partial h_{dk_1t}}{\partial \eta_{dk_2t}}\right)_{k_1,k_2=1,\dots,q-1} = \hat{N}_{dt}[\text{diag}(\boldsymbol{p}_{dt}) - \boldsymbol{p}_{dt}\boldsymbol{p}_{dt}'].$$

A Taylor expansion of $h_{dkt}(\hat{\boldsymbol{\eta}}_{dt})$ around $\boldsymbol{\eta}_{dt}$ yields to

$$h_{dkt}(\hat{\boldsymbol{\eta}}_{dt}) - h_{dkt}(\boldsymbol{\eta}_{dt}) \approx \sum_{\ell=1}^{q-1}\frac{\partial h_{dkt}}{\partial \eta_{d\ell t}}(\hat{\boldsymbol{\eta}}_{d\ell t} - \boldsymbol{\eta}_{d\ell t}), \quad h_{dt}(\hat{\boldsymbol{\eta}}_{dt}) - h_{dt}(\boldsymbol{\eta}_{dt}) \approx \boldsymbol{H}_{dt}(\hat{\boldsymbol{\eta}}_{dt} - \boldsymbol{\eta}_{dt}).$$

In matrix notation, we have

$$h(\hat{\boldsymbol{\eta}}) - h(\boldsymbol{\eta}) \approx \boldsymbol{H}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}), \quad \boldsymbol{H} = \underset{1\leq d_1\leq D}{\text{diag}}\,(\underset{1\leq t_1\leq m_{d_1}}{\text{diag}}\,(\boldsymbol{H}_{d_1t_1})).$$

As $\hat{\boldsymbol{m}}_d = \boldsymbol{A}\boldsymbol{H}\hat{\boldsymbol{\eta}}$, with $\boldsymbol{A} = \underset{1\leq d_1\leq D}{\text{col}'}(\underset{1\leq t_1\leq m_{d_1}}{\text{col}'}(\delta_{tt_1}\delta_{dd_1}\boldsymbol{I}_{q-1}))$, $\hat{\boldsymbol{\eta}}$ can be viewed as a vector of EBLUPs in the lineal mixed model (5.2.2), we propose applying the methodology of

Prasad and Rao (1990) to approximate the MSE of $\hat{\boldsymbol{m}}_{dt}$. For this sake, we first define the variance matrices

$$\begin{aligned}
\boldsymbol{V} &= \boldsymbol{Z}\boldsymbol{V}_u\boldsymbol{Z}' + \boldsymbol{W}^{-1}, \quad \boldsymbol{W} = \underset{1\le d\le D}{\text{diag}}\,(\boldsymbol{W}_d), \quad \boldsymbol{W}_d = \underset{1\le t\le T}{\text{diag}}\,(\boldsymbol{W}_{dt}),\\
\boldsymbol{V}_u &= var(\boldsymbol{u}) = \text{diag}(\boldsymbol{V}_{u_1}, \boldsymbol{V}_{u_2}),\ \boldsymbol{V}_{u_1} = \underset{1\le d\le D}{\text{diag}}\,(\boldsymbol{V}_{u_1d}), \quad \boldsymbol{V}_{u_1d} = \underset{1\le k\le q-1}{\text{diag}}\,(\varphi_{1k}),\\
\boldsymbol{V}_{u_2} &= \underset{1\le d\le D}{\text{diag}}\,(\boldsymbol{V}_{u_2d}), \quad \boldsymbol{V}_{u_2d} = \underset{1\le t\le T}{\text{diag}}\,(\boldsymbol{V}_{u_2dt}), \quad \boldsymbol{V}_{u_2dt} = \underset{1\le k\le q-1}{\text{diag}}\,(\varphi_{2k}).
\end{aligned}$$

The MSE is approximated by

$$MSE(\hat{m}_{dkt}) \approx \mathcal{G}_1(\boldsymbol{\varphi}) + \mathcal{G}_2(\boldsymbol{\varphi}) + \mathcal{G}_3(\boldsymbol{\varphi}),$$

where

$$\begin{aligned}
\mathcal{G}_1(\boldsymbol{\varphi}) &= \boldsymbol{A}\boldsymbol{H}\boldsymbol{Z}\boldsymbol{T}\boldsymbol{Z}'\boldsymbol{H}'\boldsymbol{A}',\\
\mathcal{G}_2(\boldsymbol{\varphi}) &= [\boldsymbol{A}\boldsymbol{H}\boldsymbol{X} - \boldsymbol{A}\boldsymbol{H}\boldsymbol{Z}\boldsymbol{T}\boldsymbol{Z}'\boldsymbol{W}\boldsymbol{X}]\boldsymbol{Q}[\boldsymbol{X}'\boldsymbol{H}'\boldsymbol{A}' - \boldsymbol{X}'\boldsymbol{W}\boldsymbol{Z}\boldsymbol{T}\boldsymbol{Z}'\boldsymbol{H}'\boldsymbol{A}'],\\
\mathcal{G}_3(\boldsymbol{\varphi}) &\approx \sum_{k_1=1}^{2(q-1)}\sum_{k_2=1}^{2(q-1)} \text{cov}(\hat{\varphi}_{k_1}, \hat{\varphi}_{k_2})\boldsymbol{A}\boldsymbol{H}\boldsymbol{L}^{(k_1)}\boldsymbol{V}\boldsymbol{L}^{(k_2)\prime}\boldsymbol{H}'\boldsymbol{A}',
\end{aligned}$$

$\varphi_k = \varphi_{1k}$ if $k \le q-1$, $\varphi_k = \varphi_{2k}$ otherwise and

$$\boldsymbol{Q} = (\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}, \boldsymbol{T} = \boldsymbol{V}_u - \boldsymbol{V}_u\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{Z}\boldsymbol{V}_u,$$

$$\boldsymbol{L}^{(k)} = (\boldsymbol{I} - \boldsymbol{R}_1)\boldsymbol{W}_k\boldsymbol{V}^{-1}, \quad \boldsymbol{W}_k = \frac{\partial \boldsymbol{V}}{\partial \varphi_k}, \quad \boldsymbol{R}_1 = \boldsymbol{Z}_1\boldsymbol{V}_{u1}\boldsymbol{Z}_1'\boldsymbol{V}^{-1}, \quad k = 1, \ldots, q-1,$$

$$\boldsymbol{L}^{(k)} = (\boldsymbol{I} - \boldsymbol{R}_2)\boldsymbol{W}_k\boldsymbol{V}^{-1}, \quad \boldsymbol{W}_k = \frac{\partial \boldsymbol{V}}{\partial \varphi_k}, \quad \boldsymbol{R}_2 = \boldsymbol{V}_{u2}\boldsymbol{V}^{-1}, \quad k = q, \ldots, 2(q-1).$$

The covariance $\text{cov}(\hat{\varphi}_{k_1}, \hat{\varphi}_{k_2})$ is obtained from the inverse of the Fisher information matrix $\boldsymbol{F}$ at the output of the algorithm B. The analytic MSE estimator is

$$mse(\hat{m}_{dkt}) = \mathcal{G}_1(\hat{\boldsymbol{\varphi}}) + \mathcal{G}_2(\hat{\boldsymbol{\varphi}}) + 2\mathcal{G}_3(\hat{\boldsymbol{\varphi}}).$$

where $\hat{\boldsymbol{\varphi}}$ is also obtained from algorithm B.

The elements of the formula $\mathcal{G}_1(\boldsymbol{\varphi}) = \boldsymbol{A}\boldsymbol{H}\boldsymbol{Z}\boldsymbol{T}\boldsymbol{Z}'\boldsymbol{H}'\boldsymbol{A}'$ are

$$\begin{aligned}
\boldsymbol{A} &= \underset{1\le d_1\le D}{\text{col}'}\big(\underset{1\le t_1\le m_{d_1}}{\text{col}'}\,(\delta_{tt_1}\delta_{dd_1}\boldsymbol{I}_{q-1})\big), \quad \boldsymbol{H} = \underset{1\le d_1\le D}{\text{diag}}\big(\underset{1\le t_1\le m_{d_1}}{\text{diag}}\,(\boldsymbol{H}_{d_1t_1})\big),\\
\boldsymbol{T} &= \boldsymbol{V}_u - \boldsymbol{V}_u\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{Z}\boldsymbol{V}_u = \begin{pmatrix} \boldsymbol{T}_{11} & \boldsymbol{T}_{12}\\ \boldsymbol{T}_{21} & \boldsymbol{T}_{22}\end{pmatrix}, \ \boldsymbol{T}_{12} = -\boldsymbol{V}_{u_1}\boldsymbol{Z}_1'\boldsymbol{V}^{-1}\boldsymbol{Z}_2\boldsymbol{V}_{u_2} = \boldsymbol{T}_{21}',\\
\boldsymbol{T}_{11} &= \boldsymbol{V}_{u_1} - \boldsymbol{V}_{u_1}\boldsymbol{Z}_1'\boldsymbol{V}^{-1}\boldsymbol{Z}_2\boldsymbol{V}_{u_1}, \quad \boldsymbol{T}_{22} = \boldsymbol{V}_{u_2} - \boldsymbol{V}_{u_2}\boldsymbol{Z}_2'\boldsymbol{V}^{-1}\boldsymbol{Z}_2\boldsymbol{V}_{u_2}.
\end{aligned}$$

We have $\boldsymbol{T}_{ab} = \underset{1\le d_1\le D}{\text{diag}}(\boldsymbol{T}_{abd})$, $a,b=1,2$, where

$$
\begin{aligned}
\boldsymbol{T}_{11d} &= \underset{1\le k\le q-1}{\text{diag}}(\varphi_{1k}) - \underset{1\le k\le q-1}{\text{diag}}(\varphi_{1k})\boldsymbol{Z}'_{1d}\boldsymbol{V}_d^{-1}\boldsymbol{Z}_{1d}\underset{1\le k\le q-1}{\text{diag}}(\varphi_{1k}),\\
\boldsymbol{T}_{12d} &= -\underset{1\le k\le q-1}{\text{diag}}(\varphi_{1k})\boldsymbol{Z}'_{1d}\boldsymbol{V}_d^{-1}\boldsymbol{Z}_{1d}\underset{1\le t_1\le m_{d_1}}{\text{diag}}(\underset{1\le k\le q-1}{\text{diag}}(\varphi_{2k})),\\
\boldsymbol{T}_{22d} &= \underset{1\le t_1\le m_{d_1}}{\text{diag}}(\underset{1\le k\le q-1}{\text{diag}}(\varphi_{2k})) - \underset{1\le t_1\le m_{d_1}}{\text{diag}}(\underset{1\le k\le q-1}{\text{diag}}(\varphi_{2k}))\boldsymbol{Z}'_{2d}\boldsymbol{V}_d^{-1}\boldsymbol{Z}_{2d}\underset{1\le t_1\le m_{d_1}}{\text{diag}}(\underset{1\le k\le q-1}{\text{diag}}(\varphi_{2k})).
\end{aligned}
$$

We calculate this product $\boldsymbol{ZTZ}'$

$$
\boldsymbol{ZTZ}' = \boldsymbol{Z}_1\boldsymbol{T}_{11}\boldsymbol{Z}'_1 + \boldsymbol{Z}_1\boldsymbol{T}_{12}\boldsymbol{Z}'_2 + \boldsymbol{Z}_2\boldsymbol{T}_{21}\boldsymbol{Z}'_1 + \boldsymbol{Z}_2\boldsymbol{T}_{22}\boldsymbol{Z}'_2 = \boldsymbol{M}_{11} + \boldsymbol{M}_{12} + \boldsymbol{M}'_{12} + \boldsymbol{M}_{22}.
$$

We have $\boldsymbol{M}_{ab} = \underset{1\le d_1\le D}{\text{diag}}(\boldsymbol{M}_{abd_1})$, where $\boldsymbol{M}_{abd_1} = \boldsymbol{Z}_{ad_1}\boldsymbol{T}_{abd_1}\boldsymbol{Z}'_{bd_1}$, $a,b=1,2$. Finally

$$
\mathcal{G}_1(\boldsymbol{\varphi}) = \boldsymbol{AH}[\boldsymbol{M}_{11} + \boldsymbol{M}_{12} + \boldsymbol{M}'_{12} + \boldsymbol{M}_{22}]\boldsymbol{H}'\boldsymbol{A}' = \boldsymbol{G}_{11} + \boldsymbol{G}_{12} + \boldsymbol{G}'_{12} + \boldsymbol{G}_{22},
$$

where

$$
\begin{aligned}
\boldsymbol{G}_{11} &= \boldsymbol{H}_{dt}\big[\underset{1\le k\le q-1}{\text{diag}}(\varphi_{1k}) - \underset{1\le k\le q-1}{\text{diag}}(\varphi_{1k})\boldsymbol{Z}'_{1d}\boldsymbol{V}_d^{-1}\boldsymbol{Z}_{1d}\underset{1\le k\le q-1}{\text{diag}}(\varphi_{1k})\big]\boldsymbol{H}'_{dt},\\
\boldsymbol{G}_{12} &= -\boldsymbol{H}_{dt}\underset{1\le k\le q-1}{\text{diag}}(\varphi_{1k})\boldsymbol{Z}'_{1d}\boldsymbol{V}_d^{-1}\boldsymbol{Z}_{1d}\underset{1\le t_1\le m_{d_1}}{\text{diag}}(\underset{1\le k\le q-1}{\text{diag}}(\varphi_{2k}))\underset{1\le t_1\le m_{d_1}}{\text{col}}(\delta_{tt_1}\boldsymbol{H}_{dt_1}),\\
\boldsymbol{G}_{22} &= \underset{1\le t_1\le m_{d_1}}{\text{col}'}(\delta_{tt_1}\boldsymbol{H}_{dt_1})\big[\underset{1\le t_1\le m_{d_1}}{\text{diag}}(\underset{1\le k\le q-1}{\text{diag}}(\varphi_{2k})) - \underset{1\le t_1\le m_{d_1}}{\text{diag}}(\underset{1\le k\le q-1}{\text{diag}}(\varphi_{2k}))\boldsymbol{Z}'_{2d}\boldsymbol{V}_d^{-1}\boldsymbol{Z}_{2d}\\
&\quad\cdot \underset{1\le t_1\le m_{d_1}}{\text{diag}}(\underset{1\le k\le q-1}{\text{diag}}(\varphi_{2k}))\big]\underset{1\le t_1\le m_{d_1}}{\text{col}}(\delta_{tt_1}\boldsymbol{H}_{dt_1}).
\end{aligned}
$$

The expression of $\mathcal{G}_2(\boldsymbol{\varphi})$ is

$$
\mathcal{G}_2(\boldsymbol{\varphi}) = [\boldsymbol{AHX} - \boldsymbol{AHZTZ}'\boldsymbol{WX}]\boldsymbol{Q}[\boldsymbol{X}'\boldsymbol{H}'\boldsymbol{A}' - \boldsymbol{X}'\boldsymbol{WZTZ}'\boldsymbol{H}'\boldsymbol{A}'] = [\boldsymbol{A}_{21} - \boldsymbol{A}_{22}]\boldsymbol{Q}[\boldsymbol{A}'_{21} - \boldsymbol{A}'_{22}],
$$

where

$$
\begin{aligned}
\boldsymbol{A}_{21} &= \boldsymbol{AHX} = \underset{1\le d_1\le D}{\text{col}'}(\underset{1\le t_1\le m_{d_1}}{\text{col}'}(\delta_{dd_1}\delta_{tt_1}\boldsymbol{I}_{q-1}))\underset{1\le d_1\le D}{\text{diag}}(\underset{1\le t_1\le m_{d_1}}{\text{diag}}(\boldsymbol{H}_{d_1t_1}))\\
&\quad\cdot \underset{1\le d_1\le D}{\text{col}}(\underset{1\le t_1\le m_{d_1}}{\text{col}}(\boldsymbol{X}_{d_1dt_1})) = \boldsymbol{I}_{q-1}\boldsymbol{H}_{dt}\boldsymbol{X}_{dt} = \boldsymbol{H}_{dt}\boldsymbol{X}_{dt},\\
\boldsymbol{A}_{22} &= \boldsymbol{AHZTZ}'\boldsymbol{WX} = \underset{1\le d_1\le D}{\text{col}'}(\underset{1\le t_1\le m_{d_1}}{\text{col}'}(\delta_{dd_1}\delta_{tt_1}\boldsymbol{I}_{q-1}\boldsymbol{H}_{d_1t_1}))\\
&\quad\cdot \underset{1\le d_1\le D}{\text{diag}}(\boldsymbol{M}_{11d_1} + \boldsymbol{M}_{12d_1} + \boldsymbol{M}_{21d_1} + \boldsymbol{M}_{22d_1})\underset{1\le d_1\le D}{\text{col}}(\underset{1\le t_1\le m_{d_1}}{\text{col}}(\delta_{dd_1}\delta_{tt_1}\boldsymbol{W}_{d_1t_1}\boldsymbol{X}_{d_1dt_1})\\
&= \underset{1\le t_1\le m_{d_1}}{\text{col}'}(\delta_{tt_1}\boldsymbol{H}_{dt_1}))(\boldsymbol{M}_{11d} + \boldsymbol{M}_{12d} + \boldsymbol{M}_{21d} + \boldsymbol{M}_{22d})\underset{1\le t_1\le m_{d_1}}{\text{col}}(\boldsymbol{W}_{dt_1}\boldsymbol{X}_{dt_1}).
\end{aligned}
$$

For the calculation of $\mathcal{G}_3(\boldsymbol{\varphi})$, we have

$$
\boldsymbol{L}^{(k)} = (\boldsymbol{I} - \boldsymbol{R}_1)\boldsymbol{W}_k\boldsymbol{V}^{-1}, \quad \boldsymbol{W}_k = \frac{\partial \boldsymbol{V}}{\partial \varphi_k}, \quad \boldsymbol{R}_1 = \boldsymbol{Z}_1\boldsymbol{V}_{u1}\boldsymbol{Z}'_1\boldsymbol{V}^{-1}, \quad k = 1,\dots,q-1.
$$

$$\boldsymbol{L}^{(k)} = (\boldsymbol{I} - \boldsymbol{R}_2)\boldsymbol{W}_k\boldsymbol{V}^{-1}, \quad \boldsymbol{W}_k = \frac{\partial \boldsymbol{V}}{\partial \varphi_k}, \quad \boldsymbol{R}_2 = \boldsymbol{V}_{u2}\boldsymbol{V}^{-1}, \quad k = q, \dots, 2(q-1).$$

Concerning the estimation of the MSE of $\hat{m}_{dkt}$, we can also use the approach of González-Manteiga et al. (2008a, 2008b) by introducing the following parametric bootstrap method.

1. Fit the model (5.1.1)-(5.1.2) and calculate $\hat{\varphi}_{1k}$, $\hat{\varphi}_{2k}$ and $\hat{\boldsymbol{\beta}}_k$, $k = 1, \dots, q-1$.
2. For $d = 1, \dots, D$, $t = 1, \dots, T$, generate the random effects $\boldsymbol{u}^*_{1,d} \sim N(\boldsymbol{0}, \underset{1\le k\le q-1}{\text{diag}} (\hat{\varphi}_{1k}))$ and $\boldsymbol{u}^*_{2,dt} \sim N(\boldsymbol{0}, \underset{1\le k\le q-1}{\text{diag}} (\hat{\varphi}_{2k}))$, and the target variable $\boldsymbol{y}^*_{dt} \sim \text{M}(\nu_{dt}, p^*_{d1t}, \dots, p^*_{dqt-1})$, where
$$p^*_{dkt} = \frac{\exp\{\eta^*_{dkt}\}}{1 + \sum_{\ell=1}^{q-1}\exp\{\eta^*_{d\ell}\}}, \quad \eta^*_{dkt} = \hat{\boldsymbol{\beta}}_k x_{dkt} + u^*_{1,dk} + u^*_{2,dkt}, \quad m^*_{dkt} = \hat{N}_{dt}p^*_{dkt}.$$
3. For $d = 1, \dots, D$, $k = 1, \dots, q-1$, $t = 1 \dots, T$, calculate $\hat{\varphi}^*_{1k}$, $\hat{\varphi}^*_{2k}$, $\hat{\boldsymbol{\beta}}^*_k$,
$$\hat{p}^*_{dkt} = \frac{\exp\{\hat{\eta}^*_{dkt}\}}{1 + \sum_{\ell=1}^{q-1}\exp\{\hat{\eta}^*_{d\ell}\}}, \quad \hat{\eta}^*_{dkt} = \hat{\boldsymbol{\beta}}^*_k x_{dkt} + \hat{u}^*_{1,dk} + \hat{u}^*_{2,dkt}, \quad \hat{m}^*_{dkt} = \hat{N}_{dt}\hat{p}^*_{dkt}.$$
4. Repeat B times steps 2-3 and calculate the bootstrap mean square error estimator
$$mse^{*1}_{dkt} = \frac{1}{B}\sum_{b=1}^{B}(\hat{m}^*_{dkt} - m^*_{dkt})^2, \quad d = 1, \dots, D,\ k = 1, \dots, q-1,\ t = 1 \dots, T.$$

We also propose two other estimators. The second bootstrap MSE estimator is based on the analytic one and follows the ideas of bagging from Breiman (1996). The bootstrap approximation of $MSE(\hat{m}_{dkt})$ is

$$mse^{*2}(\hat{m}_{dkt}) = E_*[\mathcal{G}_1(\hat{\boldsymbol{\varphi}}^*) + \mathcal{G}_2(\hat{\boldsymbol{\varphi}}^*) + 2\mathcal{G}_3(\hat{\boldsymbol{\varphi}}^*)],$$

with the Monte Carlo approximation

$$mse^{*2}(\hat{m}_{dkt}) = \frac{1}{B}\sum_{b=1}^{B}(\mathcal{G}^{*b}_1(\hat{\boldsymbol{\varphi}}^*) + \mathcal{G}^{*b}_2(\hat{\boldsymbol{\varphi}}^*) + 2\mathcal{G}^{*b}_3(\hat{\boldsymbol{\varphi}}^*)).$$

5.4 Simulation study

In this section we present two simulation experiments. The first experiment is designed to analyze the behavior of the estimators $\hat{\boldsymbol{\beta}}_k$, $\hat{\varphi}_{1k}$, $\hat{\varphi}_{2k}$ and $\hat{m}_{dkt} = \hat{N}_{dt}p_{dkt}$. The second simulation studies the behavior of the proposed MSE estimators.

5.4.1 Sample simulation

We take $\hat{N}_{dt} = 1000$ and we consider a multinomial logit mixed model with three model categories $(q-1=2)$. For $d=1,\dots,D$, $k=1,2$ and $t=1,\dots,T$, we generate the explanatory variables

$$U_{dkt} = \frac{1}{3}\left(\frac{d-D}{D} + \frac{k}{q-1} + \frac{t}{T}\right),\ x_{d1t} = \mu_1 + \sigma_{x11}^{1/2} U_{d1t},\ x_{d2t} = \mu_2 + \sigma_{x22}^{1/2}\big[\rho_x U_{d1} + \sqrt{1-\rho_x^2}\, U_{d2t}\big],$$

where $\mu_1 = \mu_2 = 1$, $\sigma_{x11} = 1$, $\sigma_{x22} = 2$ and $\rho_x = 0$. The random effects are $u_{1,dk} \sim N(0, \varphi_{1k})$ with $\varphi_{11} = 1$, $\varphi_{12} = 2$ and $u_{2,dkt} \sim N(0, \varphi_{2k})$ with $\varphi_{21} = 0.25$, $\varphi_{22} = 0.5$. The target variable is $\boldsymbol{y}_{dt} = (y_{d1t}, y_{d2t})' \sim \mathrm{M}(\nu_{dt}, p_{d1t}, p_{d2t})$, where

$$p_{dkt} = \frac{\exp\{\eta_{dkt}\}}{1 + \exp\{\eta_{d1t}\} + \exp\{\eta_{d2t}\}},\ \eta_{dkt} = \beta_{0k} + \beta_{1k} x_{dkt} + u_{1,dk} + u_{2,dkt}, \tag{5.4.1}$$

$\nu_{dt} = 100$, $\beta_{01} = 1.3$, $\beta_{02} = -1$, $\beta_{11} = -1.6$ and $\beta_{12} = 1$. We can write the model (5.4.1) in the matrix form

$$\left(\begin{array}{c} \eta_{111} \\ \eta_{121} \\ \hline \eta_{112} \\ \eta_{122} \\ \hline \vdots \\ \hline \eta_{D11} \\ \eta_{D21} \\ \hline \eta_{D12} \\ \eta_{D22} \end{array}\right) = \left(\begin{array}{cccc} 1 & x_{111} & 0 & 0 \\ 0 & 0 & 1 & x_{121} \\ \hline 1 & x_{112} & 0 & 0 \\ 0 & 0 & 1 & x_{122} \\ \hline \vdots & & & \\ \hline 1 & x_{D11} & 0 & 0 \\ 0 & 0 & 1 & x_{D21} \\ \hline 1 & x_{D12} & 0 & 0 \\ 0 & 0 & 1 & x_{D22} \end{array}\right) \begin{pmatrix} \beta_{01} \\ \beta_{11} \\ \beta_{02} \\ \beta_{12} \end{pmatrix} + \left(\begin{array}{c} u_{1,11} \\ u_{1,12} \\ \hline u_{1,11} \\ u_{1,12} \\ \hline \vdots \\ \hline u_{1,D1} \\ u_{1,D2} \\ \hline u_{1,D1} \\ u_{1,D2} \end{array}\right) + \left(\begin{array}{c} u_{2,111} \\ u_{2,121} \\ \hline u_{2,112} \\ u_{2,122} \\ \hline \vdots \\ \hline u_{2,D11} \\ u_{2,D21} \\ \hline u_{2,D12} \\ u_{2,D22} \end{array}\right),$$

or, more concisely, in the form

$$\begin{pmatrix} \boldsymbol{\eta}_{11} \\ \boldsymbol{\eta}_{12} \\ \vdots \\ \boldsymbol{\eta}_{D1} \\ \boldsymbol{\eta}_{D2} \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}_{11} \\ \boldsymbol{X}_{12} \\ \vdots \\ \boldsymbol{X}_{D1} \\ \boldsymbol{X}_{D2} \end{pmatrix} \boldsymbol{\beta} + \boldsymbol{Z}_1 \boldsymbol{u}_1 + \boldsymbol{Z}_2 \boldsymbol{u}_2,$$

where $\boldsymbol{Z}_1 = \underset{1\le d\le D}{\operatorname{diag}}(\boldsymbol{Z}_{1d})$, $\boldsymbol{Z}_2 = \boldsymbol{I}_{2D}$,

$$\boldsymbol{\eta}_{dt} = \begin{pmatrix} \eta_{d1t} \\ \eta_{d2t} \end{pmatrix},\ \boldsymbol{X}_{dt} = \begin{pmatrix} 1 & x_{d1t} & 0 & 0 \\ 0 & 0 & 1 & x_{d2t} \end{pmatrix},\ \boldsymbol{I}_{2D} = \begin{pmatrix} \boldsymbol{I}_2 & & 0 \\ & \ddots & \\ 0 & & \boldsymbol{I}_2 \end{pmatrix},\ \boldsymbol{Z}_{1d} = \begin{pmatrix} \boldsymbol{Z}_{1d1} \\ \boldsymbol{Z}_{1d2} \end{pmatrix},$$

$$\boldsymbol{Z}_{1dt} = \boldsymbol{I}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},\ \boldsymbol{u}_1 \sim N(0, \boldsymbol{\varphi}_1 \boldsymbol{I}_{2D}),\ \boldsymbol{u}_{2dt} \sim N(0, \boldsymbol{\varphi}_2 \boldsymbol{I}_2).$$

5.4.2 Simulation experiment 1

The objective of this experiment is to analyze the behavior of the estimators of $\boldsymbol{\beta}_k$, $\boldsymbol{\varphi}_{1k}$, $\boldsymbol{\varphi}_{2k}$ and m_{dkt}. As efficiency measures, we use the relative empirical bias (RBIAS) and the relative mean squared error (RMSE). The simulation is described below.

1. Repeat $I = 1000$ times ($i = 1, \ldots, 1000$)

    1.1. For $d = 1, \ldots, D$, $k = 1, 2$, $t = 1, \ldots, T$, generate (y_{dkt}, x_{dkt}).

    1.2 For $d = 1, \ldots, D$, $k = 1, 2$, $t = 1, \ldots, T$, $j = 0, 1$, $r = 1, 2$, calculate $\hat{\beta}_{jk}^{(i)}$,$\hat{\varphi}_{rk}^{(i)}$, $\hat{m}_{dkt}$.

2. For $j = 0, 1$, $k = 1, 2$, $r = 1, 2$, $(d, t) = (1, 1), (D/2, T), (D, T)$, calculate

$$BIAS(\hat{\beta}_{jk}) = \frac{1}{I}\sum_{i=1}^{I}(\hat{\beta}_{jk}^{(i)} - \beta_{jk}),\ BIAS(\hat{\varphi}_{rk}) = \frac{1}{I}\sum_{i=1}^{I}(\hat{\varphi}_{rk}^{(i)} - \varphi_k),$$

$$MSE(\hat{\beta}_{jk}) = \frac{1}{I}\sum_{i=1}^{I}(\hat{\beta}_{jk}^{(i)} - \beta_k)^2,\ MSE(\hat{\varphi}_{rk}) = \frac{1}{I}\sum_{i=1}^{I}(\hat{\varphi}_{rk}^{(i)} - \varphi_k)^2,$$

$$BIAS_{dkt} = \frac{1}{I}\sum_{i=1}^{I}(\hat{m}_{dkt}^{(i)} - m_{dkt}^{(i)}),\ MSE_{dkt} = \frac{1}{I}\sum_{i=1}^{I}(\hat{m}_{dkt}^{(i)} - m_{dkt}^{(i)})^2,$$

$$RBIAS(\hat{\beta}_{jk}) = \frac{BIAS(\hat{\beta}_{jk})}{|\beta_{jk}|},\ RBIAS(\hat{\varphi}_{rk}) = \frac{BIAS(\hat{\varphi}_{rk})}{\varphi_k},$$

$$RMSE(\hat{\beta}_{jk}) = \frac{\sqrt{MSE(\hat{\beta}_{jk})}}{|\beta_{jk}|},\ RMSE(\hat{\varphi}_{rk}) = \frac{\sqrt{MSE(\hat{\varphi}_{rk})}}{\varphi_k},$$

$$MEAN_{dkt} = \frac{1}{I}\sum_{i=1}^{I} m_{dkt}^{(i)},\ RBIAS_{dkt} = \frac{BIAS_{dkt}}{|MEAN_{dkt}|},\ RMSE_{dkt} = \frac{\sqrt{MSE_{dkt}}}{|MEAN_{dkt}|}.$$

Table 5.4.1 and Table 5.4.2 present the RMSE-values of the model parameter estimators for $T = 2$ time periods and for $D = 100$ domains respectively. In the first case ($T = 2$), as D increases from 50 to 200 we observe a reduction in RMSE of approximately 50% for the $\boldsymbol{\beta}_k$'s and around 40% for $\boldsymbol{\varphi}_1$, but there is no reduction in $\boldsymbol{\varphi}_2$. In the second case ($D = 100$), as T increases from 2 to 8, Table 5.4.2 shows that RMSE decreases by approximately 70% for $\boldsymbol{\varphi}_2$, but there is no reduction for $\boldsymbol{\varphi}_1$. This simulation suggests that the proposed multinomial mixed model should be used with more than two time periods.

D	50	100	200	D	50	100	200
$RMSE(\hat{\beta}_{01})$	0.61	0.46	0.33	$RBIAS(\hat{\beta}_{01})$	-0.04	0.01	-0.04
$RMSE(\hat{\beta}_{02})$	0.63	0.47	0.33	$RBIAS(\hat{\beta}_{02})$	0.05	0.00	0.05
$RMSE(\hat{\beta}_{11})$	0.62	0.44	0.32	$RBIAS(\hat{\beta}_{11})$	0.05	0.06	0.06
$RMSE(\hat{\beta}_{12})$	0.62	0.43	0.31	$RBIAS(\hat{\beta}_{12})$	-0.04	-0.05	-0.05
$RMSE(\hat{\varphi}_{11})$	0.24	0.17	0.12	$RBIAS(\hat{\varphi}_{11})$	-0.05	-0.04	-0.03
$RMSE(\hat{\varphi}_{12})$	0.23	0.17	0.14	$RBIAS(\hat{\varphi}_{12})$	-0.07	-0.08	-0.08
$RMSE(\hat{\varphi}_{21})$	0.56	0.60	0.62	$RBIAS(\hat{\varphi}_{21})$	-0.54	-0.60	-0.62
$RMSE(\hat{\varphi}_{22})$	0.51	0.55	0.57	$RBIAS(\hat{\varphi}_{22})$	-0.49	-0.54	-0.57

Table 5.4.1: RMSE and RBIAS for $T = 2$.

T	2	4	8	T	2	4	8
$RMSE(\hat{\beta}_{01})$	0.43	0.30	0.20	$RBIAS(\hat{\beta}_{01})$	-0.07	-0.05	-0.03
$RMSE(\hat{\beta}_{02})$	0.45	0.31	0.21	$RBIAS(\hat{\beta}_{02})$	0.08	-0.06	0.03
$RMSE(\hat{\beta}_{11})$	0.42	0.28	0.21	$RBIAS(\hat{\beta}_{11})$	0.09	0.06	0.07
$RMSE(\hat{\beta}_{12})$	0.41	0.28	0.20	$RBIAS(\hat{\beta}_{12})$	0.08	-0.05	-0.05
$RMSE(\hat{\varphi}_{11})$	0.16	0.16	0.14	$RBIAS(\hat{\varphi}_{11})$	-0.01	-0.01	-0.01
$RMSE(\hat{\varphi}_{12})$	0.18	0.18	0.16	$RBIAS(\hat{\varphi}_{12})$	-0.1	-0.11	-0.09
$RMSE(\hat{\varphi}_{21})$	0.60	0.31	0.16	$RBIAS(\hat{\varphi}_{21})$	-0.59	-0.31	-0.15
$RMSE(\hat{\varphi}_{22})$	0.56	0.30	0.17	$RBIAS(\hat{\varphi}_{22})$	-0.55	-0.29	-0.16

Table 5.4.2: RMSE and RBIAS for $D = 100$.

Table 5.4.3 and Table 5.4.4 present the RMSE and RBIAS values of $\hat{m}_{dkt}$ for $T = 2$ and $D = 100$ respectively. We observe that all the RMSE values are below 14%, which indicates a good behavior.

	D	50	100	200		D	50	100	200
	(d, t)					(d, t)			
$RMSE_{d1t}$	(1,1)	0.09	0.10	0.10	$RBIAS_{d1t}$	(1,1)	-0.0027	0.0001	-0.0044
	$(D/2, 2)$	0.14	0.13	0.12		$(D/2, 2)$	0.0087	0.0032	-0.0012
	$(D, 2)$	0.14	0.14	0.14		$(D, 2)$	-0.0032	0.0007	0.0076
$RMSE_{d2t}$	(1,1)	0,13	0,13	0,13	$RBIAS_{d2t}$	(1,1)	-0.0043	-0.0004	0.0060
	$(D/2, 2)$	0.09	0.10	0.10		$(D/2, 2)$	-0.0187	-0.0050	0.0001
	$(D, 2)$	0.08	0.09	0.08		$(D, 2)$	-0.0049	0.0023	-0.0032

Table 5.4.3: $RMSE_{dkt}$ and $RBIAS_{dkt}$ for $T = 2$.

	T	2	4	8		T	2	4	8
	(d,t)					(d,t)			
$RMSE_{d1t}$	(1,1)	0.10	0.09	0.08	$RBIAS_{d1t}$	(1,1)	0.001	0.004	0.003
	$(D/2,T)$	0.14	0.12	0.13		$(D/2,T)$	-0.002	-0.0003	0.002
	(D,T)	0.14	0.12	0.12		(D,T)	-0.003	-0.0003	0.002
$RMSE_{d2t}$	(1,1)	0.13	0.12	0.13	$RBIAS_{d1t}$	(1,1)	-0.003	-0,004	-0,001
	$(D/2,T)$	0.09	0.08	0.09		$(D/2,T)$	0.001	-0.0005	-0.0004
	(D,T)	0.09	0.08	0.09		(D,T)	0.004	-0.0005	-0.0004

Table 5.4.4: $RMSE_{dkt}$ and $RBIAS_{dkt}$ for $D = 100$.

5.4.3 Simulation experiment 2

The second simulation experiment is designed to study the behavior of the three mean square error estimators (analytic and bootstrap). In this case we take $D = 50$ and $T = 2$. The steps of the simulation are

1. Repeat $I = 500$ times ($i = 1,\dots,500$)

   1.1. For $d = 1,\dots,50$, $k = 1,2$, $t = 1,2$, generate $(y_{dkt}^{(i)}, x_{dkt}^{(i)})$.

   1.2. For $d = 1,\dots,D$, $k = 1,2$, $t = 1,2$, calculate $\hat{p}_{dkt}^{(i)}$, $\hat{m}_{dkt}^{(i)}$, $\hat{\boldsymbol{\varphi}}^{(i)}$, $\hat{\boldsymbol{\beta}}^{(i)}$ and

$$mse_{dkt}^{(i)} = \mathcal{G}_{1dkt}^{(i)}(\hat{\boldsymbol{\varphi}}^{(i)}) + \mathcal{G}_{2dkt}^{(i)}(\hat{\boldsymbol{\varphi}}^{(i)}) + 2\mathcal{G}_{dkt}^{(i)}(\hat{\boldsymbol{\varphi}}^{(i)}).$$

   1.3. Repeat $B = 500$ times ($b = 1,\dots,B$)

   1.3.1. For $d = 1,\dots,D$, $k = 1,2$, $t = 1,2$, generate $\boldsymbol{u}_{1,d}^{*(ib)}$, $\boldsymbol{u}_{2,dt}^{*(ib)}$,

$$\boldsymbol{y}_{dt}^{*(ib)} = (y_{1dkt}^{*(ib)}, y_{2dkt}^{*(ib)})' \sim \mathrm{M}(\nu_{dt}, p_{d1t}^{*(ib)}, p_{d2t}^{*(ib)}),$$

   where

$$p_{dkt}^{*(ib)} = \frac{\exp\{\eta_{dkt}^{*(ib)}\}}{1+\exp\{\eta_{d1t}^{*(ib)}\}+\exp\{\eta_{d2t}^{*(ib)}\}}, \quad \eta_{dkt}^{*(ib)} = \hat{\beta}_{0k}^{(i)} + \hat{\beta}_{1k}^{(i)} x_{dkt}^{(i)} + u_{1,dk}^{*(ib)} + u_{2,dkt}^{*(ib)}.$$

   1.3.2. For $d = 1,\dots,D$, $k = 1,2$, $t = 1,2$, calculate $\hat{\boldsymbol{\varphi}}^{*(ib)}$, $\hat{\boldsymbol{\beta}}^{*(ib)}$, $\hat{p}_{dkt}^{*(ib)}$, $\tilde{p}_{dkt}^{*(ib)}$, $\hat{m}_{dkt}^{*(ib)}$, $\tilde{m}_{dkt}^{*(ib)}$.

   1.4 For $d = 1,\dots,D$, $k = 1,2$, $t = 1,2$, calculate

$$mse_{dkt}^{*1(i)} = \frac{1}{B}\sum_{b=1}^{B}(\hat{m}_{dkt}^{*(ib)} - m_{dkt}^{*(ib)})^2,$$

$$mse_{dkt}^{*2(i)} = \frac{1}{B}\sum_{b=1}^{B}\left\{\mathcal{G}_{1dkt}^{(*ib)}(\hat{\varphi}_1^{*(ib)}, \hat{\varphi}_2^{*(ib)}) + \mathcal{G}_{2dkt}^{*(ib)}(\hat{\varphi}_1^{*(ib)}, \hat{\varphi}_2^{*(ib)}) + 2\mathcal{G}_{3dkt}^{*(ib)}(\hat{\varphi}_1^{*(ib)}, \hat{\varphi}_2^{*(ib)})\right\}.$$

2. Calculate

$$B^0_{dkt} = \frac{1}{I}\sum_{i=1}^{I}(mse^{(i)}_{dkt} - MSE_{dkt}),\ B^\ell_{dkt} = \frac{1}{I}\sum_{i=1}^{I}(mse^{*\ell(i)}_{dkt} - MSE_{dkt}),\ \ell = 1, 2,$$

$$E^0_{dkt} = \frac{1}{I}\sum_{i=1}^{I}(mse^{(i)}_{dkt} - MSE_{dkt})^2,\ E^\ell_{dkt} = \frac{1}{I}\sum_{i=1}^{I}(mse^{*\ell(i)}_{dkt} - MSE_{dkt})^2,\ \ell = 1, 2,$$

$$RB^\ell_{dkt} = \frac{B^{*\ell}_{dkt}}{MSE_{dkt}},\ RE^\ell_{dkt} = \frac{\sqrt{E^{*\ell}_{dkt}}}{MSE_{dkt}},\ \ell = 0, 1, 2,$$

where the MSE_{dkt} are taken from the output of the first simulation.

Figure 5.4.1 presents the box-plots of the values of the $mse^{(i)}_{dkt}, mse^{*\ell(i)}_{dkt}$, $k = 1, 2$, $\ell = 1, 2$, $i = 1, \ldots, 500$, in $(d, t) = (1, 1), (25, 2), (50, 2)$. The true MSE is plotted in a horizontal line. We observe that mse have the largest variability and that the estimators behaving best are mse^{*1} and mse^{*2}.

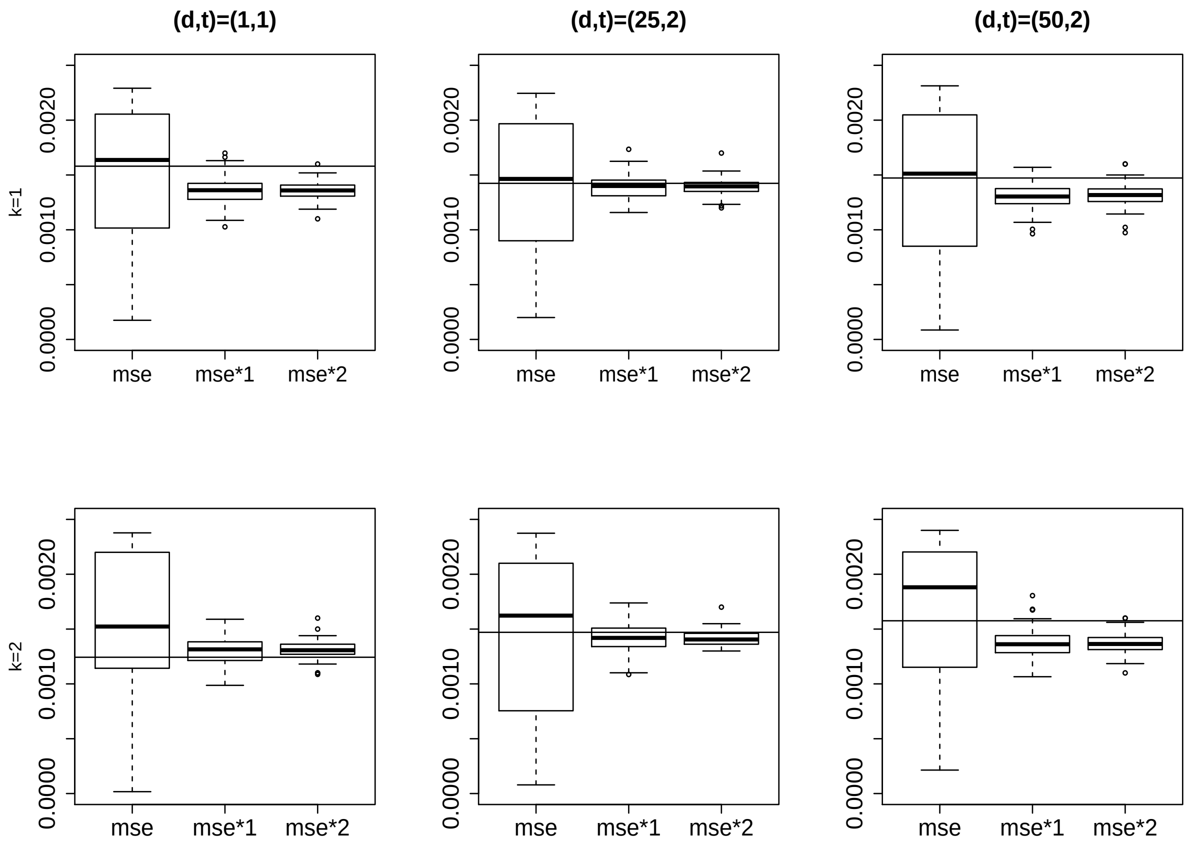


Figure 5.4.1: Boxplots of Simulation 2 for $(d, t) = (1, 1), (25, 2), (50, 2)$.

5.5 Application to real data

5.5.1 Data description

We are interested in estimating the totals of employed and unemployed people, and the unemployment rates per sex in the counties of Galicia. We deal with data from the SLFS of Galicia from the third quarter of 2009 to the fourth quarter of 2011.

As there are 51 counties in the SLFS of Galicia in this time period, we have $D = 102$ domains, denoted by P_{dt} at time t, and they are partitioned in the subsets P_{d1t}, P_{d2t} and P_{dt3} of employed, unemployed and inactive people. Our target population parameters are the totals of employed and unemployed people and the unemployment rate, this is to say

$$Y_{dkt} = \sum_{j \in P_{dkt}} y_{dktj}, \quad R_{dt} = \frac{Y_{d2t}}{Y_{d1t} + Y_{d2t}}, \qquad k = 1, 2,$$

where $y_{dktj} = 1$ if individual j of domain d at period t is in labour category k and $y_{dktj} = 0$ otherwise.

The SLFS does not produce official estimates at the domain level, but the analogous direct estimators of the total Y_{dkt}, the mean $\bar{Y}_{dkt} = Y_{dkt}/N_{dt}$, the size N_{dt} and the rate R_{dt} are

$$\hat{Y}_{dkt}^{dir} = \sum_{j \in S_{dt}} w_{dtj}\, y_{dktj}, \;\; \hat{\bar{Y}}_{dkt}^{dir} = \hat{Y}_{dkt}^{dir}/\hat{N}_{dt}^{dir}, \;\; \hat{N}_{dt}^{dir} = \sum_{j \in S_{dt}} w_{dtj}, \;\; \hat{R}_{dt}^{dir} = \frac{\hat{Y}_{d2t}^{dir}}{\hat{Y}_{d1t}^{dir} + \hat{Y}_{d2t}^{dir}}, \;\; k = 1, 2, \tag{5.5.1}$$

where S_{dt} is the domain sample at time period t and the w_{dtj}'s are the official calibrated sampling weights. The design-based covariance $\text{cov}_\pi(\hat{Y}_{dkt_1}^{dir}, \hat{Y}_{dkt_2}^{dir})$, $k_1, k_2 = 1, 2$, can be estimated by

$$\hat{\text{cov}}_\pi(\hat{Y}_{dkt_1}^{dir}, \hat{Y}_{dkt_2}^{dir}) = \sum_{j \in S_{dt}} w_{dtj}(w_{dtj} - 1)(y_{dkt_1 j} - \hat{\bar{Y}}_{dkt_1}^{dir})(y_{dkt_2 j} - \hat{\bar{Y}}_{dkt_2}^{dir}), \tag{5.5.2}$$

where the case $k_1 = k_2 = k$ denotes estimated variance, i.e. $\hat{V}_\pi(\hat{Y}_{dkt}^{dir}) = \hat{\text{cov}}_\pi(\hat{Y}_{dkt}^{dir}, \hat{Y}_{dkt}^{dir})$. The last formulas are obtained from Särndal et al. (1992), pp. 43, 185 and 391, with the simplifications $w_{dtj} = 1/\pi_{dtj}$, $\pi_{dtj,dtj} = \pi_{dtj}$ and $\pi_{dti,dtj} = \pi_{dti}\pi_{dtj}$, $i \neq j$ in the second order inclusion probabilities. The design-based variance of $\hat{R}_{dt}^{dir}$ can be approximated by Taylor linearization, i.e.

$$\hat{V}_\pi(\hat{R}_{dt}^{dir}) = \frac{(\hat{Y}_{d1t}^{dir})^2}{(\hat{Y}_{d1t}^{dir} + \hat{Y}_{d2t}^{dir})^4} \hat{V}_\pi(\hat{Y}_{d2t}^{dir}) + \frac{\hat{Y}_{d2t}^2}{(\hat{Y}_{d2t}^{dir} + \hat{Y}_{d1t}^{dir})^4} \hat{V}_\pi(\hat{Y}_{d1t}^{dir}) - \frac{2\hat{Y}_{d1t}^{dir}\hat{Y}_{d2t}^{dir}}{(\hat{Y}_{d1t}^{dir} + \hat{Y}_{d2t}^{dir})^4} \hat{\text{cov}}_\pi(\hat{Y}_{d1t}^{dir}, \hat{Y}_{d2t}^{dir}).$$

In the fourth quarter of 2011 the domain sample sizes lie all in the interval $(13, 1554)$, with median 97. Therefore, the direct estimates in (5.5.1) and (5.5.2) are not reliable and small area estimation methods are needed.

In this chapter we employ area-level models using auxiliary information from administrative registers. More concretely, we use as auxiliary variables the domain proportions of individuals within the categories of the following grouping variables.

- SEXAGE: Combinations of sex and age groups, with 6 values. SEX is coded 1 for men and 2 for women and AGE is categorized in 3 groups with codes 1 for 16-24, 2 for 25-54 and 3 for $\geq$55. The codes $1, 2, \ldots, 6$ are used for the pairs of *sex-age* $(1, 1), (1, 2), \ldots, (2, 3)$.
- STUD: This variable describes the achieved education level, with values 1-3 for the illiterate and the primary, the secondary and the higher education level respectively.
- REG: This variable indicates if an individual is registered or not as unemployed in the administrative register of employment claimants.
- SS: This variable indicates if an individual is registered or not in the social security system.

The last two variables (REG and SS) are the same as those used in the previous chapter. Table 5.5.1 shows the variability of the proportions of employed and unemployed people over the SEXAGE and STUD categories in the fourth quarter of 2011. Figure 5.5.1 shows the scatterplots of the log-rates of employed over inactive people against the proportions of people in social security system (left) and the log-rates of unemployed over inactive people against the proportions of people registered as unemployed (right). We observe that, despite the large variability observed in both plots, the log-rates of the two considered proportions seem to increase linearly with the proportions of people in the social security system and registered as unemployed respectively.

	SEXAGE						STUD		
	1	2	3	4	5	6	1	2	3
Employed	0.250	0.754	0.214	0.222	0.657	0.143	0.158	0.541	0.720
Unemployed	0.169	0.153	0.026	0.121	0.156	0.018	0.055	0.137	0.104

Table 5.5.1: Proportions of employed and unemployed people by SEXAGE and STUD categories in the fourth quarter of 2011.

Concerning Figure 5.5.1, we would like to say that the linearity of these clouds of dots can also be measure via the estimated Pearson correlations of variables. They are 0.78 (left) and 0.42 (right) respectively. Further, the corresponding 95% confidence intervals are (0.78, 0.81) and (0.36,0.47) respectively. Therefore, the proportions of people in the social security system and registered as unemployed could probably be good covariates for modeling the two probabilities. Indeed, after fitting the multinomial mixed model, tests of significance for the regression parameters and diagnosis of residuals confirmed the explanatory power of the auxiliary variables that were selected for each model.

5.5.2 Model estimation

We consider the multinomial mixed model (5.1.1)-(5.1.2) with $q = 3$ categories (employed, unemployed and inactive people) and we choose inactive people as reference (third) category. The multinomial size is $\nu_{dt} = n_{dt}$, where n_{dt} is the size of the domain sample S_{dt} in time t. The target variable is $\boldsymbol{y}_{dt} = (y_{d1t}, y_{d2t})'$, where y_{dkt} is the sample total

$$y_{dkt} = \sum_{j \in S_d} y_{dktj},$$

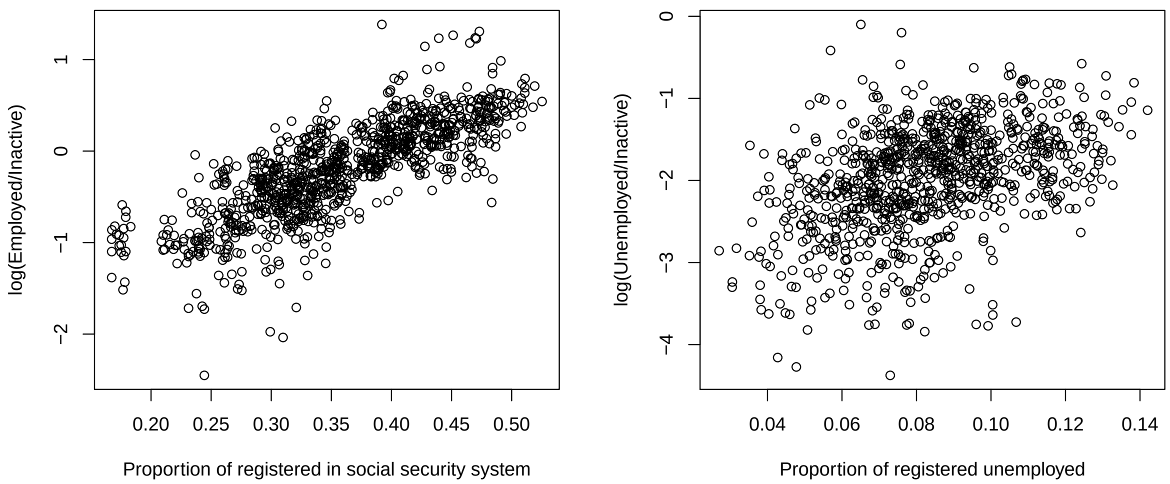


Figure 5.5.1: Proportions of employed (left) and unemployed (right) over inactive people versus the proportion of people registered as unemployed.

$y_{dktj} = 1$ if individual j is in category k $(k = 1, 2)$ and $y_{dktj} = 0$ otherwise.

The explanatory variables are the domain means of the indicators of the categories of SEXAGE, REG, SS and STUD. Their values have been taken from the SLFS and from administrative registers. The model is firstly fitted to the complete data set. An analysis of residuals is then carried out and six counties are marked as outliers. These six counties correspond with the counties of A Coruña, Eume, Ferrol, Noia, Pontevedra and Vigo. A Coruña, Ferrol, Pontevedra and Vigo are four of the most populous cities of Galicia where the relationships between the auxiliary variables SS and REG with the employment and unemployment status are typically weaker than in less populated counties. The model is finally fitted to reduced data set. The sample sizes of A Coruña, Ferrol, Pontevedra and Vigo are large enough to produce reliable direct estimates, then no model estimates are given for this counties and direct estimates are used for them. For Eume and Noia we use the synthetic estimator.

For each category, Table 5.5.1 and 5.5.2 presents the estimates of the regression parameters and their standard deviations. It also presents the p-values for testing the hypothesis $H_0 : \beta_{kr} = 0$. The estimates of the model variances and their standard deviations are presented at the bottom of Tables 5.5.1 and 5.5.2.

	Employed people			Unemployed people	
Variable	Estimate	$p-$value	Variable	Estimate	$p-$value
CONSTANT	-1.43	0.00	CONSTANT	-3.87	0.00
SEXAGE=1	0.92	0.02	SEXAGE=1	1.88	0.01
SEXAGE=2	2.05	0.00	SEXAGE=2	2.35	0.00
SEXAGE=3	0.15	0.38	SEXAGE=3	-0.48	0.11
SEXAGE=4	0.48	0.28	SEXAGE=4	1.57	0.07
SEXAGE=5	1.68	0.00	SEXAGE=5	1.51	0.00
STUD=1	-0.82	0.00	STUD=1	-0.49	0.12
SS	1.49	0.00	REG	11.84	0.00
σ	Estimate	Std.Dev.		Estimate	Std.Dev.
φ_1	0.031	0.0032		0.086	0.0186
φ_2	0.013	0.0002		0.104	0.011

Table 5.5.2: Model parameter estimates for the full data set.

	Employed people			Unemployed people	
Variable	Estimate	$p-$value	Variable	Estimate	$p-$value
CONSTANT	-1.48	0.00	CONSTANT	-3.84	0.00
SEXAGE=1	0.93	0.02	SEXAGE=1	1.83	0.02
SEXAGE=2	2.04	0.00	SEXAGE=2	2.33	0.00
SEXAGE=3	0.20	0.26	SEXAGE=3	-0.51	0.10
SEXAGE=4	0.79	0.08	SEXAGE=4	1.53	0.10
SEXAGE=5	1.69	0.00	SEXAGE=5	1.46	0.01
STUD=1	-0.82	0.00	STUD=1	-0.50	0.14
SS	1.59	0.00	REG	11.84	0.00
σ	Estimate	Std.Dev.		Estimate	Std.Dev.
φ_1	0.033	0.0055		0.091	0.017
φ_2	0.014	0.0021		0.108	0.012

Table 5.5.2: Model parameter estimates for the reduced data set.

Figure 5.5.2 plots the domain standardized residuals of models fitted to the full (left) and reduced (right) data set. The dots outside the interval (-3,3) correspond to the men and women counts in marked counties (A Coruña, Eume, Ferrol, Noia, Pontevedra and Vigo). The remaining model-based statistical analysis is carried out for the reduced data set.

Figure 5.5.3 plots the domain standardized residuals of employment (left) and unemployment (right) categories versus the proportions of people registered in the social security system (SS) and registered as unemployed (REG). The residuals are randomly distributed above and below zero and no rare pattern is observed. Therefore no diagnostics problems are found for the two main explanatory variables: SS and REG.

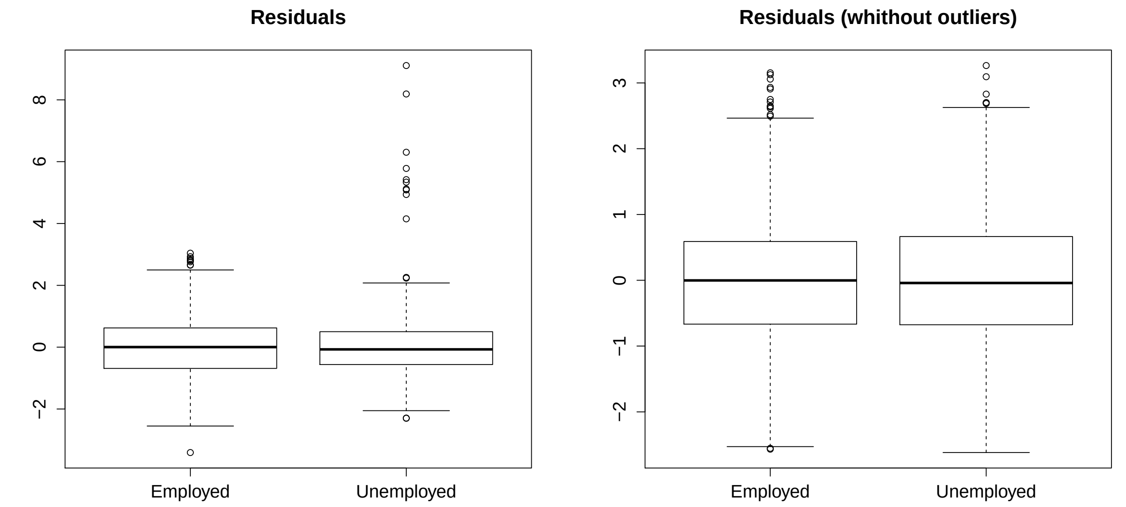


Figure 5.5.2: Boxplots of standardized residuals of models fitted to the full (left) and reduced (right) data set.

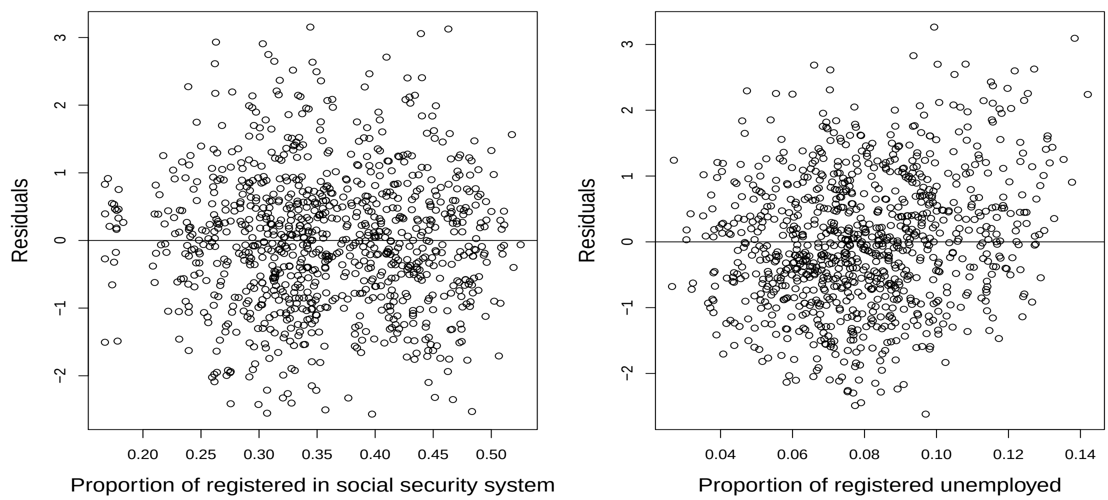


Figure 5.5.3: Domain standardized residuals of employment (left) and unemployment (right) categories versus proportions of people registered in the social security system (SS) and registered as unemployed (REG).

5.5.3 Model diagnostics

For carrying out the diagnosis of the model, we calculate the predicted sample totals $\hat{y}_{dkt} = n_{dt}\hat{p}_{dkt}$ and the domain residuals

$$r_{dkt} = \frac{y_{dkt} - \hat{y}_{dkt}}{y_{dkt}}, \quad d = 1, \dots, 102, \quad k = 1, 2, \quad t = 1, \dots, 4.$$

Figure 5.5.4 plots the domain residuals versus the predicted sample totals of employed and unemployed people. We observe that the model residuals are symmetrically situated above and below zero, so there is no prediction bias. Further, the variability of the residuals decreases as predicted employed or unemployed sample totals increase. This pattern is due to the fact that domains with greater amount of employed and unemployed people also have greater sample sizes. We also observe that there are no high residuals in absolute value or any other unusual pattern. Therefore, the fitted model seems to properly describe the data.

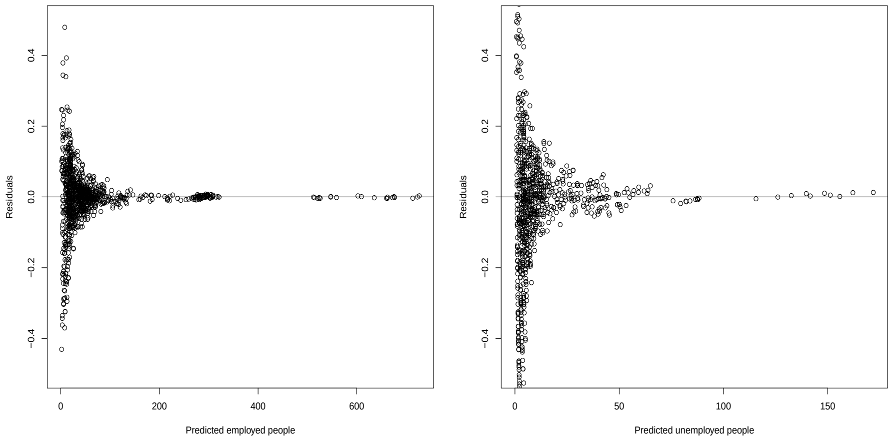


Figure 5.5.4: Domain residuals versus predicted sample totals

Figure 5.5.5 plots the direct estimates ($\hat{Y}_{dkt}^{dir}$) versus the model-based estimates ($\hat{m}_{dkt} = N_{dt}\hat{p}_{dkt}$) of the population totals of employed and unemployed people (in a logarithmic scale). We observe that both type of estimates behave quite similarly for employed people. This is because the population of employed people is quite large and there are plenty of sampled observations within this category. However, the direct and model-based estimates behave slightly different for unemployed people, which is due to the lower number of sampled observations within the category. We also observe that the model-based estimates are lower than the direct ones for large values of the direct estimates. This is a typical and desirable smoothing effect of model-based estimators.

5.5.4 Small area estimates and RMSE

Figures 5.5.6 and 5.5.7 plot the estimated women employment totals and women unemployment rates respectively for all quarters of 2011, with the counties sorted by sample size. We observe that the direct and the model-based estimators tend to be closer as soon as the sample size increases. The same pattern is observe in the case of men. For the sake of brevity we skip the corresponding figures.

Figures 5.5.8 and 5.5.9 plot the parametric bootstrap estimates (based on mse^{*1}) of the relative mean squared errors (RMSE) of the model-based estimators of totals of employed women and of women unemployment rates. The RMSEs of the corresponding

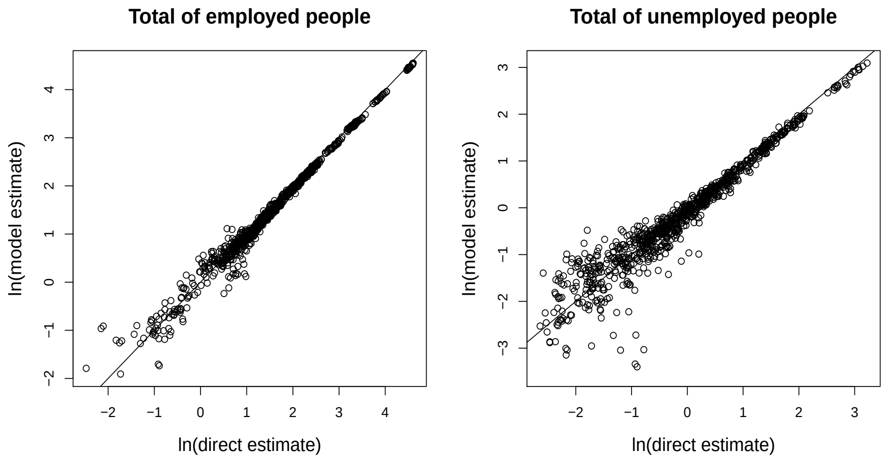


Figure 5.5.5: Direct versus model-based estimates.

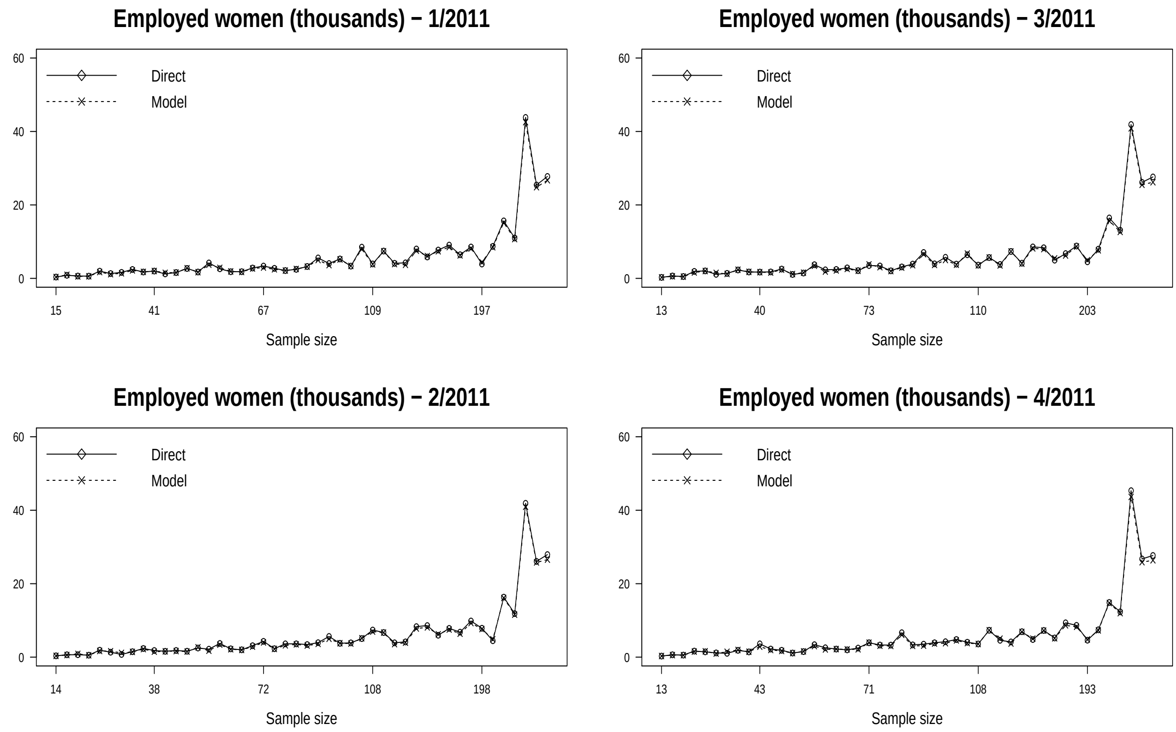


Figure 5.5.6: Direct (∘) and model-based (∗) estimates of totals of employed women in each quarter of 2011.

direct estimates are much higher than their model-based counterparts. This is the reason why they have not been plotted in Figure 5.5.9.

Tables 5.5.3 and 5.5.4 present some condensed numerical results for men and women respectively and for the fourth quarter of 2011. The tables has been constructed in two steps. We sort the domains by province and after that, in each province, we sort the domains by sample size, starting by the domain with smallest sample size. We present the

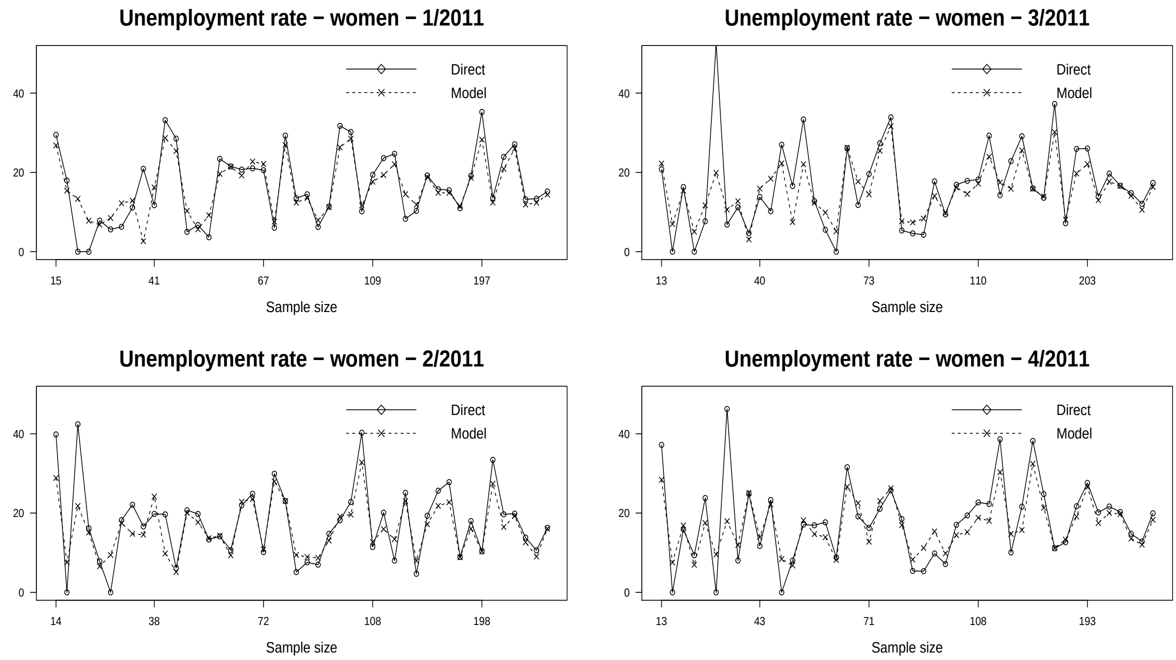


Figure 5.5.7: Direct (∘) and model-based (∗) estimates of women unemployment rates in each quarter of 2011.

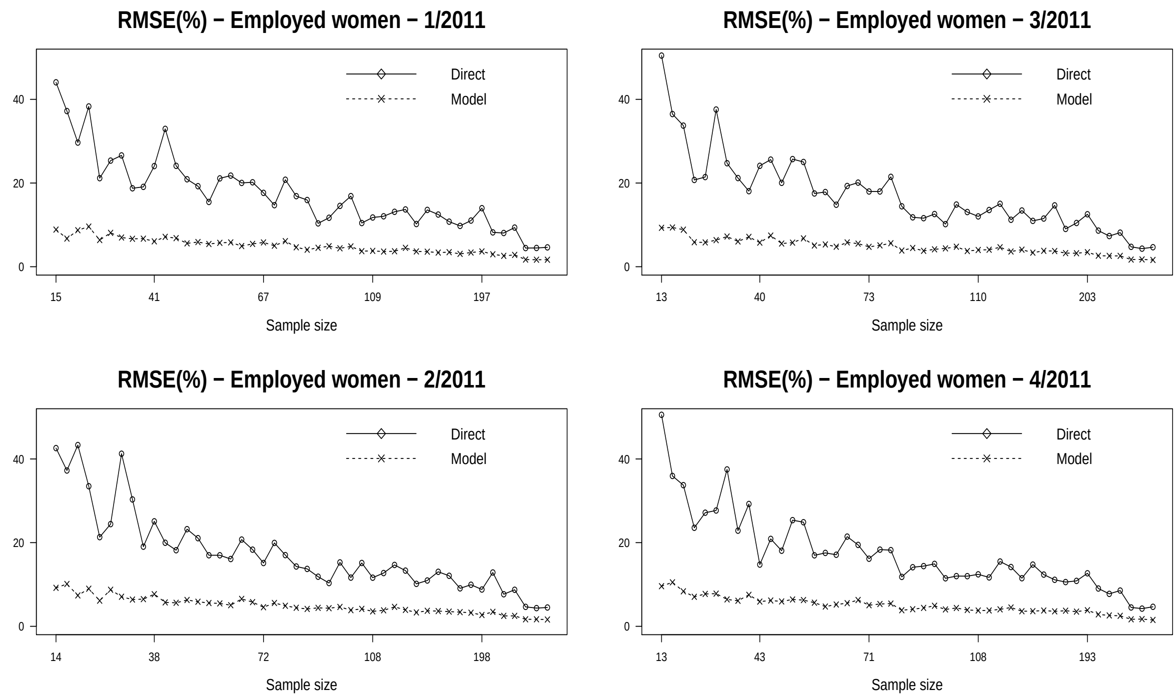


Figure 5.5.8: RMSEs of direct (∘) and model-based (∗) estimates of women employment totals in each quarter of 2011.

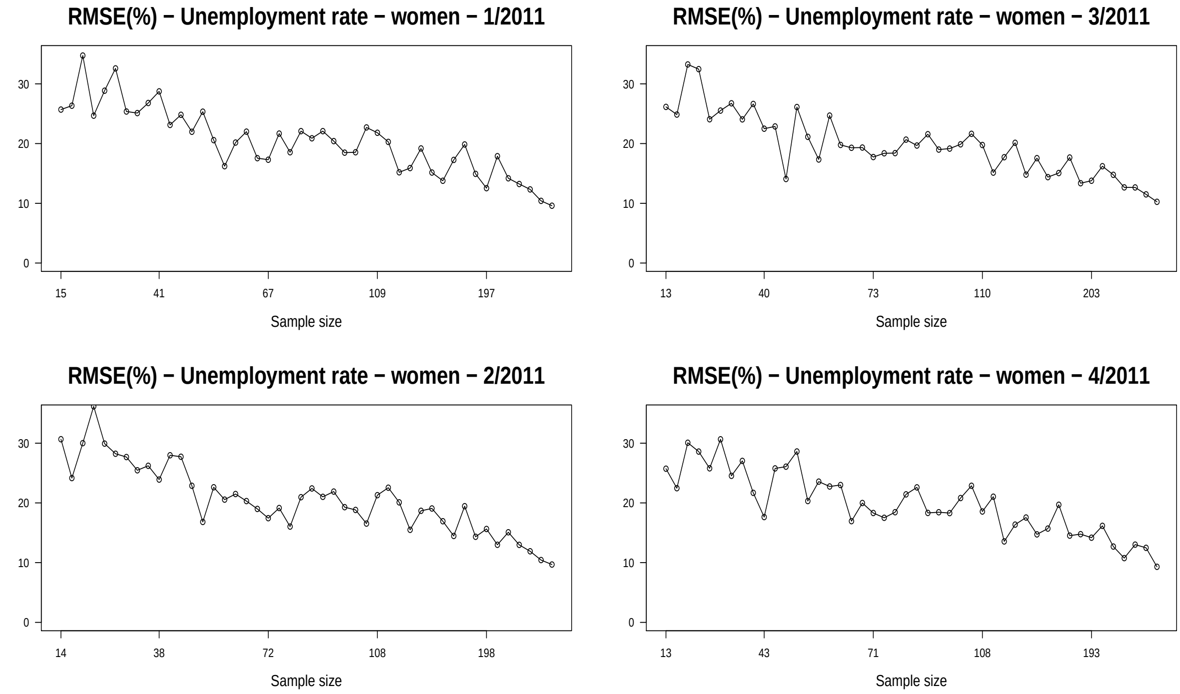


Figure 5.5.9: RMSEs of model-based estimates of women unemployment rates in each quarter of 2011.

results of the direct and the model-based estimates (labeled by n, dir, mod, re$_D$ and re$_M$ respectively) and the corresponding RMSE estimates for five domains in each province. The chosen domains correspond to the quintiles. The provinces are labeled by p and the sample sizes by n. Table 5.5.3 and 5.5.4 are partitioned in three vertical sections dealing with the estimation of totals of employed and unemployed people and with unemployment rates. The RMSEs of the model-based estimators are calculated by using the parametric bootstrap method 1 (mse^{*1}). The RMSEs of the corresponding direct estimates are estimated by using (5.5.2) and they are much greater than their counterparts. By observing the columns of RMSEs we conclude that model-based estimators are preferred to the direct ones. At this point it is good to keep in mind that the Office for National Statistics (ONS) in the United Kingdom considers that an estimate is publishable in the labour force statistics, and therefore official, if the coefficient of variation is less than 20% (ONS, 2004).

In domains d with $n_d = 0$, direct estimates cannot be calculated. In those cases, model-based estimates are calculated by using the synthetic part of the linear predictor, i.e. $\hat{\eta}_{dkt}^{synth} = \boldsymbol{x}_{dkt}\hat{\boldsymbol{\beta}}_k$. Tables 5.5.3 and 5.5.4 present blank spaces in domains where $y_{dk} = 0, k = 1, 2$.

		Employed people				Unemployed people				Unemployment rates			
p	n	dir	mod	re$_D$	re$_M$	dir	mod	re$_D$	re$_M$	dir	mod	re$_D$	re$_M$
1	21	2638	2151	12.44	11.02	155	273	97.82	28.67	5.55	15.95	201.87	27.81
1	37	2602	2518	18.18	7.54	328	299	68.98	27.07	11.21	14.32	129.74	25.52
1	56	3515	3406	16.31	8.00	1316	1104	32.56	20.30	27.25	12.44	34.83	18.85
1	100	8994	8746	8.96	6.05	948	921	43.61	17.61	9.54	9.01	78.16	13.22
1	503	48106	45676	3.90	2.86	8005	7094	15.30	11.68	14.27	3.65	21.03	7.30
2	12	410	176	40.10	11.21	93	57	95.60	31.24	18.62	25.61	146.41	26.72
2	62	2613	2524	14.78	6.39	206	285	69.15	26.30	7.32	11.67	174.27	25.96
2	79	4315	3962	11.28	6.62	623	590	44.38	19.01	12.63	10.80	70.06	18.50
2	95	4243	4054	11.06	5.67	447	371	50.17	19.59	9.54	9.98	96.01	18.74
2	592	28402	27214	3.91	2.83	4661	4444	14.14	8.97	14.10	3.49	21.18	8.63
3	17	394	408	51.03	11.27	406	33	50.45	30.19	50.73	25.48	31.04	28.59
3	58	2784	2863	16.55	6.84	687	596	45.51	21.33	19.80	12.74	56.39	19.51
3	58	1993	2088	21.43	6.41	733	595	39.38	23.94	26.90	14.83	44.93	22.62
3	107	4276	4137	13.54	4.97	1705	1551	25.21	19.63	28.51	10.74	26.48	19.00
3	184	6739	6231	9.74	5.30	2124	2006	21.46	13.92	23.97	7.67	25.11	11.76
4	47	2593	1188	18.64	7.57	572	522	50.66	22.79	18.09	17.28	71.93	21.67
4	94	6271	5650	8.84	6.26	383	533	57.27	17.26	5.76	10.45	137.13	18.80
4	138	8114	7767	9.89	4.80	2690	2453	21.75	16.95	24.90	8.40	24.44	14.23
4	262	17193	16390	6.58	4.04	5985	5392	14.59	13.03	25.82	5.75	15.68	13.38
4	413	24814	23767	4.93	3.53	6578	5600	14.06	9.94	20.95	4.56	16.08	7.65

Table 5.5.3: Estimated totals and rates with their RMSE's in the fourth quarter of 2011 for men.

		Employed people				Unemployed people				Unemployment rates			
p	n	dir	mod	re$_D$	re$_M$	dir	mod	re$_D$	re$_M$	dir	mod	re$_D$	re$_M$
1	20	1652	1512	23.55	7.05	170	113	97.23	35.59	9.37	20.60	208.55	28.60
1	32	1859	1957	22.86	6.10	162	265	97.99	30.45	8.04	17.34	241.60	27.03
1	56	3397	3004	16.97	5.65	691	517	48.55	25.41	16.92	14.39	71.25	23.57
1	79	3277	3051	18.21	5.44	1135	1086	36.36	21.09	25.72	12.41	41.00	18.44
1	149	7272	7219	12.36	3.78	2396	1958	27.05	17.59	24.78	9.77	29.87	15.70
2	13	291	311	50.55	9.59	172	123	65.95	33.19	37.20	26.66	58.99	25.73
2	54	1127	1105	25.37	6.41	98	80	98.45	26.72	8.03	15.28	267.95	28.61
2	97	3339	3068	14.12	4.09	190	277	70.05	23.29	5.40	11.51	234.42	22.60
2	107	4061	3839	11.99	3.91	977	687	37.24	24.25	19.39	10.09	42.52	22.87
2	643	26735	25852	4.24	1.73	3936	3533	15.64	12.98	12.84	4.10	25.65	12.48
3	14	644	555	35.94	10.52		45		30.07		27.62		22.46
3	54	1511	1576	24.87	6.27	311	350	72.13	23.33	17.09	17.05	101.72	20.31
3	69	2474	2137	19.46	6.29	586	620	43.60	21.69	19.17	14.78	67.89	19.98
3	108	3594	3586	12.42	3.78	1054	832	29.80	20.19	22.68	10.38	34.85	18.56
3	193	4616	4818	12.68	3.85	1760	1746	22.57	14.92	27.61	9.22	25.20	14.18
4	40	1428	1475	29.24	7.52	473	493	55.94	22.51	24.90	17.05	68.63	21.68
4	115	4586	5030	15.49	4.05	2887	2188	23.22	15.51	38.63	11.64	17.74	13.55
4	144	4803	5040	14.76	3.62	2968	2420	20.75	15.20	38.20	9.70	16.72	14.73
4	311	14887	14724	7.76	2.60	4113	3686	17.61	11.94	21.65	6.59	22.97	12.70
4	324	12342	11984	8.55	2.51	3139	2943	19.75	11.42	20.28	7.03	27.59	10.76

Table 5.5.4: Estimated totals and rates with their RMSE's in the fourth quarter of 2011 for women.

The Spanish Statistics Institute (INE) publishes LFS estimates of employed and unemployed totals at province level. In the case of extending these publications to the more disaggregated levels, the Statistical Offices might be interested in publishing data with the property that the sum of the estimated totals in all the domains within a province coincide with the official province total estimate. In order to fulfil this consistency criterion, we propose a modification of all the considered small area estimators for this model. Let $\widehat{Y}_{pt}^{dir}$ be the SLFS estimator of the total Y_{pt} of a variable y in the province p and the time period t. Assume that the province p is partitioned in D_p domains, labelled by $d = 1, \dots, D_p$. Let $\widehat{Y}_{p,1t}, \dots, \widehat{Y}_{p,D_pt}$ be some given estimators of the totals $Y_{p,1t}, \dots, Y_{p,D_pt}$ of the variable y in the domains $d = 1, \dots, D_p$ and the period t. In general, the consistency property

$$\widehat{Y}_{pt}^{dir} = \sum_{d=1}^{D_p} \widehat{Y}_{p,dt}$$

does not hold. In such cases, $\widehat{Y}_{p,1t}, \dots, \widehat{Y}_{p,D_pt}$ can be transformed into consistent estimators by

$$\widehat{Y}_{p,dt}^{c} = \lambda_{pt} \widehat{Y}_{p,dt}, \quad \lambda_{pt} = \frac{\widehat{Y}_{pt}^{dir}}{\sum_{d=1}^{D_p} \widehat{Y}_{p,dt}}.$$

Table 5.5.5 presents the direct and model-based estimator for employed and unemployed people at the province level for men (top) and women (bottom) for the fourth quarter of 2011 in the SLFS. In this table we can see also the consistency factors λ_p. We observe that the deviations from the SLFS estimation at province level are at most of 6% for employed people. However, the deviations from the SLFS estimation are bigger but under 14%.

			Employed people			Unemployed people		
Sex	Province	n	dir	mod	λ_p	dir	mod	λ_p
Men	1	2874	248516	234541	1.06	45969	40280	1.14
	2	1541	72917	68457	1.07	9307	9448	0.99
	3	1336	59824	57573	1.04	13883	12432	1.12
	4	2988	192842	181464	1.06	53477	47165	1.13
Women	1	3258	225489	220277	1.02	44696	41029	1.09
	2	1659	63163	60034	1.05	9304	8494	1.10
	3	1479	51726	50497	1.02	12418	11847	1.05
	4	3427	158415	157926	1.00	51303	48157	1.07

Table 5.5.5: Estimated men and women province totals for IV/2011.

Figure 5.5.10 maps the estimates of unemployment rates per sex in each county of Galicia for the fourth quarter of 2011. The colors are more intense in areas with higher unemployment rates. We observe that the counties of the west coast are those that, in general terms, have higher unemployment rates. That area is the most dynamic part of Galicia and the least aged. In that area live the 75% of the Galician population and the unemployment rates are also high because companies can not absorb as many workers. In these figures we also observe that the unemployment rates for women are higher than for men. More than half of the Galician counties have higher unemployment rates for women than for men. Even in 14 counties there are more than five points of difference. This leads us to conclude that we are still far from being close to equality in the labour market.

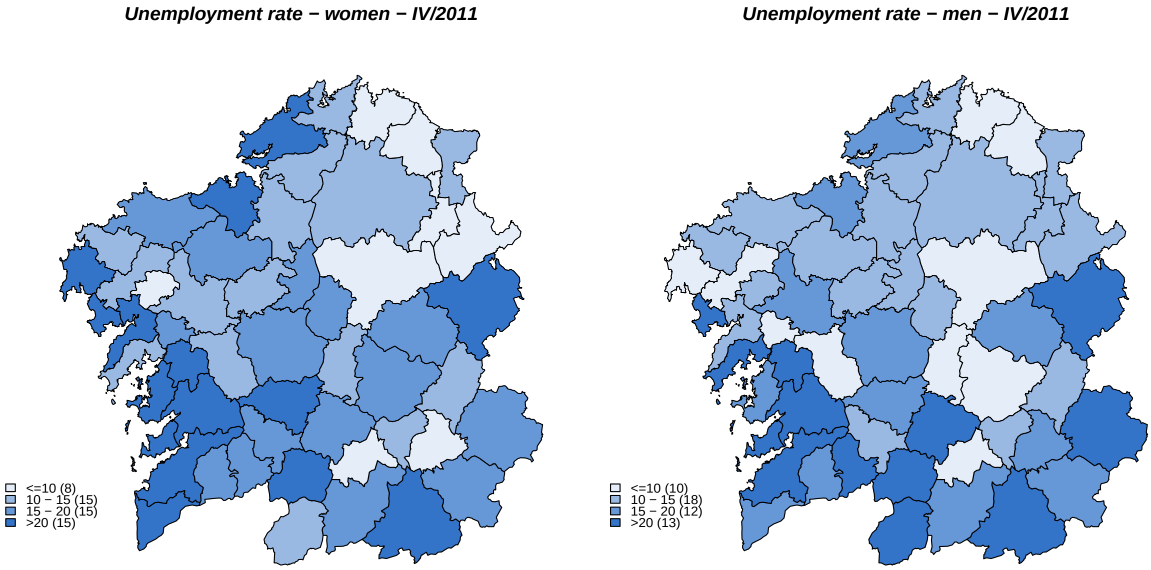


Figure 5.5.10: Estimates of women (left) and man (right) unemployment rates in Galician counties in the last quarter of 2011.

We further investigate if there are improvements in the estimates due to the inclusion of all the quarters in the model. Figure 5.5.11 compares the RMSEs of the unemployment rate estimates for the fourth quarter of 2011 based on the proposed model (Model A, defined by (5.1.1)-(5.1.2), using the samples of the ten time periods) with the corresponding ones based on the model that only uses the last time period (Model B). As the RMSEs are lower for the Model A, we conclude that it is worthwhile to use the past data to improve the estimates for the fourth quarter of 2011.

Another way of measuring the benefits of using Model A is to check the stability of the estimates along the ten time periods. For this sake we also consider a multinomial mixed model using data from only one time period, and we apply this model to the ten considered quarters (Model B) in the small domains. Figure 5.5.12 plots the women unemployment rates for six selected counties under the Model B and under the Model A. For this purpose, we select the domains that divide the sample size distribution below its median into six equal parts. In the considered counties, especially those with smaller sample size, we observe that the results of Model A are much more stable than the results of the ten separate models for each quarter. Stability is a property highly valued by the Statistical Offices when publishing the survey results. For Statistical Offices it is hard to justify that there are more than three points (in %) of difference between the unemployment rates of two consecutive quarters in a given county. Therefore, stability is an important property in public statistics.

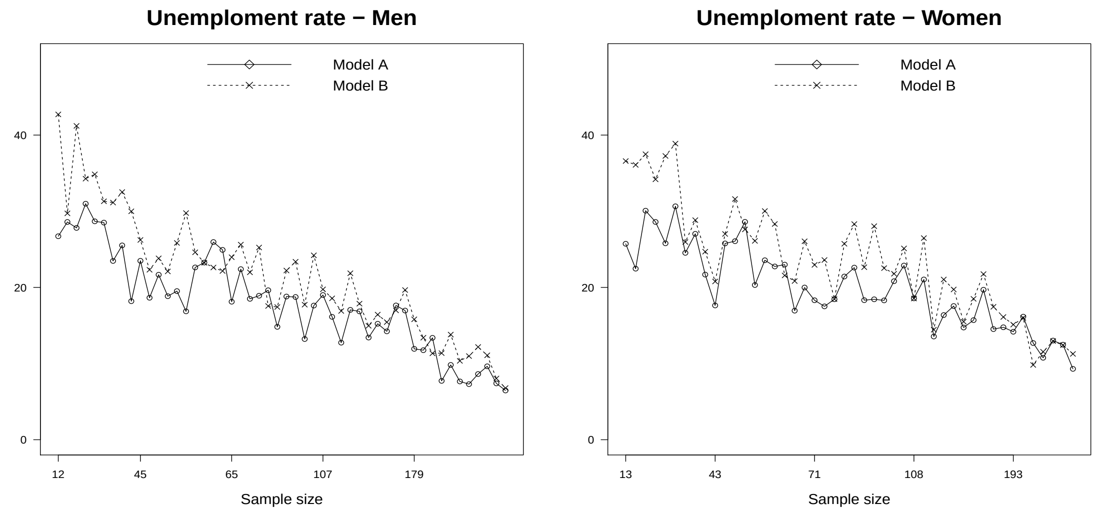


Figure 5.5.11: RMSEs of unemployment rates for models A and B in the fourth quarter of 2011

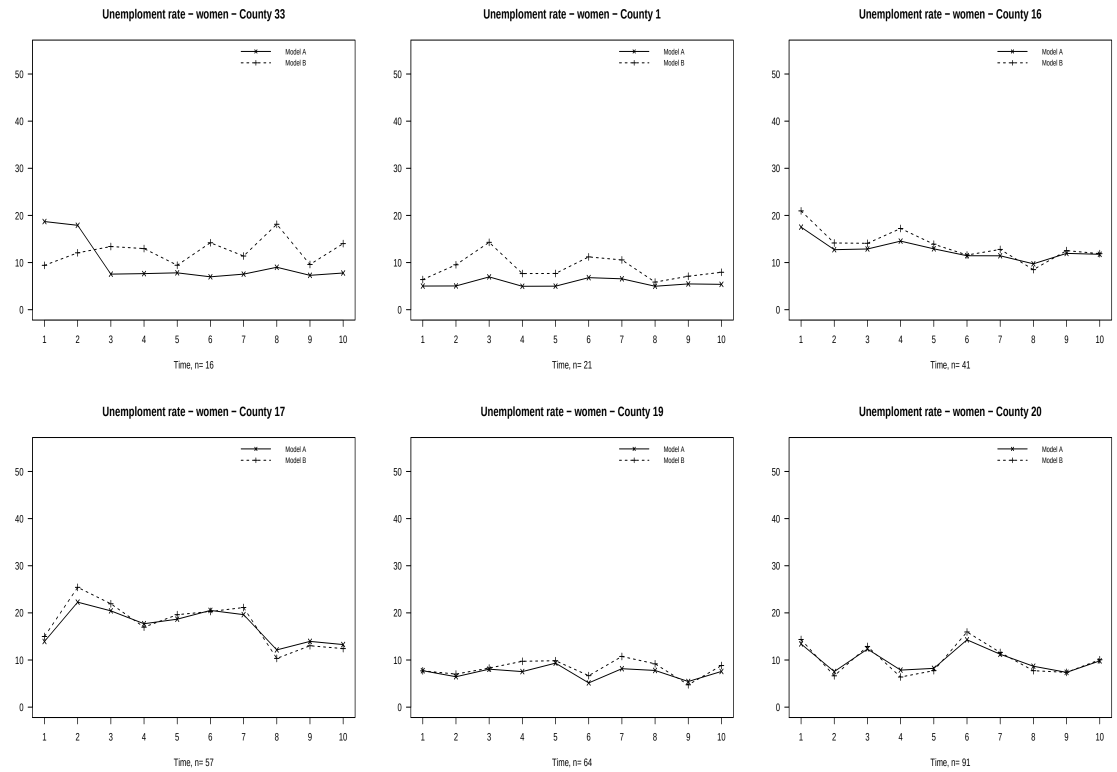


Figure 5.5.12: Women unemployment rates for models A and B and for all the periods

Chapter 6

Multinomial logit mixed model with correlated time and area effects

This chapter introduces a multinomial logit mixed model with correlated time and area effects. Like in the model of Chapter 5 we employ a multinomial model with two random effects, one associated with the category of employed people and the other associated with the category of unemployed people. This is due to the different modeling requirements for each labour category in the Galician data. Indeed, to take advantage from the availability of survey data from different time periods we propose a multinomial model with correlated time and domain random effects.

The chapter is organized as follows. Section 6.1 introduces the multinomial logit mixed model with correlated time and domain random effects. Section 6.2 develops the proposed model-based estimators and the corresponding MSE estimation procedures and Section 6.3 presents two simulation experiments. The first simulation studies the behavior of the estimates of the regression parameters, the variance components and the target population indicators. The second simulation studies the performance of the two introduced MSE estimation methods. Finally, Section 6.4 applies the proposed methodology to data from the LFS in Galicia.

6.1 The model

The multinomial models guarantee that the sum of estimated totals of employed and unemployed people is lower than the total of people. This is because we introduce a multinomial logit mixed model for estimating the domain totals of employed, unemployed and inactive people in a coherent way.

Like in the previous chapter, let us use indexes $d = 1, \dots, D$, $k = 1, \dots, q-1$ and $t = 1, \dots, T$ for the D domains, for the T time periods and for the categories of the target variable respectively. In the real data there are $q = 3$ categories, i.e. employed, unemployed and inactive people. However, there are only $q - 1 = 2$ categories in the multinomial model, i.e. employed and unemployed people. We take inactive people as the reference. Let $u_{1,dk}$ and $u_{2,dkt}$ be the random effects associated to category k, domain

d and time t. In vector notation, the random effects

$$\begin{aligned}
\boldsymbol{u}_1 &= \underset{1\le d\le D}{\mathrm{col}}(\boldsymbol{u}_{1,d}), \quad \boldsymbol{u}_{1,d} = \underset{1\le k\le q-1}{\mathrm{col}}(u_{1,dk}),\\
\boldsymbol{u}_2 &= \underset{1\le d\le D}{\mathrm{col}}(\boldsymbol{u}_{2,d}), \quad \boldsymbol{u}_{2,d} = \underset{1\le k\le q-1}{\mathrm{col}}(\boldsymbol{u}_{2,dk}), \quad \boldsymbol{u}_{2,dk} = \underset{1\le t\le T}{\mathrm{col}}(u_{2,dkt}).
\end{aligned}$$

We suppose that

1. $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ are independent,
2. $\boldsymbol{u}_1 \sim N(\boldsymbol{0}, \boldsymbol{V}_{u_1})$, where $\boldsymbol{V}_{u_1} = \underset{1\le d\le D}{\mathrm{diag}}\,(\underset{1\le k\le q-1}{\mathrm{diag}}\,(\varphi_{1k}))$, $k = 1,\dots,q-1$.
3. $\boldsymbol{u}_{2,dk} \sim N(\boldsymbol{0}, \boldsymbol{V}_{u_{2,dk}})$, $d = 1,\dots,D$, $k = 1,\dots,q-1$, are independent with covariance matrix AR(1); ie, $\boldsymbol{V}_{u_{2,dk}} = \varphi_{2k}\Omega_d(\phi_k)$ and

$$\Omega_d(\phi_k) = \Omega_{d,k} = \frac{1}{1-\phi_k^2}\begin{pmatrix}
1 & \phi_k & \dots & \phi_k^{T-2} & \phi_k^{T-1}\\
\phi_k & 1 & \ddots & & \phi_k^{T-2}\\
\vdots & \ddots & \ddots & \ddots & \vdots\\
\phi_k^{T-2} & & \ddots & 1 & \phi_k\\
\phi_k^{T-1} & \phi_k^{T-2} & \dots & \phi_k & 1
\end{pmatrix}_{T\times T}.$$

It holds that $\boldsymbol{V}_u = \mathrm{var}(\boldsymbol{u}) = \mathrm{diag}(\boldsymbol{V}_{u_1}, \boldsymbol{V}_{u_2})$, where $\boldsymbol{V}_{u_2} = \mathrm{var}(\boldsymbol{u}_2) = \underset{1\le d\le D}{\mathrm{diag}}\,(\underset{1\le k\le q-1}{\mathrm{diag}}\,(\boldsymbol{V}_{u_{2,dk}}))$. We also assume that the response vectors $\boldsymbol{y}_{dt} = \underset{1\le k\le q-1}{\mathrm{col}}(y_{dkt})$, conditioned to $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$, are independent with multinomial distributions

$$\boldsymbol{y}_{dt}|_{\boldsymbol{u}_1,\boldsymbol{u}_2} \sim \mathrm{M}(\nu_{dt}, p_{d1t},\dots,p_{dqt-1}),\ d = 1,\dots,D,\ t = 1,\dots,T. \tag{6.1.1}$$

where the ν_{dt}'s are known integer numbers which are equal to n_{dt} in the considered real data case. The covariance matrix of $\boldsymbol{y}_{dt}$ conditioned to $\boldsymbol{u}_{1,d}$ and $\boldsymbol{u}_{2,dt}$ is $\mathrm{var}(\boldsymbol{y}_{dt}|\boldsymbol{u}_{1,d}, \boldsymbol{u}_{2,dt}) = \boldsymbol{W}_{dt} = \nu_{dt}[\mathrm{diag}(\boldsymbol{p}_{dt}) - \boldsymbol{p}_{dt}\boldsymbol{p}_{dt}']$, where $\boldsymbol{p}_{dt} = \underset{1\le k\le q-1}{\mathrm{col}}(p_{dkt})$ and $\mathrm{diag}(\boldsymbol{p}_{dt}) = \underset{1\le k\le q-1}{\mathrm{diag}}\,(p_{dkt})$. For the natural parameters $\eta_{dkt} = \log\frac{p_{dkt}}{p_{dqt}}$, $k = 1,\dots,q-1$, we assume the model

$$\eta_{dkt} = \boldsymbol{x}_{dkt}\boldsymbol{\beta}_k + u_{1,dk} + u_{2,dkt},\ \ d = 1,\dots,D,\ k = 1,\dots,q-1,\ t = 1,\dots,T, \tag{6.1.2}$$

where $\boldsymbol{x}_{dkt} = \underset{1\le r\le p_r}{\mathrm{col}'}(x_{dktr})$, $\boldsymbol{\beta}_k = \underset{1\le r\le p_k}{\mathrm{col}}(\beta_{kr})$ and $p = \sum_{k=1}^{q-1} p_k$.

In matrix notation, the model is

$$\boldsymbol{\eta} = \boldsymbol{X\beta} + \boldsymbol{Z}_1\boldsymbol{u}_1 + \boldsymbol{Z}_2\boldsymbol{u}_2 = \boldsymbol{X\beta} + \boldsymbol{Zu},$$

where $\boldsymbol{Z} = (\boldsymbol{Z}_1', \boldsymbol{Z}_2')'$, $\boldsymbol{\eta} = \underset{1\le d\le D}{\mathrm{col}}(\boldsymbol{\eta}_d)$, $\boldsymbol{X} = \underset{1\le d\le D}{\mathrm{col}}(\boldsymbol{X}_d)$, $\boldsymbol{Z}_1 = \underset{1\le d\le D}{\mathrm{diag}}\,(\boldsymbol{Z}_{1d})$, $\boldsymbol{Z}_2 = \underset{1\le d\le D}{\mathrm{diag}}\,(\boldsymbol{Z}_{2d})$,

$$\begin{aligned}
\boldsymbol{\eta}_d &= \underset{1\le t\le T}{\mathrm{col}}(\underset{1\le k\le q-1}{\mathrm{col}}(\eta_{dkt})), \quad \boldsymbol{X}_d = \underset{1\le k\le q-1}{\mathrm{diag}}\,(\underset{1\le t\le T}{\mathrm{col}}(\boldsymbol{x}_{dkt})),\\
\boldsymbol{Z}_{1d} &= \underset{1\le k\le q-1}{\mathrm{diag}}\,(\boldsymbol{1}_T), \quad \boldsymbol{Z}_{2d} = \underset{1\le t\le T}{\mathrm{diag}}\,(\underset{1\le k\le q-1}{\mathrm{diag}}\,(1)) = \boldsymbol{I}_{T(q-1)}, \quad \boldsymbol{1}_T = \underset{1\le t\le T}{\mathrm{col}}(1).
\end{aligned}$$

Alternatively the model (6.1.2) can be expressed as

$$\boldsymbol{\eta}_{dt} = \boldsymbol{X}_{dt}\boldsymbol{\beta} + \mathsf{Z}_{1,dt}\mathsf{u}_1 + \mathsf{Z}_{2,dt}\mathsf{u}_2 = \boldsymbol{X}_{dt}\boldsymbol{\beta} + \mathsf{Z}_{dt}\mathsf{u}, \quad d=1,\ldots,D,\ t=1,\ldots,T, \qquad (6.1.3)$$

where $\mathsf{u} = (\mathsf{u}_1', \mathsf{u}_2')'$, $\mathsf{u}_1 = \underset{1\le k\le q-1}{\text{col}}(\boldsymbol{u}_{1,k})$, $\mathsf{u}_2 = \underset{1\le k\le q-1}{\text{col}}(\boldsymbol{u}_{2,k})$, the $\boldsymbol{u}_{1,k} = \underset{1\le d\le D}{\text{col}}(u_{1,dk}) \sim N(\boldsymbol{0}, \varphi_{1k}\boldsymbol{I}_D)$ and the $\boldsymbol{u}_{2,dk} \sim N(\boldsymbol{0}, \boldsymbol{V}_{u_{2,dk}})$ where $\boldsymbol{V}_{u_{2,dk}} = \varphi_{2k}\Omega_d(\phi_k)$, $\mathsf{Z}_{dt} = (\mathsf{Z}_{1,dt}, \mathsf{Z}_{2,dt})$, $\mathsf{Z}_{1,dt} = \underset{1\le k\le q-1}{\text{diag}}\,(\underset{1\le \ell\le D}{\text{col}'}(\delta_{\ell d}))$, $\mathsf{Z}_{2,dt} = \underset{1\le k\le q-1}{\text{diag}}\,(\underset{1\le \ell\le D}{\text{col}'}(\underset{1\le t\le T}{\text{col}'}(\delta_{\ell d}\delta_{st})))$. In matrix notation (6.1.3) can be expressed as

$$\boldsymbol{\eta} = \boldsymbol{X}\boldsymbol{\beta} + \mathsf{Z}_1\mathsf{u}_1 + \mathsf{Z}_2\mathsf{u}_2 = \boldsymbol{X}\boldsymbol{\beta} + \mathsf{Z}\mathsf{u}, \qquad (6.1.4)$$

where $\mathsf{Z}_r = \underset{1\le d\le D}{\text{col}}(\underset{1\le t\le T}{\text{col}}(\mathsf{Z}_{r,dt}))$, $r=1,2$, and $\mathsf{Z} = (\mathsf{Z}_1, \mathsf{Z}_2)$. (6.1.2) and (6.1.4) are the same model. The difference is only in the management of vector elements $\boldsymbol{u}$ (or u) of random effects.

6.2 The PQL-REML fitting algorithm

To fit the model we combine the PQL method, described by Breslow and Clayton (1996) for estimating and predicting the $\boldsymbol{\beta}_k$'s, the $\boldsymbol{u}_{1d}$'s,the $\boldsymbol{u}_{2d}$'s and the ϕ_k's, with the REML method for estimating the variance components $\boldsymbol{\varphi}_1$, $\boldsymbol{\varphi}_2$ and ϕ. The presented method was described in Section 3.2. It is based on a normal approximation to the joint probability distribution of the vector $(\boldsymbol{y}, \boldsymbol{u})$. The combined algorithm was first introduced by Schall (1991) and later used by Saei and Chambers (2003) and Molina et al. (2007) in applications of generalized linear mixed models to small area estimation problems. In this chapter, we adapt the combined algorithm to the multinomial logit mixed model defined by (6.1.1) and (6.1.2). The loglikelihood of $\boldsymbol{y}$ conditioned to $\boldsymbol{u}$ is

$$\begin{aligned} l_1(\boldsymbol{y}|\boldsymbol{u}) &= \sum_{d=1}^{D}\sum_{t=1}^{T}\left\{\sum_{k=1}^{q-1} y_{dkt}\log\frac{p_{dkt}}{p_{dqt}} + \nu_{dt}\log p_{dqt} + \log\frac{\nu_{dt}}{y_{d1t}!\cdots y_{dqt}!}\right\} \\ &= \sum_{d=1}^{D}\sum_{t=1}^{T}\left\{\sum_{k=1}^{q-1} y_{dkt}\eta_{dkt} - \nu_{dt}\log\big(1 + \sum_{k=1}^{q-1}\exp\{\eta_{dkt}\}\big) + \log\frac{\nu_{dt}}{y_{d1t}!\cdots y_{dqt}!}\right\}. \end{aligned}$$

The partial derivatives of

$$\eta_{dkt} = \sum_{r=1}^{p_k} x_{dktr}\beta_{kr} + u_{1,dk} + u_{2,dkt}, \quad d=1,\ldots,D,\ t=1,\ldots,T,\ k=1,\ldots,q-1,$$

with respect to β_{kr}, $u_{1,dk}$ and $u_{2,dkt}$ are

$$\frac{\partial\eta_{dkt}}{\partial\beta_{kr}} = x_{dktr}, \quad \frac{\partial\eta_{dkt}}{\partial u_{1,dk}} = 1, \quad \frac{\partial\eta_{dkt}}{\partial u_{2,dkt}} = 1.$$

The first order partial derivatives of l_1 are

$$\begin{aligned}
S_{1,\beta_{kr}} &= \frac{\partial l_1(\boldsymbol{y}|\boldsymbol{u})}{\partial \beta_{kr}} = \sum_{d=1}^{D}\sum_{t=1}^{T}\left\{x_{dktr}y_{dkt} - \frac{\nu_{dt}x_{dktr}\exp\{\eta_{dkt}\}}{1+\sum_{\ell=1}^{q-1}\exp\{\eta_{d\ell t}\}}\right\} \\
&= \sum_{d=1}^{D}\sum_{t=1}^{T} x_{dktr}(y_{dkt}-\mu_{dkt}), \\
S_{1,u_{1,dk}} &= \frac{\partial l_1(\boldsymbol{y}|\boldsymbol{u})}{\partial u_{1,dk}} = \sum_{t=1}^{T}(y_{dkt}-\mu_{dkt}), \quad S_{1,u_{2,dkt}} = \frac{\partial l_1(\boldsymbol{y}|\boldsymbol{u})}{\partial u_{2,dkt}} = (y_{dkt}-\mu_{dkt}),
\end{aligned}$$

The vector expressions of the first order partial derivatives of l_1 are

$$\begin{aligned}
\boldsymbol{S}_{1,\beta} &= \underset{1\le k\le q-1}{\text{col}}\big(\underset{1\le r\le p_k}{\text{col}}(S_{1,\beta_{kr}})\big) = \underset{1\le k\le q-1}{\text{col}}\left(\sum_{d=1}^{D}\sum_{t=1}^{T}\boldsymbol{x}'_{dkt}(y_{dkt}-\mu_{dkt})\right), \\
\boldsymbol{S}_{1,u_1} &= \underset{1\le d\le D}{\text{col}}\big(\underset{1\le k\le q-1}{\text{col}}(S_{1,u_{1,dk}})\big), \quad \boldsymbol{S}_{1,u_2} = \underset{1\le d\le D}{\text{col}}\big(\underset{1\le k\le q-1}{\text{col}}\big(\underset{1\le t\le T}{\text{col}}(S_{1,u_{2,dkt}})\big)\big).
\end{aligned}$$

The second order partial derivatives of l_1 are

$$\begin{aligned}
H_{1,\beta_{kr}\beta_{ks}} &= \frac{\partial^2 l_1(\boldsymbol{y}|\boldsymbol{u})}{\partial \beta_{kr}\partial \beta_{ks}} = -\sum_{d=1}^{D}\sum_{t=1}^{T}\nu_{dt}x_{dktr}x_{dkts}p_{dkt}(1-p_{dkt}), \\
H_{1,u_{1,dk}\beta_{kr}} &= \frac{\partial^2 l_1(\boldsymbol{y}|\boldsymbol{u})}{\partial u_{1,dk}\partial \beta_{kr}} = -\sum_{t=1}^{T}\nu_{dt}x_{dktr}p_{dkt}(1-p_{dkt}), \\
H_{1,u_{2,dkt}\beta_{kr}} &= \frac{\partial^2 l_1(\boldsymbol{y}|\boldsymbol{u})}{\partial u_{2,dkt}\partial \beta_{kr}} = -\nu_{dt}x_{dktr}p_{dkt}(1-p_{dkt}), \\
H_{1,u_{1,dk}u_{1,dk}} &= \frac{\partial^2 l_1(\boldsymbol{y}|\boldsymbol{u})}{\partial u_{1,dk}\partial u_{1,dk}} = -\sum_{t=1}^{T}\nu_{dt}p_{dkt}(1-p_{dkt}), \\
H_{1,u_{2,dkt}u_{2,dkt}} &= \frac{\partial^2 l_1(\boldsymbol{y}|\boldsymbol{u})}{\partial u_{2,dkt}\partial u_{2,dkt}} = -\nu_{dt}p_{dkt}(1-p_{dkt}), \\
H_{1,u_{1,dk}u_{2,dkt}} &= \frac{\partial^2 l_1(\boldsymbol{y}|\boldsymbol{u})}{\partial u_{1,dk}\partial u_{2,dkt}} = -\nu_{dt}p_{dkt}(1-p_{dkt}).
\end{aligned}$$

For the case $k_1 \neq k_2$, we have that

$$\begin{aligned}
H_{1,\beta_{k_1r}\beta_{k_2s}} &= \frac{\partial^2 l_1(\boldsymbol{y}|\boldsymbol{u})}{\partial \beta_{k_1r}\partial \beta_{k_2s}} = \sum_{d=1}^{D}\sum_{t=1}^{T}\nu_{dt}x_{dk_1tr}x_{dk_2s}p_{dk_1t}p_{dk_2t}, \\
H_{1,u_{1,dk_1}\beta_{k_2s}} &= \frac{\partial^2 l_1(\boldsymbol{y}|\boldsymbol{u})}{\partial u_{1,dk_1}\partial \beta_{k_2s}} = \sum_{t=1}^{T}\nu_{dt}x_{dk_2ts}p_{dk_1t}p_{dk_2t}, \\
H_{1,u_{2,dk_1t}\beta_{k_2s}} &= \frac{\partial^2 l_1(\boldsymbol{y}|\boldsymbol{u})}{\partial u_{2,dk_1t}\partial \beta_{k_2s}} = \nu_{dt}x_{dk_2ts}p_{dk_1t}p_{dk_2t},
\end{aligned}$$

and that

$$
\begin{aligned}
H_{1,u_{1,dk_1}u_{1,dk_2}} &= \frac{\partial^2 l_1(\boldsymbol{y}|\boldsymbol{u})}{\partial u_{1,dk_1}\partial u_{1,dk_2}} = \sum_{t=1}^{T} \nu_{dt} p_{dk_1 t} p_{dk_2 t}, \\
H_{1,u_{2,dk_1t}u_{2,dk_2t}} &= \frac{\partial^2 l_1(\boldsymbol{y}|\boldsymbol{u})}{\partial u_{2,dk_1t}\partial u_{2dk_2t}} = \nu_{dt} p_{dk_1 t} p_{dk_2 t}, \\
H_{1,u_{1,dk_1}u_{2,dk_2t}} &= \frac{\partial^2 l_1(\boldsymbol{y}|\boldsymbol{u})}{\partial u_{1,dk_1}\partial u_{2,dk_2t}} = \nu_{dt} p_{dk_1 t} p_{dk_2 t}.
\end{aligned}
$$

Finally, for the case $d_1 \neq d_2$, we have

$$
\begin{aligned}
H_{1,u_{1,d_1k_1}u_{1,d_2k_2}} &= \frac{\partial^2 l_1(\boldsymbol{y}|\boldsymbol{u})}{\partial u_{1,d_1k_1}\partial u_{1,d_2,k_2}} = 0,\ d_1 \neq d_2, \\
H_{1,u_{2,d_1k_1t_1}u_{2,d_2k_2t_2}} &= \frac{\partial^2 l_1(\boldsymbol{y}|\boldsymbol{u})}{\partial u_{2,d_1k_1t_1}\partial u_{2,d_2k_2t_2}} = 0,\ d_1 \neq d_2 \text{ or } t_1 \neq t_2, \\
H_{1,u_{1,d_1k_1}u_{2,d_2k_2t}} &= \frac{\partial^2 l_1(\boldsymbol{y}|\boldsymbol{u})}{\partial u_{1,d_1k_1}\partial u_{2,d_2k_2t}} = 0,\ d_1 \neq d_2.
\end{aligned}
$$

The matrix of the second order partial derivatives of l_1 are

$$
\begin{aligned}
\boldsymbol{H}_{1,\beta\beta} &= \Big(H_{1,\beta_{k_1r}\beta_{k_2s}}\Big)_{\substack{k_1=1,\dots,q-1;\, r=1,\dots,p_{k_1}\\ k_2=1,\dots,q-1;\, s=1,\dots,p_{k_2}}}, \\
\boldsymbol{H}_{1,\beta u_1} &= \Big(H_{1,\beta_{k_1r}u_{1,dk_2}}\Big)_{\substack{k_1=1,\dots,q-1;\, r=1,\dots,p_{k_1}\\ d=1,\dots,D;\, k_2=1,\dots,q-1}}, \\
\boldsymbol{H}_{1,\beta u_2} &= \Big(H_{1,\beta_{k_1r}u_{2,dk_2t}}\Big)_{\substack{k_1=1,\dots,q-1;\, r=1,\dots,p_{k_1}\\ d=1,\dots,D;\, k_2=1,\dots,q-1;\, t=1,\dots,T}}, \\
\boldsymbol{H}_{1,u_1u_1} &= \Big(H_{1,u_{1,d_1k_1}u_{1,d_2k_2}}\Big)_{\substack{d_1=1,\dots,D;\, k_1=1,\dots,q-1\\ d_2=1,\dots,D;\, k_2=1,\dots,q-1}}, \\
\boldsymbol{H}_{1,u_2u_2} &= \Big(H_{1,u_{2,d_1k_1t_1}u_{2,d_2k_2t_2}}\Big)_{\substack{d_1=1,\dots,D;\, k_1=1,\dots,q-1;\, t_1=1,\dots,T\\ d_2=1,\dots,D;\, k_2=1,\dots,q-1;\, t_2=1,\dots,T}}, \\
\boldsymbol{H}_{1,u_1u_2} &= \Big(H_{1,u_{1,d_1k_1}u_{2,d_2k_2t}}\Big)_{\substack{d_1=1,\dots,D;\, k_1=1,\dots,q-1\\ d_2=1,\dots,D;\, k_2=1,\dots,q-1;\, t=1,\dots,T}}.
\end{aligned}
$$

The likelihood of $\boldsymbol{u}$ is

$$
f_2(\boldsymbol{u}) = \frac{1}{(\sqrt{2\pi})^{M(q-1)}|\boldsymbol{V}_u|^{1/2}} \exp\left\{-\frac{1}{2}\boldsymbol{u}'\boldsymbol{V}_u^{-1}\boldsymbol{u}\right\},
$$

where $M = DT$. The loglikelihood of $\boldsymbol{u}$ is

$$
l_2(\boldsymbol{u}) = \kappa - \sum_{k=1}^{q-1}\left\{\frac{D}{2}\log\varphi_{1k} + \frac{M}{2}\log\varphi_{2k} + \frac{1}{2}\sum_{d=1}^{D}\log|\Omega_{dk}| + \frac{\boldsymbol{u}_{1,\cdot k}'\boldsymbol{u}_{1,\cdot k}}{2\varphi_{1k}} + \sum_{d=1}^{D}\frac{\boldsymbol{u}_{2,dk}'\Omega_{dk}^{-1}\boldsymbol{u}_{2,dk}}{2\varphi_{2k}}\right\},
$$

where $\kappa = -\frac{M(q-1)}{2}\log 2\pi$ and $\boldsymbol{u}_{1,\cdot k} = \underset{1\le d\le D}{\text{col}}(u_{1,dk})$.

The first order partial derivatives of l_2 are

$$S_{2,u_{1,dk}} = \frac{\partial l_2(\boldsymbol{u})}{\partial u_{1,dk}} = -\frac{1}{\varphi_{1k}} u_{1,dk}, \quad S_{2,u_{2,dkt}} = -\frac{1}{\varphi_{2k}} \operatorname*{col'}_{1\le \ell \le T} (\delta_{t\ell}) \Omega_{dk}^{-1} \boldsymbol{u}_{2,dk}.$$

The vector expressions of the partial derivatives of l_2 are

$$\boldsymbol{S}_{2,u_1} = \operatorname*{col}_{1\le d\le D}\big(\operatorname*{col}_{1\le k\le q-1}(S_{2,u_{1,dk}})\big), \quad \boldsymbol{S}_{2,u_2} = \operatorname*{col}_{1\le d\le D}\big(\operatorname*{col}_{1\le k\le q-1}\big(\operatorname*{col}_{1\le t\le T}(S_{2,u_{2,dkt}})\big)\big).$$

The second order partial derivatives of l_2 are

$$\begin{aligned}
H_{2,u_{1,dk}u_{1,dk}} &= -\frac{1}{\varphi_{1k}}, \quad H_{2,u_{2,dkt_1}u_{2,dkt_2}} = -\frac{1}{\varphi_{2k}} \operatorname*{col'}_{1\le \ell\le T}(\delta_{t_1\ell})\Omega_{dk}^{-1} \operatorname*{col}_{1\le \ell\le T}(\delta_{t_2\ell}) = -\frac{\nu_{t_1t_2}^{dk}}{\varphi_{2k}}, \\
H_{2,u_{1,d_1k_1}u_{1,d_2k_2}} &= H_{2,u_{2,d_1k_1t_1}u_{2,d_2k_2t_2}} = 0, \ k_1 \neq k_2 \text{ or } d_1 \neq d_2, \\
H_{2,u_{1,d_1k_1},u_{2,d_2k_2t}} &= 0, \ d_1, d_2 = 1, \dots, D, \ k_1, k_2 = 1, \dots, q-1, \ t = 1, \dots, T,
\end{aligned}$$

where $\nu_{t_1t_2}^{dk}$ is the element (t_1, t_2) of the matrix Ω_{dk}^{-1}.

The matrix expressions of the second order partial derivatives of l_2 are

$$\begin{aligned}
\boldsymbol{H}_{2,u_1u_1} &= \Big(H_{2,u_{1,d_1k_1}u_{1,d_2k_2}}\Big)_{\substack{d_1=1,\dots,D;\, k_1=1,\dots,q-1; \\ d_2=1,\dots,D;\, k_2=1,\dots,q-1}} \\
\boldsymbol{H}_{2,u_2u_2} &= \Big(H_{2,u_{2,d_1k_1t_1}u_{2,d_2k_2t_2}}\Big)_{\substack{d_1=1,\dots,D;\, k_1=1,\dots,q-1;\, t_1=1,\dots,T \\ d_2=1,\dots,D;\, k_2=1,\dots,q-1;\, t_2=1,\dots,T}} \\
\boldsymbol{H}_{2,u_1u_2} &= \Big(H_{2,u_{1,d_1k_1}u_{2,d_2k_2t}}\Big)_{\substack{d_1=1,\dots,D;\, k_1=1,\dots,q-1 \\ d_2=1,\dots,D;\, k_2=1,\dots,q-1;\, t=1,\dots,T}} = \boldsymbol{0}.
\end{aligned}$$

The loglikelihood of $(\boldsymbol{y}, \boldsymbol{u})$ is

$$l(\boldsymbol{y}, \boldsymbol{u}) = l_1(\boldsymbol{y}|\boldsymbol{u}) + l_2(\boldsymbol{u}).$$

The first order partial derivatives of l are

$$\boldsymbol{S}_\beta = \boldsymbol{S}_{1,\beta}, \quad \boldsymbol{S}_{u_1} = \boldsymbol{S}_{1,u_1} + \boldsymbol{S}_{2,u_1}, \quad \boldsymbol{S}_{u_2} = \boldsymbol{S}_{1,u_2} + \boldsymbol{S}_{2,u_2}.$$

The blocks of the Fisher information matrix associated to l are

$$\begin{aligned}
\boldsymbol{F}_{\beta\beta} &= -\boldsymbol{H}_{1,\beta\beta}, \ \boldsymbol{F}_{u_1u_1} = -\boldsymbol{H}_{1,u_1u_1} - \boldsymbol{H}_{2,u_1u_1}, \ \boldsymbol{F}_{u_2u_2} = -\boldsymbol{H}_{1,u_2u_2} - \boldsymbol{H}_{2,u_2u_2}, \\
\boldsymbol{F}_{u_1\beta} &= -\boldsymbol{H}_{1,u_1\beta}, \ \boldsymbol{F}_{\beta u_1} = \boldsymbol{F}'_{u_1,\beta}, \ \boldsymbol{F}_{u_2\beta} = -\boldsymbol{H}_{1,u_2\beta}, \ \boldsymbol{F}_{\beta u_2} = \boldsymbol{F}'_{u_2\beta}, \\
\boldsymbol{F}_{u_1u_2} &= -\boldsymbol{H}_{1,u_1u_2}, \ \boldsymbol{F}_{u_2u_1} = \boldsymbol{F}'_{u_1u_2}.
\end{aligned}$$

We define

$$\boldsymbol{S} = \begin{pmatrix} \boldsymbol{S}_\beta \\ \boldsymbol{S}_u \end{pmatrix}, \ \boldsymbol{S}_u = \begin{pmatrix} \boldsymbol{S}_{u_1} \\ \boldsymbol{S}_{u_1} \end{pmatrix}, \ \boldsymbol{F}_{uu} = \begin{pmatrix} \boldsymbol{F}_{u_1u_1} & \boldsymbol{F}_{u_1u_2} \\ \boldsymbol{F}_{u_2u_1} & \boldsymbol{F}_{u_2u_2} \end{pmatrix}, \ \boldsymbol{F}_{u\beta} = \begin{pmatrix} \boldsymbol{F}_{u_1\beta} \\ \boldsymbol{F}_{u_2\beta} \end{pmatrix}, \ \boldsymbol{F}_{\beta u} = \boldsymbol{F}'_{u\beta},$$

$$\boldsymbol{F} = \begin{pmatrix} \boldsymbol{F}_{\beta\beta} & \boldsymbol{F}_{\beta u} \\ \boldsymbol{F}_{u\beta} & \boldsymbol{F}_{uu} \end{pmatrix}, \ \boldsymbol{F}^{-1} = \begin{pmatrix} \boldsymbol{F}^{\beta\beta} & \boldsymbol{F}^{\beta u} \\ \boldsymbol{F}^{u\beta} & \boldsymbol{F}^{uu} \end{pmatrix}, \ \boldsymbol{F}_{uu}^{-1} = \begin{pmatrix} \boldsymbol{F}^{u_1u_1} & \boldsymbol{F}^{u_1u_2} \\ \boldsymbol{F}^{u_2u_1} & \boldsymbol{F}^{u_2u_2} \end{pmatrix}.$$

It holds that

$$\begin{aligned}
\boldsymbol{F}^{\beta\beta} &= (\boldsymbol{F}_{\beta\beta} - \boldsymbol{F}_{\beta u}\boldsymbol{F}_{uu}^{-1}\boldsymbol{F}_{u\beta})^{-1}, \quad \boldsymbol{F}^{\beta u} = -\boldsymbol{F}^{\beta\beta}\boldsymbol{F}_{\beta u}\boldsymbol{F}_{uu}^{-1},\\
\boldsymbol{F}^{u\beta} &= (\boldsymbol{F}^{\beta u})', \quad \boldsymbol{F}^{u} = \boldsymbol{F}_{uu}^{-1} + \boldsymbol{F}_{uu}^{-1}\boldsymbol{F}_{u\beta}\boldsymbol{F}^{\beta\beta}\boldsymbol{F}_{\beta u}\boldsymbol{F}_{uu}^{-1},\\
\boldsymbol{F}^{u_1u_1} &= (\boldsymbol{F}_{u_1u_1} - \boldsymbol{F}_{u_1u_2}\boldsymbol{F}_{u_2u_2}^{-1}\boldsymbol{F}_{u_2u_1})^{-1}, \quad \boldsymbol{F}^{u_1u_2} = -\boldsymbol{F}^{u_1u_1}\boldsymbol{F}_{u_1u_2}\boldsymbol{F}_{u_2u_2}^{-1},\\
\boldsymbol{F}^{u_2u_1} &= (\boldsymbol{F}^{u_1u_2})', \quad \boldsymbol{F}^{u_2u_2} = \boldsymbol{F}_{u_2u_2}^{-1} + \boldsymbol{F}_{u_2u_2}^{-1}\boldsymbol{F}_{u_2u_1}\boldsymbol{F}^{u_1u_1}\boldsymbol{F}_{u_1u_2}\boldsymbol{F}_{u_2u_2}^{-1}.
\end{aligned}$$

The algorithm has two parts. In the first part it updates the values of $\boldsymbol{\beta}_k$, $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$. In the second part it updates the variance components.

Algorithm A. Let $\varphi_{1k}, \varphi_{2k}, \phi_k$, $k = 1, \ldots, q-1$, known. The PQL estimator of $\boldsymbol{\beta}$ and $\boldsymbol{u}$ are calculated by using the Fisher-scoring algorithm, which is described below.

(A1) *Beginning:* Assign the initial values $\boldsymbol{\beta}^{(0)} = \boldsymbol{\beta}^{initial}$ and $\boldsymbol{u}^{(0)} = \boldsymbol{u}^{initial}$.

(A2) *Iteration $r+1$:* For $d = 1, \ldots, D$, $k = 1, \ldots, q-1$, $t = 1, \ldots, T$, calculate

$$\boldsymbol{\eta}^{(r)} = \boldsymbol{X}\boldsymbol{\beta}^{(r)} + \boldsymbol{Z}_1\boldsymbol{u}_1^{(r)} + \boldsymbol{Z}_2\boldsymbol{u}_2^{(r)}, \; p_{dkt}^{(r)} = \frac{\exp\{\eta_{dkt}^{(r)}\}}{1 + \exp\{\eta_{dkt}^{(r)}\}}, \; \mu_{dkt}^{(r)} = \nu_{dt}p_{dkt}^{(r)}.$$

Update the values of $\boldsymbol{\beta}^{(r)}$ and $\boldsymbol{u}^{(r)}$ by using the equation

$$\begin{bmatrix} \boldsymbol{\beta}^{(r+1)} \\ \boldsymbol{u}^{(r+1)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\beta}^{(r)} \\ \boldsymbol{u}^{(r)} \end{bmatrix} + \boldsymbol{F}^{-1}(\boldsymbol{\beta}^{(r)}, \boldsymbol{u}^{(r)})\boldsymbol{S}(\boldsymbol{\beta}^{(r)}, \boldsymbol{u}^{(r)}).$$

(A3) *End:* Repeat the step (A.2) until convergence of $\boldsymbol{\beta}^{(r)}$ and $\boldsymbol{u}^{(r)}$.

The updated variance components are obtained by applying the Fisher-scoring algorithm to the REML loglikelihood of $\boldsymbol{\eta} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u} + \boldsymbol{e}$, where $\boldsymbol{e} \sim N(\boldsymbol{0}, \boldsymbol{W}^{-1})$ and $\boldsymbol{u} \sim N(\boldsymbol{0}, \boldsymbol{V}_u)$ are independent with

$$\begin{aligned}
\boldsymbol{V} &= \text{var}(\boldsymbol{\eta}) = \boldsymbol{Z}\boldsymbol{V}_u\boldsymbol{Z}' + \boldsymbol{W}^{-1} = \boldsymbol{Z}_1\boldsymbol{V}_{u_1}\boldsymbol{Z}_1' + \boldsymbol{Z}_2\boldsymbol{V}_{u_2}\boldsymbol{Z}_2' + \boldsymbol{W}^{-1},\\
\boldsymbol{W} &= \boldsymbol{W}(\boldsymbol{\eta}) = \text{var}(\boldsymbol{y}|\boldsymbol{u}) = \underset{1\le d\le D}{\text{diag}}\,(\boldsymbol{W}_d), \; \boldsymbol{W}_d = \boldsymbol{W}_{d1} + \boldsymbol{W}_{d2},\\
\boldsymbol{W}_{d1} &= \underset{1\le k\le q-1}{\text{diag}}\,(\underset{1\le t\le T}{\text{diag}}\,(\nu_{dt}p_{dkt})), \; \boldsymbol{W}_{d2} = \underset{1\le k_1,k_2\le q-1}{\text{matrix}}\,(\underset{1\le t\le T}{\text{diag}}\,(-\nu_{dt}p_{dk_1t}p_{dk_2t})).
\end{aligned}$$

The REML loglikelihood is

$$l_{reml}(\boldsymbol{\eta}) = -\frac{1}{2}(n-p)\log 2\pi - \frac{1}{2}\log|\boldsymbol{K}'\boldsymbol{V}\boldsymbol{K}| - \frac{1}{2}\boldsymbol{\eta}'\boldsymbol{P}\boldsymbol{\eta},$$

where $\boldsymbol{P} = \boldsymbol{V}^{-1} - \boldsymbol{V}^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}^{-1}$ and $\boldsymbol{K} = \boldsymbol{W} - \boldsymbol{W}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{W}$. We define

$$\begin{aligned}
\boldsymbol{V}_{1k} &= \frac{\partial \boldsymbol{V}}{\partial \varphi_{1k}} = \boldsymbol{Z}_1\frac{\partial \boldsymbol{V}_{u_1}}{\partial \varphi_{1k}}\boldsymbol{Z}_1' = \underset{1\le d\le D}{\text{diag}}\,(\boldsymbol{V}_{1dk}), \; \boldsymbol{V}_{1dk} = \underset{1\le \ell\le q-1}{\text{diag}}\,(\boldsymbol{1}_T)\underset{1\le \ell\le q-1}{\text{diag}}\,(\delta_{\ell k})\underset{1\le \ell\le q-1}{\text{diag}}\,(\boldsymbol{1}_T'),\\
\boldsymbol{V}_{1dk} &= \underset{1\le \ell\le q-1}{\text{diag}}\,(\delta_{\ell k}\boldsymbol{1}_T\boldsymbol{1}_T'),\\
\boldsymbol{V}_{2k} &= \frac{\partial \boldsymbol{V}}{\partial \varphi_{2k}} = \boldsymbol{Z}_2\frac{\partial \boldsymbol{V}_{u_2}}{\partial \varphi_{2k}}\boldsymbol{Z}_2' = \underset{1\le d\le D}{\text{diag}}\,(\boldsymbol{V}_{2dk}), \; \boldsymbol{V}_{2dk} = \underset{1\le \ell\le q-1}{\text{diag}}\,(\delta_{\ell k}\Omega_d(\phi_\ell)),\\
\boldsymbol{V}_{3k} &= \frac{\partial \boldsymbol{V}}{\partial \phi_k} = \boldsymbol{Z}_2\frac{\partial \boldsymbol{V}_{u_2}}{\partial \phi_k}\boldsymbol{Z}_2' = \underset{1\le d\le D}{\text{diag}}\,(\boldsymbol{V}_{3dk}), \; \boldsymbol{V}_{3dk} = \underset{1\le \ell\le q-1}{\text{diag}}\,(\delta_{\ell k}\varphi_{2\ell}\dot{\Omega}_d(\phi_\ell)).
\end{aligned}$$

We also define

$$S_{1k}=\frac{\partial l_{reml}}{\partial \varphi_{1k}},\quad S_{2k}=\frac{\partial l_{reml}}{\partial \varphi_{2k}},\quad S_{2k}=\frac{\partial l_{reml}}{\partial \phi_{k}}.$$

The elements of the vector of scores are

$$S_{ak}=-\frac{1}{2}\,\mathrm{tr}\{\boldsymbol{P}\boldsymbol{V}_{ak}\}+\frac{1}{2}\,\boldsymbol{\eta}'\boldsymbol{P}\boldsymbol{V}_{ak}\boldsymbol{P}\boldsymbol{\eta},\quad a=1,2,3,\ k=1,\dots,q-1.$$

We define

$$F_{ak_1,1k_2}=-E\left[\frac{\partial S_{ak_1}}{\partial \varphi_{1k_2}}\right],\ F_{ak_1,2k_2}=-E\left[\frac{\partial S_{ak_1}}{\partial \varphi_{2k_2}}\right],\ F_{ak_1,3k_2}=-E\left[\frac{\partial S_{ak_1}}{\partial \phi_{k_2}}\right],\quad a=1,2,3.$$

The elements of the Fisher information matrix are

$$F_{ak_1,bk_2}=\frac{1}{2}\,\mathrm{tr}\{\boldsymbol{P}\boldsymbol{V}_{ak_1}\boldsymbol{P}\boldsymbol{V}_{bk_2}\},\quad a,b=1,2,3,\ k_1,k_2=1,\dots,q-1.$$

The Fisher-scoring method use this updating formula

$$\boldsymbol{\sigma}^{(\ell+1)}=\boldsymbol{\sigma}^{(\ell)}+\boldsymbol{F}^{-1}(\boldsymbol{\sigma}^{(\ell)})\boldsymbol{S}(\boldsymbol{\sigma}^{(\ell)}).$$

where $\boldsymbol{\sigma}^{(\ell)}$ is the actual version of vector $\boldsymbol{\sigma}=(\varphi_{11},\dots,\varphi_{1q-1},\varphi_{21},\dots,\varphi_{2q-1},\phi_1,\dots,\phi_{q-1})$, $\boldsymbol{S}=\underset{1\le a\le 3}{\mathrm{col}}(\underset{1\le k\le q-1}{\mathrm{col}}(\boldsymbol{S}_{ak}))$ is the vector of scores and $\boldsymbol{F}=\underset{1\le a,b\le 3}{\mathrm{matrix}}(\boldsymbol{F}_{ab})$ is the Fisher information matrix, with components $\boldsymbol{F}_{ab}=\underset{1\le k_1,k_2\le q-1}{\mathrm{matrix}}(\boldsymbol{F}_{ak_1,bk_2})$.

The update of the variance components can be done by applying the Fisher-Scoring algorithm to the REML log-likelihood. The algorithm is described below.

Algorithm B.

(B1) Do $\ell=1$. Initiate the values $\boldsymbol{\beta}^{(0)}$, $\boldsymbol{u}^{(0)}$ and $\boldsymbol{\sigma}^{(1)}$.

(B2) Run the *Algorithm A* by using $\boldsymbol{\sigma}^{(\ell)}$ as known value of the vector of variance components, and $\boldsymbol{\beta}^{(\ell-1)}$, $\boldsymbol{u}^{(\ell-1)}$ as initial values. Let $\boldsymbol{\beta}^{(\ell)}$ and $\boldsymbol{u}^{(\ell)}$ be the obtained estimators and predictors.

(B3) Calculate $\boldsymbol{\eta}^{(\ell)}=\boldsymbol{X}\boldsymbol{\beta}^{(\ell)}+\boldsymbol{Z}_1\boldsymbol{u}_1^{(\ell)}+\boldsymbol{Z}_2\boldsymbol{u}_2^{(\ell)}$. Apply the updating equation

$$\boldsymbol{\sigma}^{(\ell+1)}=\boldsymbol{\sigma}^{(\ell)}+\boldsymbol{F}^{-1}(\boldsymbol{\sigma}^{(\ell)})\boldsymbol{S}(\boldsymbol{\sigma}^{(\ell)}).$$

(B4) Repeat the steps (B.2)-(B.3) until the convergence of $\boldsymbol{\beta}^{(\ell)}$, $\boldsymbol{u}^{(\ell)}$ and $\boldsymbol{\sigma}^{(\ell)}$.

The updated variance components φ_{rk}, $r=1,2$, $k=1,\dots,q-1$, can also be calculated by applying the formula

$$\widehat{\varphi}_{rk}=\frac{\widehat{\boldsymbol{u}}_{rk}'\boldsymbol{\Sigma}_{u_{rk}}^{-1}\widehat{\boldsymbol{u}}_{rk}}{\dim(\boldsymbol{u}_{rk})-\tau_{rk}},\quad r=1,2,$$

i.e.

$$\widehat{\varphi}_{1k}=\frac{\widehat{\boldsymbol{u}}_{1k}'\widehat{\boldsymbol{u}}_{1k}}{A_1-\tau_{1k}},\quad \widehat{\varphi}_{2k}=\frac{\widehat{\boldsymbol{u}}_{2k}'\boldsymbol{\Sigma}_{u_{2k}}^{-1}\widehat{\boldsymbol{u}}_{2k}}{\dim(\boldsymbol{u}_{2k})-\tau_{2k}},$$

where $\boldsymbol{\Sigma}_{u_{1k}} = \boldsymbol{I}_D$, $\boldsymbol{\Sigma}_{u_{2k}} = \underset{1\leq d\leq D}{\operatorname{diag}}\,(\Omega_d(\phi_k))$, $A_1 = D$, $A_2 = DT$, $\tau_{rk} = \frac{1}{\widehat{\varphi}_{rk}}\operatorname{tr}\big(\boldsymbol{\Sigma}_{u_{rk}}^{-1}\widehat{\mathsf{T}}_{rkk}^{rml}\big)$ and T_{rkk}^{rml} is the block (k,k) of dimension A_r of the matrix

$$\widehat{\mathsf{T}}_r^{rml} = \widehat{\mathsf{T}}_r + \widehat{\mathsf{T}}_r\mathsf{Z}_r'\boldsymbol{W}\boldsymbol{X}(\boldsymbol{X}'\widehat{\boldsymbol{V}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{W}\mathsf{Z}_r\widehat{\mathsf{T}}_r,$$

with $\widehat{\boldsymbol{V}} = \mathsf{Z}\widehat{\boldsymbol{\Sigma}}_u\mathsf{Z}' + \boldsymbol{W}^{-1}$, $\boldsymbol{\Sigma}_u = \operatorname{diag}(\boldsymbol{\Sigma}_{u_1}, \boldsymbol{\Sigma}_{u_2})$, $\widehat{\mathsf{T}}_r = (\mathsf{Z}_r\boldsymbol{W}\mathsf{Z}_r' + \widehat{\boldsymbol{\Sigma}}_{u_r}^{-1})^{-1}$ and

$\widehat{\boldsymbol{\Sigma}}_{u_1} = \operatorname{diag}(\varphi_{11}\boldsymbol{I}_{A_1}, \ldots, \varphi_{1q-1}\boldsymbol{I}_{A_1})$, $\widehat{\boldsymbol{\Sigma}}_{u_2} = \operatorname{diag}(\underset{1\leq d\leq D}{\operatorname{diag}}\,(\varphi_{21}\Omega_d(\phi_1)), \ldots, \underset{1\leq d\leq D}{\operatorname{diag}}\,(\varphi_{2q-1}\Omega_d(\phi_{q-1})))$.

The update of ϕ_k, $k = 1, \ldots, q-1$, can be calculated by applying the formula

$$\widehat{\phi}_k = -\frac{\widehat{\varphi}_{2k}^{-1}(\operatorname{tr}(\widehat{\mathsf{T}}_{2kk}^{rml}\boldsymbol{F}) + \widehat{\boldsymbol{u}}_{2k}'\boldsymbol{F}\widehat{\boldsymbol{u}}_{2k})}{\frac{2}{1-\widehat{\phi}_k^2} + 2\widehat{\varphi}_{2k}^{-1}(\operatorname{tr}(\widehat{\mathsf{T}}_{2kk}^{rml}\boldsymbol{E}) + \widehat{\boldsymbol{u}}_{2k}'\boldsymbol{E}\widehat{\boldsymbol{u}}_{2k})}.$$

The following calculations are useful.

$$\dot{\Omega}_d(\phi_k) = \frac{1}{1-\phi_k^2}\begin{pmatrix} 0 & 1 & \ldots & \ldots & (T-1)\phi_k^{T-2} \\ 1 & 0 & \ddots & & (T-2)\phi_k^{T-3} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ (T-2)\phi_k^{T-3} & & \ddots & 0 & 1 \\ (T-1)\phi_k^{T-2} & \ldots & \ldots & 1 & 0 \end{pmatrix} + \frac{2\phi_k\Omega_d(\phi_k)}{1-\phi_k^2},$$

$$\Omega_d^{-1}(\phi_k) = \begin{pmatrix} 1 & -\phi_k & 0 & \ldots & \ldots & 0 \\ -\phi_k & 1+\phi_k^2 & -\phi_k & 0 & & 0 \\ 0 & -\phi_k & 1+\phi_k^2 & -\phi_k & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & & \ddots & -\phi_k & 1+\phi_k^2 & -\phi_k \\ 0 & \ldots & \ldots & 0 & -\phi_k & 1 \end{pmatrix} = \boldsymbol{I}_T + \phi_k^2\boldsymbol{E} - \phi_k\boldsymbol{F},$$

where $\boldsymbol{I}_T$ is the identity matrix of dimension T, $\boldsymbol{E}$ is a diagonal matrix with diagonal elements $0, 1, \ldots, 1, 0$, and $\boldsymbol{F}$ is a matrix with the diagonal elements immediately above and below the main diagonal equal to -1 and the rest of elements equal to 0.

Algorithm B (alternative).

(B.1) Do $\ell = 1$. Initiate the values $\boldsymbol{\beta}^{(0)}$, $\boldsymbol{u}^{(0)}$ and $\boldsymbol{\sigma}^{(1)}$.

(B.2) Run the *Algorithm A* by using $\boldsymbol{\sigma}^{(\ell)}$ as known value of the vector of variance components, and $\boldsymbol{\beta}^{(\ell-1)}$ and $\mathsf{u}^{(\ell-1)}$ as initial values. Let $\boldsymbol{\beta}^{(\ell)}$ and $\mathsf{u}^{(\ell)}$ be the obtained estimators and predictors.

(B.3) Calculate $\boldsymbol{\eta}_{dt}^{(\ell)} = \boldsymbol{X}_{dt}\boldsymbol{\beta}^{(\ell)} + \mathsf{Z}_{1dt}\mathsf{u}_1^{(\ell)} + \mathsf{Z}_{2dt}\mathsf{u}_2^{(\ell)}$, $d = 1, \ldots, D$, $t = 1, \ldots, T$. Calculate

$$p_{dkt}^{(\ell)} = \frac{\exp(\eta_{dkt}^{(\ell)})}{1 + \sum_{k=1}^{q-1}\exp(\eta_{dkt}^{(\ell)})},\ \boldsymbol{p}_{dt}^{(\ell)} = \underset{1\leq k\leq q-1}{\operatorname{col}}(p_{dkt}^{(\ell)}),\ \boldsymbol{W}^{(\ell)} = \underset{1\leq d\leq D}{\operatorname{diag}}\,(\boldsymbol{W}_d),$$

$$\boldsymbol{W}_d = \underset{1\leq t\leq T}{\text{diag}}\,(\nu_{dt}[\text{diag}(\boldsymbol{p}_{dt}^{(\ell)}) - \boldsymbol{p}_{dt}^{(\ell)}\boldsymbol{p}_{dt}^{(\ell)\prime}]),\ \boldsymbol{\Sigma}_{u_1}^{(\ell)} = \text{diag}\big(\varphi_{11}^{(\ell)}\boldsymbol{I}_{A_1},\dots,\varphi_{1q-1}^{(\ell)}\boldsymbol{I}_{A_1}\big),$$

$$\boldsymbol{\Sigma}_{u_2}^{(\ell)} = \text{diag}\big(\underset{1\leq d\leq D}{\text{diag}}\,(\varphi_{21}^{(\ell)}\Omega_d(\phi_1^{(\ell)})),\dots,\underset{1\leq d\leq D}{\text{diag}}\,(\varphi_{2q-1}^{(\ell)}\Omega_d(\phi_{q-1}^{(\ell)}))\big),$$

$$\mathsf{T}_r^{(\ell)} = (\mathsf{Z}_r'\boldsymbol{W}\mathsf{Z}_r + \boldsymbol{\Sigma}_{u_r}^{(\ell)-1})^{-1},\ \boldsymbol{V}^{(\ell)} = \mathsf{Z}\boldsymbol{\Sigma}_u^{(\ell)}\mathsf{Z}' + \boldsymbol{W}^{(\ell)-1},$$

$$\mathsf{T}_r^{rml(\ell)} = \mathsf{T}_r^{(\ell)} + \mathsf{T}_r^{(\ell)}\mathsf{Z}_r'\boldsymbol{W}^{(\ell)}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{V}^{(\ell)-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{W}^{(\ell)}\mathsf{Z}_r\mathsf{T}_r^{(\ell)},\ \tau_{rk}^{(\ell)} = \big(\varphi_{rk}^{(\ell)}\big)^{-1}\text{tr}\big(\boldsymbol{\Sigma}_{u_{rk}}^{-1(\ell)}\mathsf{T}_{rkk}^{rml(\ell)}\big).$$

(B.4) Update φ_{rk} and ϕ_k using the equations

$$\widehat{\varphi}_{rk}^{(\ell+1)} = \frac{\widehat{\boldsymbol{u}}_{rk}^{(\ell)\prime}\boldsymbol{\Sigma}_{u_{rk}}^{-1(\ell)}\widehat{\boldsymbol{u}}_{rk}^{(\ell)}}{A_r - \tau_{rk}^{(\ell)}},\quad r=1,2,\ k=1,\dots,q-1,$$

$$\widehat{\phi}_k^{(\ell)} = -\frac{\widehat{\varphi}_{2k}^{-1(\ell)}(\text{tr}(\widehat{\mathsf{T}}_{2kk}^{rml(\ell)}\boldsymbol{F}) + \widehat{\boldsymbol{u}}_{2k}^{(\ell)\prime}\boldsymbol{F}\widehat{\boldsymbol{u}}_{2k}^{\ell})}{\frac{2}{1-\widehat{\phi}_k^{2(\ell-1)}} + 2\widehat{\varphi}_{2k}^{-1(\ell)}(\text{tr}(\widehat{\mathsf{T}}_{2kk}^{rml(\ell)}\boldsymbol{E}) + \widehat{\boldsymbol{u}}_{2k}^{(\ell)\prime}\boldsymbol{E}\widehat{\boldsymbol{u}}_{2k}^{\ell})},\quad k=1,\dots,q-1.$$

(B.5) Repeat the steps (B.2)-(B.4) until the convergence of $\boldsymbol{\beta}^{(\ell)}$, $\boldsymbol{u}^{(\ell)}$ (o $\mathsf{u}^{(\ell)}$) y $\boldsymbol{\sigma}^{(\ell)}$.

Like in chapter 4 and 5 the difference between this algorithm and the previous is that th last is a fixed-point algorithm and the previous is an iterative Fisher-Scoring algorithm.

The above described algorithms requires initial values of $\boldsymbol{\beta}$, $\boldsymbol{u}_1$, $\boldsymbol{u}_2$ and $\boldsymbol{\sigma}$. We suggest employing some easy-to-calculate estimates. More concretely, we use $\boldsymbol{u}_1^{(0)} = \boldsymbol{u}_2^{(0)} = \boldsymbol{0}$, $\phi_1 = \ldots = \phi_{q-1} = 0$ and $\boldsymbol{\beta}^{(0)} = \tilde{\boldsymbol{\beta}}$, where $\tilde{\boldsymbol{\beta}}$ is obtained by fitting the non mixed variant of the model without the random effects $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$. The non mixed model is also used for calculating φ_{rk}, by means of the formula

$$\varphi_{rk} = \frac{1}{2(D-1)}\sum_{d=1}^{D}\sum_{t=1}^{T}(\tilde{\eta}_{dkt}^{(dir)} - \tilde{\eta}_{dkt})^2,\quad k=1,\dots,q-1, r=1,2, \tag{6.2.1}$$

where

$$\tilde{\eta}_{dkt} = \tilde{\beta}_k x_{dkt},\quad \tilde{\eta}_{dkt}^{(dir)} = \log\frac{y_{dkt}}{y_{dqt}},\quad d=1,\dots,D,\ k=1,\dots,q-1,\ t=1,\dots,T.$$

We also propose using as initial values for ϕ_1, …,ϕ_{q-1} a consistent moment estimator proposed in Rao and Yu (1994) that is given by

$$\phi_k = \frac{\sum_{i=1}^{D}\sum_{t=1}^{T}\hat{o}_{dt}(\hat{o}_{d,t+1} - \hat{o}_{d,t+2})}{\sum_{i=1}^{D}\sum_{t=1}^{T-2}\hat{o}_{dt}(\hat{o}_{d,t} - \hat{o}_{d,t+1})},\quad k=1,\dots,q-1,$$

where

$$\hat{o}_{dt} = \tilde{\eta}_{dt}^{(dir)} - \tilde{\eta}_{dt}.$$

Under regularity conditions the asymptotic distribution of the REML-PQL estimators $\hat{\boldsymbol{\beta}}$ is multivariate normal $N(\boldsymbol{\beta}, \boldsymbol{F}^{\beta\beta})$, where $\boldsymbol{F}^{\beta\beta} = (q_{rr})_{r=1,\dots,l_k}$ is the block sub-matrix of the Fisher information matrix in the output of Algorithm A. Therefore, an approximate $(1-\alpha)$-level confidence interval for β_{kr} is

$$\hat{\beta}_{kr} \pm z_{\alpha/2} q_{rr}, \quad r = 1, \dots, l_k,$$

where z_α is the α-quantile of the normal distribution $N(0,1)$. If we use $\hat{\beta}_{kr}$ to test $H_0: \beta_{kr} = 0$ and we observe the realization $\hat{\beta}_{kr} = \beta_0$, the approximate p-value is

$$p = 2P_{H_0}(\hat{\beta}_{kr} > |\beta_0|) = 2P(Z > |\beta_0| / \sqrt{q_{rr}})$$

where Z follows a standard normal distribution..

6.3 Model-based small area estimation

Let us consider the population quantity $\boldsymbol{m}_{0dt} = N_{dt} p_{dt}$, where $N_{dt} = \#(P_{dt})$. In practice N_{dt} is unknown and it is estimated by combining administrative registers and population projections models, so we are more rigorously interested in estimating $\boldsymbol{m}_{dt} = \hat{N}_{dt} p_{dt}$. In the case of the SLFS, we use $\hat{N}_{dt} = \hat{N}_{dt}^{dir}$ because the sample weights are calibrated to population projections by sex and age groups at the province level. We estimate $\boldsymbol{m}_{dt}$ by means of $\hat{\boldsymbol{m}}_{dt} = \hat{N}_{dt} \hat{\boldsymbol{p}}_{dt}$, where

$$\hat{\boldsymbol{p}}_{dt} \;=\; \underset{1\le k\le q-1}{\text{col}}(\hat{p}_{dkt}), \quad \hat{p}_{dkt} = \frac{\exp\{\hat{\eta}_{dkt}\}}{1+\sum_{\ell=1}^{q-1}\exp\{\hat{\eta}_{d\ell t}\}}, \quad \hat{\eta}_{dkt} = \boldsymbol{x}_{dkt}\hat{\boldsymbol{\beta}}_k + \hat{u}_{1,dk} + \hat{u}_{2,dkt},$$

and $\hat{\boldsymbol{\beta}}_k$, $\hat{u}_{1,d}$ and $\hat{u}_{2,dt.}$ are obtained from the output of the fitting algorithm.

For deriving an approximation to the mean squared error of $\hat{m}_{dkt}$, we treat $\hat{N}_{dt}$ as a know constant. Let us write

$$m_{dkt} = h_{dkt}(\boldsymbol{\eta}_{dt}) = \hat{N}_{dt} p_{dkt} = \hat{N}_{dt} \frac{\exp\{\eta_{dkt}\}}{1+\sum_{\ell=1}^{q-1}\exp\{\eta_{d\ell t}\}}.$$

The partial derivatives of h_{dkt} are

$$\frac{\partial h_{dkt}}{\partial \eta_{dkt}} = \nu_{dt} p_{dkt}(1-p_{dkt}), \quad \frac{\partial h_{dt_1k}}{\partial \eta_{dt_2k}} = -\nu_{dt} p_{dt_1k} p_{dt_2k}, \quad k_1 \neq k_2.$$

We define

$$\begin{aligned}
\boldsymbol{\mu} &= h(\boldsymbol{\eta}) = \underset{1\le d\le D}{\text{col}}(\underset{1\le t\le T}{\text{col}}(\underset{1\le k\le q-1}{\text{col}}(h_{dkt}))),\\
\boldsymbol{H} &= \boldsymbol{W} = \boldsymbol{W}(\boldsymbol{\eta}) = \underset{1\le d\le D}{\text{diag}}(\boldsymbol{W}_d), \ \boldsymbol{W}_d = \boldsymbol{W}_{d1} + \boldsymbol{W}_{d2},\\
\boldsymbol{W}_{d1} &= \underset{1\le k\le q-1}{\text{diag}}(\underset{1\le t\le T}{\text{diag}}(\nu_{dt} p_{dkt})), \ \boldsymbol{W}_{d2} = \underset{1\le k_1,k_2\le q-1}{\text{matrix}}(\underset{1\le t\le T}{\text{diag}}(-\nu_{dt} p_{dk_1t} p_{dk_2t})).
\end{aligned}$$

In matrix notation, we have

$$h(\hat{\boldsymbol{\eta}}) - h(\boldsymbol{\eta}) \approx \boldsymbol{H}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}).$$

Let $\boldsymbol{\sigma} = (\varphi_{11}, \ldots, \varphi_{1q-1}, \varphi_{21}, \ldots, \varphi_{2q-1}, \phi_1, \ldots, \phi_{q-1}) = \underset{1\leq k\leq 3(q-1)}{\text{diag}}(\theta_k)$ the vector of variance components. As $\boldsymbol{m}_{d\cdot t} = \boldsymbol{A}_{dt}\boldsymbol{m}$, where $\boldsymbol{A}_{dt} = \underset{1\leq d_1\leq D}{\text{col}'}(\underset{1\leq k\leq q-1}{\text{diag}}(\underset{1\leq t_1\leq T}{\text{col}'}(\delta_{tt_1}\delta_{dd_1})))$, $\hat{\boldsymbol{\eta}}$ can be viewed as a vector of EBLUPs in the lineal mixed model (6.1.2), we propose applying the methodology of Prasad and Rao (1990) to approximate the MSE of $\hat{\boldsymbol{m}}_{d\cdot t}$. Then the MSE of $\hat{\boldsymbol{m}}_{d\cdot t}$ is approximated by

$$MSE(\hat{\boldsymbol{m}}_{dkt}) = \mathcal{G}_{1dt}(\boldsymbol{\sigma}) + \mathcal{G}_{2dt}(\boldsymbol{\sigma}) + \mathcal{G}_{3dt}(\boldsymbol{\sigma}),$$

where

$$\begin{aligned}
\mathcal{G}_{1dt}(\boldsymbol{\sigma}) &= \boldsymbol{A}_{dt}\boldsymbol{H}\boldsymbol{Z}\boldsymbol{T}\boldsymbol{Z}'\boldsymbol{H}'\boldsymbol{A}'_{dt},\\
\mathcal{G}_{2dt}(\boldsymbol{\sigma}) &= [\boldsymbol{A}_{dt}\boldsymbol{H}\boldsymbol{X} - \boldsymbol{A}_{dt}\boldsymbol{H}\boldsymbol{Z}\boldsymbol{T}\boldsymbol{Z}'\boldsymbol{W}\boldsymbol{X}]\boldsymbol{Q}[\boldsymbol{X}'\boldsymbol{H}'\boldsymbol{A}'_{dt} - \boldsymbol{X}'\boldsymbol{W}\boldsymbol{Z}\boldsymbol{T}\boldsymbol{Z}'\boldsymbol{H}'\boldsymbol{A}'_{dt}],\\
\mathcal{G}_{3dt}(\boldsymbol{\sigma}) &\approx \sum_{k_1=1}^{3(q-1)}\sum_{k_2=1}^{3(q-1)} \text{cov}(\hat{\theta}_{k_1}, \hat{\theta}_{k_2})\boldsymbol{A}_{dt}\boldsymbol{H}\boldsymbol{L}^{(k_1)}\boldsymbol{V}\boldsymbol{L}^{(k_2)\prime}\boldsymbol{H}'\boldsymbol{A}'_{dt}.
\end{aligned}$$

The covariance $\text{cov}(\hat{\theta}_{k_1}, \hat{\theta}_{k_2})$ is obtained from the inverse of the Fisher information matrix $\boldsymbol{F}$ at the output of the algorithm B.

$$\begin{aligned}
\boldsymbol{V} &= \text{var}(\boldsymbol{\eta}) = \boldsymbol{Z}\boldsymbol{V}_u\boldsymbol{Z}' + \boldsymbol{W}^{-1},\ \boldsymbol{T} = \boldsymbol{V}_u - \boldsymbol{V}_u\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{Z}\boldsymbol{V}_u,\ \boldsymbol{Q} = (\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\\
\boldsymbol{L}^{(k)} &= (\boldsymbol{I} - \boldsymbol{R}_1)\boldsymbol{V}_{1k}\boldsymbol{V}^{-1},\ \boldsymbol{V}_{1k} = \frac{\partial \boldsymbol{V}}{\partial \varphi_{1k}},\quad \boldsymbol{R}_1 = \boldsymbol{Z}_1\boldsymbol{V}_{u1}\boldsymbol{Z}_1'\boldsymbol{V}^{-1},\quad k = 1, \ldots, q-1,\\
\boldsymbol{L}^{(k)} &= (\boldsymbol{I} - \boldsymbol{R}_2)\boldsymbol{V}_{2k}\boldsymbol{V}^{-1},\ \boldsymbol{V}_{2k} = \frac{\partial \boldsymbol{V}}{\partial \varphi_{2k}},\quad \boldsymbol{R}_2 = \boldsymbol{V}_{u2}\boldsymbol{V}^{-1},\quad k = q, \ldots, 2(q-1),\\
\boldsymbol{L}^{(k)} &= (\boldsymbol{I} - \boldsymbol{R}_2)\boldsymbol{V}_{3k}\boldsymbol{V}^{-1},\ \boldsymbol{V}_{3k} = \frac{\partial \boldsymbol{V}}{\partial \phi_k},\quad \boldsymbol{R}_2 = \boldsymbol{V}_{u2}\boldsymbol{V}^{-1},\quad k = 2q-1, \ldots, 3(q-1).
\end{aligned}$$

The estimator of the MSE is

$$mse(\hat{\boldsymbol{m}}_{dkt}) = \mathcal{G}_{1dt}(\hat{\boldsymbol{\sigma}}) + \mathcal{G}_{2dt}(\hat{\boldsymbol{\sigma}}) + 2\mathcal{G}_{3dt}(\hat{\boldsymbol{\sigma}}).$$

The elements of the formula $\mathcal{G}_{1dt}(\boldsymbol{\sigma}) = \boldsymbol{A}_{dt}\boldsymbol{H}\boldsymbol{Z}\boldsymbol{T}\boldsymbol{Z}'\boldsymbol{H}'\boldsymbol{A}'_{dt}$ are

$$\begin{aligned}
\boldsymbol{A}_{dt} &= \underset{1\leq d_1\leq D}{\text{col}'}(\underset{1\leq k\leq q-1}{\text{diag}}(\underset{1\leq t_1\leq T}{\text{col}'}(\delta_{tt_1}\delta_{dd_1}))),\quad \boldsymbol{H} = \boldsymbol{W},\\
\boldsymbol{T} &= \boldsymbol{V}_u - \boldsymbol{V}_u\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{Z}\boldsymbol{V}_u = \begin{pmatrix} \boldsymbol{T}_{11} & \boldsymbol{T}_{12}\\ \boldsymbol{T}_{21} & \boldsymbol{T}_{22}\end{pmatrix},\ \boldsymbol{T}_{12} = -\boldsymbol{V}_{u_1}\boldsymbol{Z}_1'\boldsymbol{V}^{-1}\boldsymbol{Z}_2\boldsymbol{V}_{u_2} = \boldsymbol{T}'_{21},\\
\boldsymbol{T}_{11} &= \boldsymbol{V}_{u_1} - \boldsymbol{V}_{u_1}\boldsymbol{Z}_1'\boldsymbol{V}^{-1}\boldsymbol{Z}_2\boldsymbol{V}_{u_1},\quad \boldsymbol{T}_{22} = \boldsymbol{V}_{u_2} - \boldsymbol{V}_{u_2}\boldsymbol{Z}_2'\boldsymbol{V}^{-1}\boldsymbol{Z}_2\boldsymbol{V}_{u_2}.
\end{aligned}$$

We have that $\boldsymbol{T}_{ab} = \underset{1\leq d\leq D}{\text{diag}}(\boldsymbol{T}_{abd})$, $a, b = 1, 2$, where

$$\begin{aligned}
\boldsymbol{T}_{11d} &= \underset{1\leq k\leq q-1}{\text{diag}}(\varphi_{1k}) - \underset{1\leq k\leq q-1}{\text{diag}}(\varphi_{1k})\boldsymbol{Z}'_{1d}\boldsymbol{V}_d^{-1}\boldsymbol{Z}_{1d}\underset{1\leq k\leq q-1}{\text{diag}}(\varphi_{1k}),\\
\boldsymbol{T}_{12d} &= -\underset{1\leq k\leq q-1}{\text{diag}}(\varphi_{1k})\boldsymbol{Z}'_{1d}\boldsymbol{V}_d^{-1}\boldsymbol{Z}_{1d}\underset{1\leq k\leq q-1}{\text{diag}}(\boldsymbol{V}_{u_{2,dk}}),\\
\boldsymbol{T}_{22d} &= \underset{1\leq k\leq q-1}{\text{diag}}(\boldsymbol{V}_{u_{2,dk}}) - \underset{1\leq k\leq q-1}{\text{diag}}(\boldsymbol{V}_{u_{2,dk}})\boldsymbol{Z}'_{2d}\boldsymbol{V}_d^{-1}\boldsymbol{Z}_{2d}\underset{1\leq k\leq q-1}{\text{diag}}(\boldsymbol{V}_{u_{2,dk}}),
\end{aligned}$$

We calculate this product $\boldsymbol{ZTZ}'$

$$\boldsymbol{ZTZ}' = \boldsymbol{Z}_1\boldsymbol{T}_{11}\boldsymbol{Z}_1' + \boldsymbol{Z}_1\boldsymbol{T}_{12}\boldsymbol{Z}_2' + \boldsymbol{Z}_2\boldsymbol{T}_{21}\boldsymbol{Z}_1' + \boldsymbol{Z}_2\boldsymbol{T}_{22}\boldsymbol{Z}_2' = \boldsymbol{M}_{11} + \boldsymbol{M}_{12} + \boldsymbol{M}_{12}' + \boldsymbol{M}_{22}.$$

We have $\boldsymbol{M}_{ab} = \underset{1\le d_1\le D}{\text{diag}}(\boldsymbol{M}_{abd_1})$, where $\boldsymbol{M}_{abd_1} = \boldsymbol{Z}_{ad_1}\boldsymbol{T}_{abd_1}\boldsymbol{Z}_{bd_1}'$, $a, b = 1, 2$;

$$\begin{aligned}
\boldsymbol{M}_{11d} &= \underset{1\le k\le q-1}{\text{diag}}(\varphi_{1k}\boldsymbol{1}_T\boldsymbol{1}_T') - \underset{1\le k\le q-1}{\text{diag}}(\varphi_{1k}\boldsymbol{1}_T)\boldsymbol{Z}_{1d}'\boldsymbol{V}_d^{-1}\boldsymbol{Z}_{1d}\underset{1\le k\le q-1}{\text{diag}}(\varphi_{1k}\boldsymbol{1}_T'),\\
\boldsymbol{M}_{12d} &= \underset{1\le k\le q-1}{\text{diag}}(\varphi_{1k}\boldsymbol{1}_T)\boldsymbol{Z}_{1d}'\boldsymbol{V}_d^{-1}\boldsymbol{Z}_{1d}\underset{1\le k\le q-1}{\text{diag}}(\boldsymbol{V}_{u_{2,dk}}),\\
\boldsymbol{M}_{22d} &= \underset{1\le k\le q-1}{\text{diag}}(\boldsymbol{V}_{u_{2,dk}}) - \underset{1\le k\le q-1}{\text{diag}}(\boldsymbol{V}_{u_{2,dk}})\boldsymbol{Z}_{2d}'\boldsymbol{V}_d^{-1}\boldsymbol{Z}_{2d}\underset{1\le k\le q-1}{\text{diag}}(\boldsymbol{V}_{u_{2,dk}}),
\end{aligned}$$

Finally

$$\mathcal{G}_{1dt}(\boldsymbol{\sigma}) = \boldsymbol{A}_{dt}\boldsymbol{H}[\boldsymbol{M}_{11} + \boldsymbol{M}_{12} + \boldsymbol{M}_{12}' + \boldsymbol{M}_{22}]\boldsymbol{H}'\boldsymbol{A}_{dt}' = \boldsymbol{G}_{11} + \boldsymbol{G}_{12} + \boldsymbol{G}_{12}' + \boldsymbol{G}_{22},$$

where

$$\begin{aligned}
\boldsymbol{G}_{ab} &= \boldsymbol{A}_{dt}\boldsymbol{H}\boldsymbol{M}_{ab}\boldsymbol{H}'\boldsymbol{A}_{dt}' = \underset{1\le d_1\le D}{\text{col}'}(\delta_{dd_1}\underset{1\le k\le q-1}{\text{diag}}(\underset{1\le t_1\le T}{\text{col}'}(\delta_{tt_1})))\\
&\cdot \underset{1\le d_1\le D}{\text{diag}}(\boldsymbol{W}_{d_1}\boldsymbol{M}_{abd_1}\boldsymbol{W}_{d_1}')\underset{1\le d_1\le D}{\text{col}}(\delta_{dd_1}\underset{1\le k\le q-1}{\text{diag}}(\underset{1\le t_1\le T}{\text{col}}(\delta_{tt_1})))\\
&= \underset{1\le k\le q-1}{\text{diag}}(\underset{1\le t_1\le T}{\text{col}'}(\delta_{tt_1}))\boldsymbol{W}_d\boldsymbol{M}_{abd}\boldsymbol{W}_d'\underset{1\le k\le q-1}{\text{diag}}(\underset{1\le t_1\le T}{\text{col}}(\delta_{tt_1})).
\end{aligned}$$

The expression of $\mathcal{G}_{2dt}(\boldsymbol{\sigma})$ is

$$\begin{aligned}
\mathcal{G}_{2dt}(\boldsymbol{\sigma}) &= [\boldsymbol{A}_{dt}\boldsymbol{HX} - \boldsymbol{A}_{dt}\boldsymbol{HZTZ}'\boldsymbol{WX}]\boldsymbol{Q}[\boldsymbol{X}'\boldsymbol{H}'\boldsymbol{A}_{dt}' - \boldsymbol{X}'\boldsymbol{WZTZ}'\boldsymbol{H}'\boldsymbol{A}_{dt}']\\
&= [\boldsymbol{A}_{21} - \boldsymbol{A}_{22}]\boldsymbol{Q}[\boldsymbol{A}_{21}' - \boldsymbol{A}_{22}'],
\end{aligned}$$

where

$$\begin{aligned}
\boldsymbol{A}_{21} &= \boldsymbol{A}_{dt}\boldsymbol{HX} = \underset{1\le k\le q-1}{\text{diag}}(\underset{1\le t_1\le T}{\text{col}'}(\delta_{tt_1}))\boldsymbol{X}_d,\\
\boldsymbol{A}_{22} &= \boldsymbol{A}_{dt}\boldsymbol{HZTZ}'\boldsymbol{WX} = \underset{1\le k\le q-1}{\text{diag}}(\underset{1\le t_1\le T}{\text{col}'}(\delta_{tt_1}))\boldsymbol{W}_d(\boldsymbol{M}_{11d} + \boldsymbol{M}_{12d} + \boldsymbol{M}_{21d} + \boldsymbol{M}_{22d})\boldsymbol{W}_d\boldsymbol{X}_d.
\end{aligned}$$

For the calculation of $\mathcal{G}_{3dt}(\boldsymbol{\sigma})$ are

$$\begin{aligned}
\boldsymbol{L}^{(k)} &= (\boldsymbol{I} - \boldsymbol{R}_1)\boldsymbol{V}_{1k}\boldsymbol{V}^{-1}, \quad \boldsymbol{V}_{1k} = \frac{\partial \boldsymbol{V}}{\partial \varphi_{1k}}, \quad \boldsymbol{R}_1 = \boldsymbol{Z}_1\boldsymbol{V}_{u1}\boldsymbol{Z}_1'\boldsymbol{V}^{-1}, \quad k = 1, \ldots, q-1,\\
\boldsymbol{L}^{(k)} &= (\boldsymbol{I} - \boldsymbol{R}_2)\boldsymbol{V}_{2k}\boldsymbol{V}^{-1}, \quad \boldsymbol{V}_{2k} = \frac{\partial \boldsymbol{V}}{\partial \varphi_{2k}}, \quad \boldsymbol{R}_2 = \boldsymbol{V}_{u2}\boldsymbol{V}^{-1}, \quad k = q, \ldots, 2(q-1),\\
\boldsymbol{L}^{(k)} &= (\boldsymbol{I} - \boldsymbol{R}_2)\boldsymbol{V}_{3k}\boldsymbol{V}^{-1}, \quad \boldsymbol{V}_{3k} = \frac{\partial \boldsymbol{V}}{\partial \phi_k}, \quad \boldsymbol{R}_2 = \boldsymbol{V}_{u2}\boldsymbol{V}^{-1}, \quad k = 2q-1, \ldots, 3(q-1).
\end{aligned}$$

The covariance $\text{cov}(\hat{\theta}_{k_1}, \hat{\theta}_{k_2})$ are obtained from the inverse of the Fisher information matrix at the output of the algorithm B.

Concerning the estimation of the MSE of $\hat{m}_{dkt}$, we can also use the approach of González-Manteiga et al. (2008a, 2008b) by introducing the following parametric bootstrap method.

1. Fit the model (6.1.1)-(6.1.2) and calculate $\hat{\varphi}_{1k}$, $\hat{\varphi}_{2k}$, $\hat{\phi}_k$ and $\hat{\boldsymbol{\beta}}_k$, $k = 1, \ldots, q-1$.
2. For $d = 1, \ldots, D$, $t = 1, \ldots, T$, generate the random effects $\boldsymbol{u}^*_{1,d} \sim N(\mathbf{0}, \underset{1\leq k\leq q-1}{\text{diag}} (\hat{\varphi}_{1k}))$ and $\boldsymbol{u}^*_{2,dk} \sim N(\mathbf{0}, \varphi_{2k}\Omega_d(\phi_k))$, and the response variable $\boldsymbol{y}^*_{dt} \sim \text{M}(\nu_{dt}, p^*_{d1t}, \ldots, p^*_{dqt-1})$, where
$$p^*_{dkt} = \frac{\exp\{\eta^*_{dkt}\}}{1 + \sum_{\ell=1}^{q-1} \exp\{\eta^*_{d\ell t}\}}, \quad \eta^*_{dkt} = \hat{\boldsymbol{\beta}}_k x_{dkt} + u^*_{1,dk} + u^*_{2,dkt}, \quad m^*_{dkt} = \hat{N}_{dt} p^*_{dkt}.$$
3. For $d = 1, \ldots, D$, $t = 1 \ldots, T$, $k = 1, \ldots, q-1$, calculate $\hat{\varphi}^*_{1k}$, $\hat{\varphi}^*_{2k}$, $\hat{\phi}^*_k$, $\hat{\boldsymbol{\beta}}^*_k$,
$$\hat{p}^*_{dkt} = \frac{\exp\{\hat{\eta}^*_{dkt}\}}{1 + \sum_{\ell=1}^{q-1} \exp\{\hat{\eta}^*_{d\ell t}\}}, \quad \hat{\eta}^*_{dkt} = \hat{\boldsymbol{\beta}}^*_k x_{dkt} + \hat{u}^*_{1,dk} + \hat{u}^*_{2,dkt}, \quad \hat{m}^*_{dkt} = \hat{N}_{dt} \hat{p}^*_{dkt}.$$
4. Repeat B times steps 2-3 and calculate the bootstrap mean square error estimator
$$mse^{*1}_{dkt} = \frac{1}{B} \sum_{b=1}^{B} (\hat{m}^*_{dkt} - m^*_{dkt})^2, \quad d = 1, \ldots, D, \; k = 1, \ldots, q-1, \; t = 1 \ldots, T.$$

In this chapter we only use one bootstrap estimates for the MSE because in previous chapters we check that the parametric bootstrap has better performance.

6.4 Simulation study

In this section we present two simulation experiments. The first experiment is designed to analyze the behavior of the estimators $\hat{\boldsymbol{\beta}}_k$, $\hat{\varphi}_{1k}$, $\hat{\varphi}_{2k}$ and $\hat{m}_{dkt} = \hat{N}_{dt} p_{dkt}$. The second simulation studies the behavior of the proposed MSE estimators.

6.4.1 Sample simulation

We take $\hat{N}_{dt} = 1000$ and we consider a multinomial logit mixed model with three model categories ($q-1 = 2$). For $d = 1, \ldots, D$, $k = 1, 2$ and $t = 1, \ldots, T$, we generate the explanatory variables
$$U_{dkt} = \frac{1}{3}\left(\frac{d-D}{D} + \frac{k}{q-1} + \frac{t}{T}\right), \; x_{d1t} = \mu_1 + \sigma^{1/2}_{x11} U_{d1t}, \; x_{d2t} = \mu_2 + \sigma^{1/2}_{x22}\big[\rho_x U_{d1} + \sqrt{1-\rho_x^2} U_{d2t}\big],$$
where $\mu_1 = \mu_2 = 1$, $\sigma_{x11} = 1$, $\sigma_{x22} = 2$ and $\rho_x = 0$. The random effects are $u_{1,dk} \sim N(0, \varphi_{1k})$ with $\varphi_{11} = 1$, $\varphi_{12} = 2$ and $u_{2,dk} \sim N(\mathbf{0}, \boldsymbol{V}_{u_{2,dk}})$ with $\varphi_{21} = 0.25$, $\varphi_{22} = 0.5$ and
$$\Omega_d(\phi_k) = \Omega_{d,k} = \frac{1}{1-\phi_k^2} \begin{pmatrix} 1 & \phi_k & \ldots & \phi_k^{T-2} & \phi_k^{T-1} \\ \phi_k & 1 & \ddots & & \phi_k^{T-2} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \phi_k^{T-2} & & \ddots & 1 & \phi_k \\ \phi_k^{T-1} & \phi_k^{T-2} & \ldots & \phi_k & 1 \end{pmatrix}_{T\times T}.$$

We take $\phi_1 = 0.5$ and $\phi_2 = 0.75$.

The target variable is $\boldsymbol{y}_{dt} = (y_{d1t}, y_{d2t})' \sim \mathrm{M}(\nu_{dt}, p_{d1t}, p_{d2t})$, where

$$p_{dkt} = \frac{\exp\{\eta_{dkt}\}}{1 + \exp\{\eta_{d1t}\} + \exp\{\eta_{d2t}\}}, \quad \eta_{dkt} = \beta_{0k} + \beta_{1k} x_{dkt} + u_{1,dk} + u_{2,dkt}, \tag{6.4.1}$$

$\nu_{dt} = 100$, $\beta_{01} = 1.3$, $\beta_{02} = -1$, $\beta_{11} = -1.6$ and $\beta_{12} = 1$.

If $T = 2$, we can write the model (6.4.1) in the matrix form

$$\left(\begin{array}{c} \eta_{111} \\ \eta_{121} \\ \hline \eta_{112} \\ \eta_{122} \\ \hline \vdots \\ \hline \eta_{D11} \\ \eta_{D21} \\ \hline \eta_{D12} \\ \eta_{D22} \end{array}\right) = \left(\begin{array}{cccc} 1 & x_{111} & 0 & 0 \\ 0 & 0 & 1 & x_{121} \\ \hline 1 & x_{112} & 0 & 0 \\ 0 & 0 & 1 & x_{122} \\ \hline \vdots & & & \\ \hline 1 & x_{D11} & 0 & 0 \\ 0 & 0 & 1 & x_{D21} \\ \hline 1 & x_{D12} & 0 & 0 \\ 0 & 0 & 1 & x_{D22} \end{array}\right) \begin{pmatrix} \beta_{01} \\ \beta_{11} \\ \beta_{02} \\ \beta_{12} \end{pmatrix} + \left(\begin{array}{c} u_{1,11} \\ u_{1,12} \\ \hline u_{1,11} \\ u_{1,12} \\ \hline \vdots \\ \hline u_{1,D1} \\ u_{1,D2} \\ \hline u_{1,D1} \\ u_{1,D2} \end{array}\right) + \left(\begin{array}{c} u_{2,111} \\ u_{2,121} \\ \hline u_{2,112} \\ u_{2,122} \\ \hline \vdots \\ \hline u_{2,D11} \\ u_{2,D21} \\ \hline u_{2,D12} \\ u_{2,D22} \end{array}\right),$$

or, equivalently, in the more concise notation

$$\begin{pmatrix} \boldsymbol{\eta}_{11} \\ \boldsymbol{\eta}_{12} \\ \vdots \\ \boldsymbol{\eta}_{D1} \\ \boldsymbol{\eta}_{D2} \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}_{11} \\ \boldsymbol{X}_{12} \\ \vdots \\ \boldsymbol{X}_{D1} \\ \boldsymbol{X}_{D2} \end{pmatrix} \boldsymbol{\beta} + \boldsymbol{Z}_1 \boldsymbol{u}_1 + \boldsymbol{Z}_2 \boldsymbol{u}_2,$$

where $\boldsymbol{Z}_1 = \underset{1 \le d \le D}{\operatorname{diag}} (\boldsymbol{Z}_{1d})$, $\boldsymbol{Z}_2 = \boldsymbol{I}_{2D}$,

$$\boldsymbol{\eta}_{dt} = \begin{pmatrix} \eta_{d1t} \\ \eta_{d2t} \end{pmatrix}, \ \boldsymbol{X}_{dt} = \begin{pmatrix} 1 & x_{d1t} & 0 & 0 \\ 0 & 0 & 1 & x_{d2t} \end{pmatrix}, \ \boldsymbol{I}_{2D} = \begin{pmatrix} \boldsymbol{I}_2 & & 0 \\ & \ddots & \\ 0 & & \boldsymbol{I}_2 \end{pmatrix}, \ \boldsymbol{Z}_{1d} = \begin{pmatrix} \boldsymbol{Z}_{1d1} \\ \boldsymbol{Z}_{1d2} \end{pmatrix},$$

$$\boldsymbol{Z}_{1dt} = \boldsymbol{I}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \ \boldsymbol{u}_1 \sim N(0, \boldsymbol{\varphi}_1 \boldsymbol{I}_{2D}), \ \boldsymbol{u}_{2dk} \sim N(0, \boldsymbol{\varphi}_{2k} \Omega_d(\phi_k)).$$

6.4.2 Simulation experiment 1

The objective of this experiment is to analyze the behavior of the estimators of $\boldsymbol{\beta}_k$, φ_{1k}, φ_{2k}, ϕ_k and m_{dkt}. As efficiency measures, we use the relative empirical bias (RBIAS) and the relative mean squared error (RMSE). The simulation is described below.

1. Repeat $I = 1000$ times $(i = 1, \dots, 1000)$

   1.1. Generate (y_{dkt}, x_{dkt}), $d = 1, \dots, D$, $k = 1, 2$, $t = 1, 2$.

   1.2 Calculate $\mu_{dkt} = \nu_{dt} p_{dkt}$, $\boldsymbol{\mu}_{dt} = (\mu_{d1t}, \mu_{d2t})'$.

1.3. Calculate $\widehat{\beta}_{jk}^{(i)}$, $\widehat{\varphi}_{rk}^{(i)}$, $j = 0,1$, $k = 1,2$, $r = 1,2$, and $\hat{\mu}_{dkt} = \nu_{dt} p_{dkt}$, $\hat{\boldsymbol{\mu}}_{dt} = (\hat{\mu}_{d1t}, \hat{\mu}_{d2t})'$.

2. Output:

$$BIAS(\widehat{\beta}_{jk}) = \frac{1}{I}\sum_{i=1}^{I}(\widehat{\beta}_{jk}^{(i)} - \beta_{jk}),\ BIAS(\widehat{\varphi}_{rk}) = \frac{1}{I}\sum_{i=1}^{I}(\widehat{\varphi}_{rk}^{(i)} - \varphi_k),\ j = 0,1, k = 1,2, r = 1,2.$$

$$MSE(\widehat{\beta}_{jk}) = \frac{1}{I}\sum_{i=1}^{I}(\widehat{\beta}_{jk}^{(i)} - \beta_k)^2,\ MSE(\widehat{\varphi}_{rk}) = \frac{1}{I}\sum_{i=1}^{I}(\widehat{\varphi}_{rk}^{(i)} - \varphi_k)^2,\ j = 0,1, k = 1,2, r = 1,2.$$

$$MSE_{dt} = \frac{1}{I}\sum_{i=1}^{I}(\hat{\boldsymbol{\mu}}_{dt}^{(i)} - \boldsymbol{\mu}_{dt}^{(i)})(\hat{\boldsymbol{\mu}}_{dt}^{(i)} - \boldsymbol{\mu}_{dt}^{(i)})', \quad d = 1, D/2, D,\ t = 2.$$

$$RBIAS(\widehat{\beta}_{jk}) = \frac{BIAS(\widehat{\beta}_{jk})}{|\beta_{jk}|},\ RBIAS(\widehat{\varphi}_{rk}) = \frac{BIAS(\widehat{\varphi}_{rk})}{\varphi_k},\ j = 0,1, k = 1,2, r = 1,2.$$

$$RMSE(\widehat{\beta}_{jk}) = \frac{\sqrt{MSE(\widehat{\beta}_{jk})}}{|\beta_{jk}|},\ RMSE(\widehat{\varphi}_{rk}) = \frac{\sqrt{MSE(\widehat{\varphi}_{rk})}}{\varphi_k},\ j = 0,1, k = 1,2, r = 1,2.$$

$$MEAN_{dt} = \frac{1}{I}\sum_{i=1}^{I}\hat{\boldsymbol{\mu}}_{dt}^{(i)} \quad RMSE_{dt} = \frac{\sqrt{MSE_{dt}}}{|MEAN_{dt}|}, \quad d = 1, D/2, D,\ t = 2.$$

Table 6.4.1 gives the RMSE-values of the model parameter estimators for $D = 100$. As T increases from 4 to 12 we observe a reduction in RMSE for the $\boldsymbol{\beta}_k$'s. This reduction is much higher for $\boldsymbol{\varphi}_k$'s and for the ϕ_k. This simulation suggests that the proposed multinomial mixed model should be used when the number of time periods is greater than eight.

T	4	8	12	T	4	8	12
$RMSE(\widehat{\beta}_{01})$	0.33	0.28	0.22	$RBIAS(\widehat{\beta}_{01})$	-0.04	-0.03	0.01
$RMSE(\widehat{\beta}_{02})$	0.34	0.28	0.23	$RBIAS(\widehat{\beta}_{02})$	-0.05	-0.02	0.01
$RMSE(\widehat{\beta}_{11})$	0.34	0.36	0.31	$RBIAS(\widehat{\beta}_{11})$	-0.04	-0.06	-0.06
$RMSE(\widehat{\beta}_{12})$	0.34	0.35	0.32	$RBIAS(\widehat{\beta}_{12})$	-0.03	-0.05	-0.05
$RMSE(\widehat{\varphi}_{11})$	0.17	0.17	0.15	$RBIAS(\widehat{\varphi}_{11})$	0.02	-0.04	-0.05
$RMSE(\widehat{\varphi}_{12})$	0.25	0.18	0.18	$RBIAS(\widehat{\varphi}_{12})$	0.15	0.07	-0.05
$RMSE(\widehat{\varphi}_{21})$	0.39	0.19	0.13	$RBIAS(\widehat{\varphi}_{21})$	-0.39	-0.18	-0.11
$RMSE(\widehat{\varphi}_{22})$	0.39	0.18	0.12	$RBIAS(\widehat{\varphi}_{22})$	-0.39	-0.17	-0.11
$RMSE(\widehat{\phi}_{11})$	0.94	0.50	0.30	$RBIAS(\widehat{\phi}_{11})$	-0.57	-0.50	-0.29
$RMSE(\widehat{\phi}_{12})$	0.79	0.39	0.24	$RBIAS(\widehat{\phi}_{12})$	-0.38	-0.33	-0.19

Table 6.4.1. MSE, RMSE, BIAS and RBIAS for $D = 100$.

Table 6.4.2 shows the RMSE and RBIAS values of $\hat{m}_{dkt}$ for $D = 100$ in different periods of time. We give the behaviour for three particular small area $d = 1, D/2, D$. We observe that all the RMSE values are below 13%, which indicates a good behavior.

	T	4	8	12		T	4	8	12
$RMSE_{1d1t}$	1	0.09	0.09	0.09	$RBIAS_{1d1t}$	1	0.004	-0.018	0.004
	$D/2$	0.11	0.11	0.11		$D/2$	-0.008	0.001	0.005
	D	0.13	0.14	0.12		D	0.002	0.015	0.010
$RMSE_{2d2t}$	1	0.12	0.12	0.13	$RBIAS_{1d1t}$	1	0.004	0.017	0.012
	$D/2$	0.09	0.10	0.10		$D/2$	0.005	0.009	-0.011
	D	0.09	0.08	0.08		D	-0.002	0.005	-0.003

Table 6.4.2. MSE, RMSE, BIAS and RBIAS for $D = 100$.

6.4.3 Simulation experiment 2

The simulation experiment is designed to study the behavior of the two mean square error estimators (analytic and bootstrap). In this case we take $D = 50$ and $T = 2, 4, 6$. We take $N_{dt} = 1000$ and we consider a multinomial logit mixed model with three model categories and the last as the reference ($q - 1 = 2$). For $d = 1, \dots, D$, $k = 1, 2$ and $t = 1 \dots, T$, we generate the explanatory variables

$$U_{dkt} = \frac{1}{3}\left(\frac{d-D}{D} + \frac{k}{q-1} + \frac{t}{T}\right),\ x_{d1t} = \mu_1 + \sigma_{x11}^{1/2} U_{d1t},\ x_{d2t} = \mu_2 + \sigma_{x22}^{1/2}\left[\rho_x U_{d1} + \sqrt{1-\rho_x^2}\, U_{d2t}\right],$$

where $\mu_1 = \mu_2 = 1$, $\sigma_{x11} = 1$, $\sigma_{x22} = 2$ and $\rho_x = 0$. The random effects are $u_{1,dk} \sim N(0, \varphi_{1k})$ with $\varphi_{11} = 1$, $\varphi_{12} = 2$ and $\boldsymbol{u}_{2,dk} \sim N(\boldsymbol{0}, \boldsymbol{V}_{u_{2,dk}})$ with $\varphi_{21} = 0.25$, $\varphi_{22} = 0.5$ and

$$\Omega_d(\phi_k) = \Omega_{d,k} = \frac{1}{1-\phi_k^2}\begin{pmatrix} 1 & \phi_k & \dots & \phi_k^{T-2} & \phi_k^{T-1} \\ \phi_k & 1 & \ddots & & \phi_k^{T-2} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \phi_k^{T-2} & & \ddots & 1 & \phi_k \\ \phi_k^{T-1} & \phi_k^{T-2} & \dots & \phi_k & 1 \end{pmatrix}_{T\times T}.$$

We take $\phi_1 = 0.5$ and $\phi_2 = 0.75$

The target variable is $\boldsymbol{y}_{dt} = (y_{d1t}, y_{d2t})' \sim \text{M}(\nu_{dt}, p_{d1t}, p_{d2t})$, where

$$p_{dkt} = \frac{\exp\{\eta_{dkt}\}}{1 + \exp\{\eta_{d1t}\} + \exp\{\eta_{d2t}\}},\ \eta_{dkt} = \beta_{0k} + \beta_{1k} x_{dkt} + u_{1,dk} + u_{2,dkt}. \tag{6.4.2}$$

$\nu_{dt} = 100$, $\beta_{01} = 1.3$, $\beta_{02} = -1$, $\beta_{11} = -1.6$ and $\beta_{12} = 1$.

The steps of the simulation are

1. Repeat $I = 500$ times ($i = 1, \dots, 500$)

   1.1. For $d = 1, \dots, 50$, $k = 1, 2$, $t = 1 \dots, T$, generate $(y_{dkt}^{(i)}, x_{dkt}^{(i)})$.

   1.2. For $d = 1, \dots, D$, $k = 1, 2$, $t = 1 \dots, T$, calculate $\hat{p}_{dkt}^{(i)}$, $\hat{m}_{dkt}^{(i)}$, $\hat{\boldsymbol{\sigma}}^{(i)}$, $\hat{\boldsymbol{\beta}}^{(i)}$ and

   $$mse_{dkt}^{(i)} = \mathcal{G}_{1dkt}^{(i)}(\hat{\boldsymbol{\sigma}}^{(i)}) + \mathcal{G}_{2dkt}^{(i)}(\hat{\boldsymbol{\sigma}}^{(i)}) + 2\mathcal{G}_{dkt}^{(i)}(\hat{\boldsymbol{\sigma}}^{(i)}).$$

1.3. Repeat $B = 500$ times ($b = 1, \ldots, B$)

1.3.1. For $d = 1, \ldots, D$, $k = 1, 2$, $t = 1 \ldots, T$, generate $\boldsymbol{u}_{1,d}^{*(ib)}$, $\boldsymbol{u}_{2,dt}^{*(ib)}$,

$$\boldsymbol{y}_{dt}^{*(ib)} = (y_{1dkt}^{*(ib)}, y_{2dkt}^{*(ib)})' \sim \mathrm{M}(\nu_{dt}, p_{d1t}^{*(ib)}, p_{d2t}^{*(ib)}),$$

where

$$p_{dkt}^{*(ib)} = \frac{\exp\{\eta_{dkt}^{*(ib)}\}}{1 + \exp\{\eta_{d1t}^{*(ib)}\} + \exp\{\eta_{d2t}^{*(ib)}\}}, \quad \eta_{dkt}^{*(ib)} = \hat{\beta}_{0k}^{(i)} + \hat{\beta}_{1k}^{(i)} x_{dkt}^{(i)} + u_{1,dk}^{*(ib)} + u_{2,dkt}^{*(ib)}.$$

1.3.2. For $d = 1, \ldots, D$, $k = 1, 2$, $t = 1 \ldots, T$, calculate $m_{dkt}^{*(ib)}$, $\hat{\boldsymbol{\sigma}}^{*(ib)}$, $\hat{\boldsymbol{\beta}}^{*(ib)}$, $\hat{m}_{dkt}^{*(ib)}$.

1.4 For $d = 1, \ldots, D$, $k = 1, 2$, $t = 1 \ldots, T$, calculate

$$mse_{dkt}^{*(i)} = \frac{1}{B} \sum_{b=1}^{B} (\hat{m}_{dkt}^{*(ib)} - m_{dkt}^{*(ib)})^2,$$

2. Output: mse_{dkt}, mse_{dkt}^{*}, $d = 1, \ldots, 50$, $k = 1, 2$, $t = 1, 2$, $i = 1, \ldots, 500$.

Figure 6.4.1 presents the box-plots of the values of the two simulated estimators $mse_{dkt}^{(i)}, mse_{dkt}^{1(i)}$, $k = 1, 2$, $i = 1, \ldots, 500$, for $d = 25$ and $t = T$, with $T = 2, 4, 6$ and $D = 50$. The first column is for $T = 2$, second column is for $T = 4$ and the last column is for $T = 6$. The true MSE is plotted in a horizontal line. It has been calculated by the Monte Carlo formula

$$MSE_{dkt} = \frac{1}{I} \sum_{i=1}^{I} (\hat{m}_{dkt}^{(i)} - m_{dkt}^{(i)})^2,$$

under $I = 1000$ iterations of the simulation experiment, excluding the bootstrap step. We observe that the analytic estimator mse has larger variability than bootstrap estimator mse^* and the best estimator is mse^*.

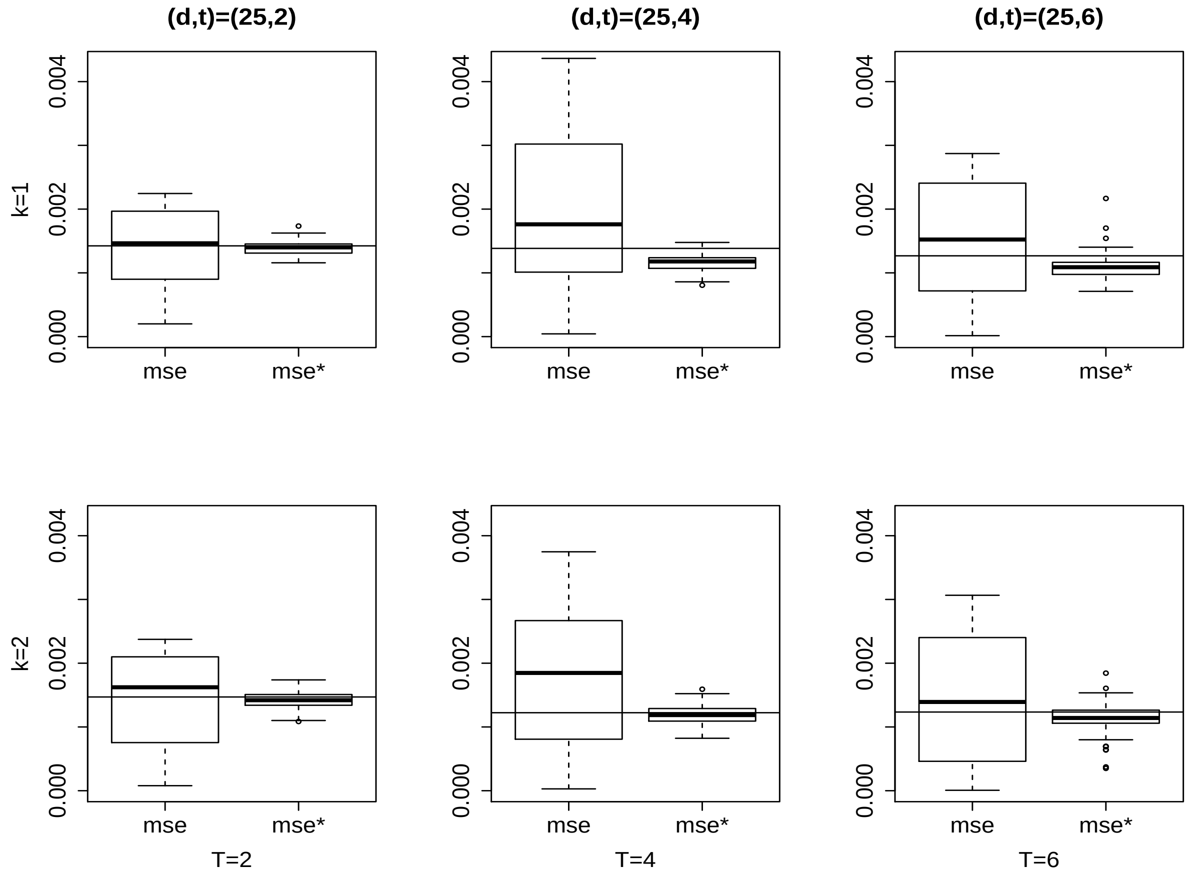


Figure 6.4.1: Boxplots of MSE estimates for $d = 25$ and $t = T$, with $T = 2, 4, 6$ and $D = 50$.

6.5 Application to real data

6.5.1 Data description

The objective of this chapter is to estimate the totals of employed, unemployed and inactive people and unemployment rates in Galician counties. We deal with data from the SLFS of Galicia from the third quarter of 2009 to the fourth quarter of 2011. Our domains of interest are the counties crossed with sex for each time period. As there are 51 counties in the SLFS of Galicia in this period of time, we have $D = 102$ domains, denoted by P_{dt} at time t, and they are partitioned in the subsets P_{d1t}, P_{d2t} and P_{d3t} of employed, unemployed and inactive people. Our target population parameters are the totals of employed and unemployed people and the unemployment rate, this is to say

$$Y_{dkt} = \sum_{j \in P_{dkt}} y_{dktj}, \quad R_{dt} = \frac{Y_{d2t}}{Y_{d1t} + Y_{d2t}}, \qquad k = 1, 2,$$

where $y_{dktj} = 1$ if individual j of domain d at period t is in labour category k and $y_{dktj} = 0$ otherwise. The LFS does not produce official estimates at the domain level, but the analogous direct estimators of the total Y_{dkt}, the mean $\bar{Y}_{dkt} = Y_{dkt}/N_{dt}$, the size N_{dt} and the rate R_{dt} are

$$\hat{Y}_{dkt}^{dir} = \sum_{j\in S_{dt}} w_{dtj}\, y_{dktj},\ \hat{\bar{Y}}_{dkt}^{dir} = \hat{Y}_{dkt}^{dir}/\hat{N}_{dt}^{dir},\ \hat{N}_{dt}^{dir} = \sum_{j\in S_{dt}} w_{dtj},\ \hat{R}_{dt}^{dir} = \frac{\hat{Y}_{d2t}^{dir}}{\hat{Y}_{d1t}^{dir} + \hat{Y}_{d2t}^{dir}},\ k = 1,2, \tag{6.5.1}$$

where S_{dt} is the domain sample at time period t and the w_{dtj}'s are the official calibrated sampling weights.

The target variable is $\boldsymbol{y}_{dt} = (y_{d1t}, y_{d2t})'$, where y_{dkt} is the sample total

$$y_{dkt} = \sum_{j\in S_d} y_{dktj},$$

$y_{dktj} = 1$ if individual j is in category k ($k = 1,2$) and $y_{dktj} = 0$ otherwise.

We employ area-level models using auxiliary information from administrative registers. We use the same auxiliary variables as in the previous chapter. More concretely, we use the domain proportions of individuals within the categories of the following grouping variables.

- SEXAGE: Combinations of sex and age groups, with 6 values. SEX is coded 1 for men and 2 for women and AGE is categorized in 3 groups with codes 1 for 16-24, 2 for 25-54 and 3 for $\geq$55. The codes $1,2,\dots,6$ are used for the pairs of *sex-age* $(1,1),(1,2),\dots,(2,3)$.

- STUD: This variable describes the achieved education level, with values 1-3 for the illiterate and the primary, the secondary and the higher education level respectively.

- REG: This variable indicates if an individual is registered or not as unemployed in the administrative register of employment claimants.

- SS: This variable indicates if an individual is registered or not in the social security system. Figure 6.5.1 shows the scatterplots of the log-rates of employed over inactive people against the proportions of people in social security system (left) and the log-rates of unemployed over inactive people against the proportions of people registered as unemployed (right).We observe that, despite the large variability observed in both plots, the log-rates of the two considered proportions seem to increase linearly with the proportions of people in the social security system and registered as unemployed respectively.

In the fourth quarter of 2011 the domain sample sizes lie all in the interval $(13, 1554)$, with median 97. Therefore, the direct estimates in (A.0.1) are not reliable and small area estimation methods are needed, as in previous chapter.

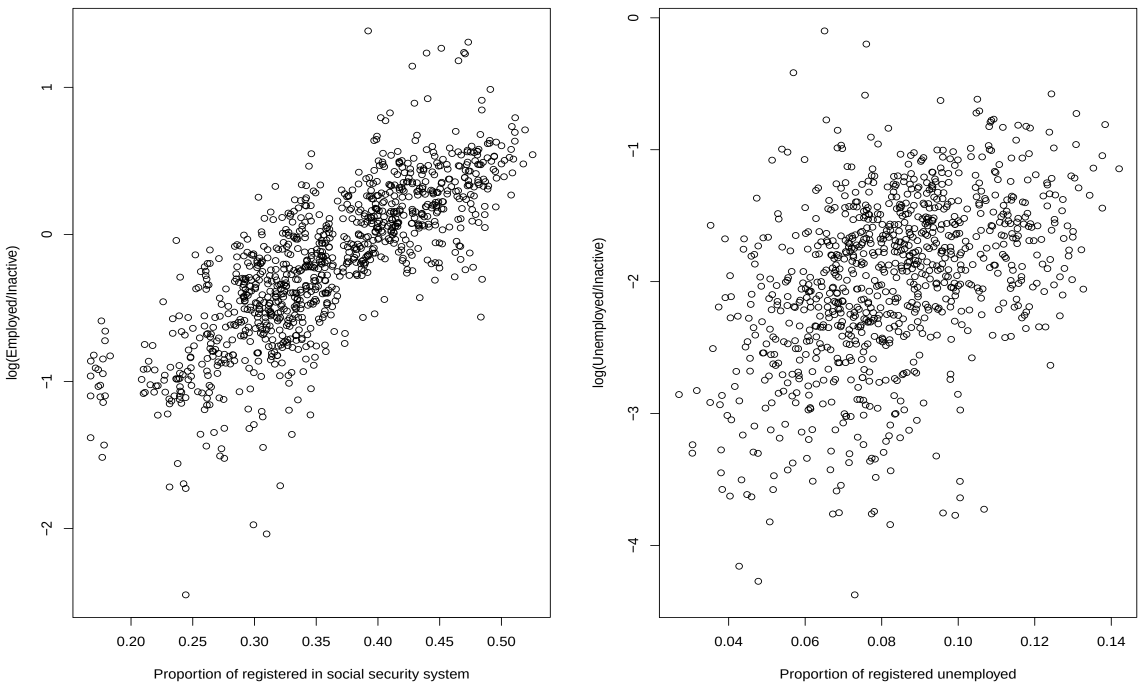


Figure 6.5.1: Log-rates of employed and unemployed over inactive people versus proportions of people in the social security system (left) and registered as unemployed (right) respectively.

6.5.2 Model estimation

We consider the Model 3 defined in (6.1.1)-(6.1.2), the Model 2 defined in (5.1.1)-(5.1.2), the Model 1 defined in (4.1.1)-(4.1.2). We also consider the Model 0 obtained by making $u_{1,d1} = \ldots = u_{1,dq-1}$ in the Model 1. Model 0 is the model studied by Molina et al. (2007). In the application to real data we apply the non temporal Models 1 and 0 to all the considered periods. We consider $q = 3$ categories (employed, unemployed and inactive people) and we choose inactive people as reference (third) category. The multinomial size is $\nu_{dt} = n_{dt}$, where n_{dt} is the size of the domain sample S_{dt} in time t.

Model 3 is firstly fitted to the complete data set. An analysis of residuals is then carried out and nine counties are marked as outliers. These nine counties correspond with the counties of A Coruña, Ferrol, Santiago de Compostela, Ourense, O Condado, O Morrazo, Pontevedra, O Salnés and Vigo. A Coruña, Ferrol, Santiago de Compostela, Ourense, Pontevedra and Vigo are six of the most populous cities of Galicia where the relationships between the auxiliary variables SS and REG with the employment and unemployment status are typically weaker than in less populated counties. The model is finally fitted to reduced data set. The sample sizes of A Coruña, Ferrol, Santiago de Compostela, Ourense, Pontevedra and Vigo are large enought to produce reliable estimates, then no model estimates are give for this counties and direct estimates are used. For O Condado, O Morrazo and O Salnés we use the synthetic estimator. For Model 3 and for each category, Table 6.5.1 presents the estimates of the regression parameters and their standard deviations with the complete data set and Table 6.5.5 with the reduced dataset.

	Employed			Unemployed	
Variable	Estimate	p-value	Variable	Estimate	p-value
CONSTANT	-1.47	0.00	CONSTANT	-4.40	0.00
SEXAGE=1	0.69	0.07	SEXAGE=1	2.26	0.00
SEXAGE=2	2.14	0.00	SEXAGE=2	2.98	0.00
SEXAGE=3	0.16	0.33	SEXAGE=3	-0.45	0.06
SEXAGE=4	0.56	0.18	SEXAGE=4	2.62	0.00
SEXAGE=5	1.71	0.00	SEXAGE=5	2.13	0.00
STUD=1	-0.78	0.00	STUD=1	-0.27	0.29
SS	1.50	0.00	REG	12.40	0.00
	Estimate	Std.Dev.		Estimate	Std.Dev.
φ_1	0.024	0.017		0.081	0.010
φ_2	0.013	0.017		0.098	0.010
ϕ	0.58	0.099		0.29	0.084

Table 6.5.1: Parameter estimates of Model 3 with the full data set.

Figure 6.5.2 plots the domain standardized residuals of model 3 fitted to the full (left) and reduced (right) data set. The dots outside the interval (-3,3) correspond to the men and women counts in marked counties (A Coruña, Ferrol, Santiago, Ourense, O Condado, O Morrazo, Pontevedra and Vigo). The remaining model-based statistical analysis is carried out for the reduced data set.

Figure 6.5.3 plots the domain standardized residuals of employment (left) and unemployment (right) categories versus the proportions of people registered in the social security system (SS) and registered as unemployed (REG). The residuals are randomly distributed above and below zero and no rare pattern is observed. Therefore no diagnostics problems are found for the two main explanatory variables: SS and REG.

For Models 0-3, the estimated regression coefficients and their corresponding p-values for testing the hypothesis $H_0 : \beta_{kr} = 0$ are presented in Tables 6.5.2-6.5.5. The estimates of the model variances and their standard deviations are presented at the bottom of Tables 6.5.1-6.5.4. The 95% confidence intervals for ϕ_1 and ϕ_2 are 0.58 ± 0.19 and 0.29 ± 0.16 respectively. These intervals do not contain zero. Then we may recommend Model 3, i.e. the model with time-correlated random effects.

6.5.3 Model diagnostics

For carrying out the diagnosis of the models, we calculate the predicted sample totals $\hat{y}_{dkt} = n_{dt}\hat{p}_{dkt}$ and the domain residuals

$$r_{dkt} = \frac{y_{dkt} - \hat{y}_{dkt}}{\hat{y}_{dkt}}, \quad d = 1, \dots, 102, \quad k = 1, 2, \quad t = 1, \dots, 4.$$

Figure 6.5.4 plots the domain residuals versus the predicted sample totals of employed and unemployed people for the four models. We observe that the model residuals are

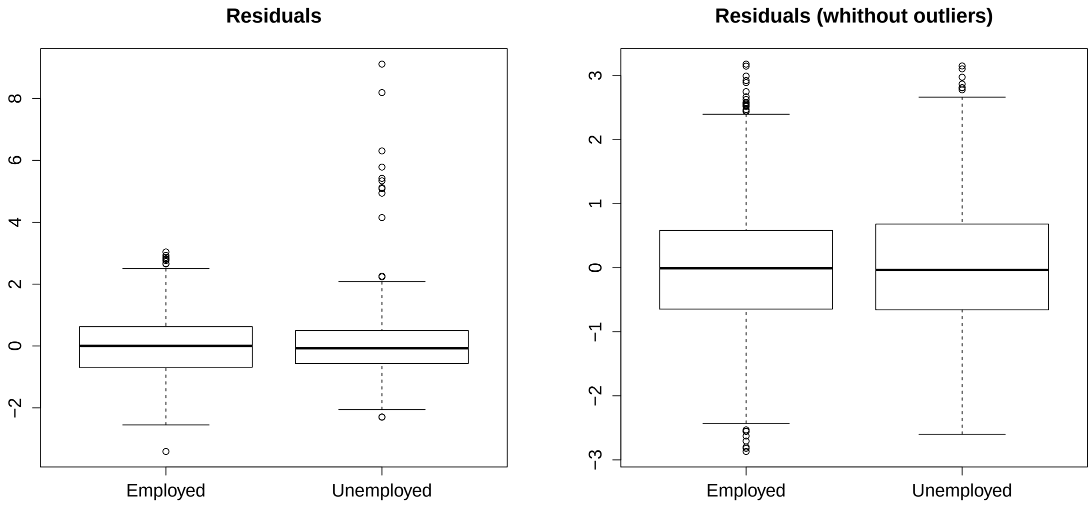


Figure 6.5.2: Boxplots of standardized residuals of models fitted to the full (left) and reduced (right) data set.

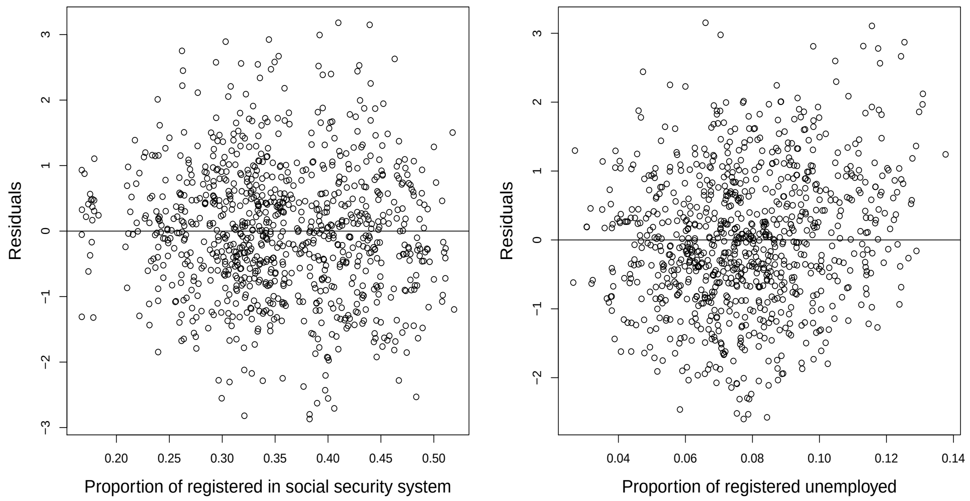


Figure 6.5.3: Domain standardized residuals of employment (left) and unemployment (right) categories versus proportions of people registered in the social security system (SS) and registered as unemployed (REG).

symmetrically situated above and below zero, so there is no prediction bias. Further, the variability of the residuals decreases as predicted employed or unemployed sample totals increase. This pattern is due to the fact that domains with greater amount of employed and unemployed people also have greater sample sizes. We also observe that there are no high residuals in absolute value or any other unusual pattern. Therefore, the fitted model seems to properly describe the data.

Figure 6.5.5 plots the direct estimates ($\hat{Y}_{dkt}^{dir}$) versus the model-based estimates ($\hat{m}_{dkt} =$

	Employed			Unemployed	
Variable	Estimate	p-value	Variable	Estimate	p-value
CONSTANT	-1.44	0.00	CONSTANT	-5.25	0.00
SEXAGE=1	-0.66	0.49	SEXAGE=1	1.62	0.36
SEXAGE=2	1.85	0.00	SEXAGE=2	3.86	0.00
SEXAGE=3	0.27	0.39	SEXAGE=3	0.79	0.19
SEXAGE=4	-0.46	0.75	SEXAGE=4	4.02	0.15
SEXAGE=5	1.64	0.00	SEXAGE=5	3.75	0.00
STUD=1	-0.80	0.04	STUD=1	0.79	0.24
SS	1.96	0.00	REG	9.33	0.00
	Estimate	Std.Dev.			
φ	0.06	0.0055			

Table 6.5.2: Parameter estimates of Model 0 for the fourth quarter of 2011.

	Employed			Unemployed	
Variable	Estimate	p-value	Variable	Estimate	p-value
CONSTANT	-1.44	0.00	CONSTANT	-5.39	0.00
SEXAGE=1	-0.66	0.49	SEXAGE=1	1.73	0.36
SEXAGE=2	1.84	0.00	SEXAGE=2	4.00	0.00
SEXAGE=3	0.26	0.41	SEXAGE=3	0.81	0.22
SEXAGE=4	-0.41	0.78	SEXAGE=4	3.71	0.22
SEXAGE=5	1.62	0.01	SEXAGE=5	3.99	0.00
STUD=1	-0.80	0.04	STUD=1	0.91	0.24
SS	1.98	0.00	REG	9.27	0.00
	Estimate	Std.Dev.		Estimate	Std.Dev.
φ_1	0.010	0.007		0.096	0.037

Table 6.5.3: Parameter estimates of Model 1 for the fourth quarter of 2011.

$N_{dt}\hat{p}_{dkt}$) of the population totals of employed (top line) and unemployed people (bottom line) in logarithmic scale in the three models. We observe, in all the models, that the direct and the model-based estimates behave quite similarly for employed people. This is because the population of employed people is quite large and there are plenty of sampled observations within this category. However, the direct and model-based estimates behave slightly different for unemployed people, which is due to the lower number of sampled observations within the category. We also observe that the model-based estimates are lower than the direct ones for large values of the direct estimates. This is a typical and desirable smoothing effect of model-based estimators.

	Employed			Unemployed	
Variable	Estimate	p-value	Variable	Estimate	p-value
CONSTANT	-1.41	0.00	CONSTANT	-4.01	0.00
SEXAGE=1	0.83	0.04	SEXAGE=1	1.92	0.02
SEXAGE=2	2.01	0.00	SEXAGE=2	2.46	0.00
SEXAGE=3	0.16	0.38	SEXAGE=3	-0.43	0.20
SEXAGE=4	0.60	0.21	SEXAGE=4	1.84	0.06
SEXAGE=5	1.66	0.00	SEXAGE=5	1.69	0.00
STUD=1	-0.92	0.00	STUD=1	-0.43	0.23
SS	1.63	0.00	REG	12.19	0.00
	Estimate	Std.Dev.		Estimate	Std.Dev.
φ_1	0.034	0.006		0.096	0.019
φ_2	0.015	0.003		0.119	0.014

Table 6.5.4: Parameter estimates of Model 2.

	Employed			Unemployed	
Variable	Estimate	p-value	Variable	Estimate	p-value
CONSTANT	-1.42	0.000	CONSTANT	-3.83	0.000
SEXAGE=1	0.89	0.027	SEXAGE=1	1.98	0.005
SEXAGE=2	1.99	0.000	SEXAGE=2	2.26	0.000
SEXAGE=3	0.15	0.398	SEXAGE=3	-0.43	0.121
SEXAGE=4	0.58	0.209	SEXAGE=4	1.39	0.117
SEXAGE=5	1.66	0.000	SEXAGE=5	1.54	0.002
STUD=1	-0.92	0.000	STUD=1	-0.47	0.114
SS	1.67	0.000	REG	11.66	0.000
	Estimate	Std.Dev.		Estimate	Std.Dev.
φ_1	0.026	0.020		0.090	0.014
φ_2	0.014	0.020		0.113	0.014
ϕ	0.58	0.117		0.29	0.0.095

Table 6.5.5: Parameter estimates of Model 3.

6.5.4 Small area estimates and RMSE

Figures 6.5.6 and 6.5.7 plot the estimated employment totals and unemployment rates respectively per sex, for the fourth quarter of 2011 and for all the models, with the counties sorted by sample size. We observe that for employed all the estimates are very similar, this is because the population in this category is quite large. In the unemployment rates the direct and the model-based estimators tend to be closer as soon as the sample size increases. The same pattern is observe in the rest of quarters. For the sake of brevity we skip the corresponding figures.

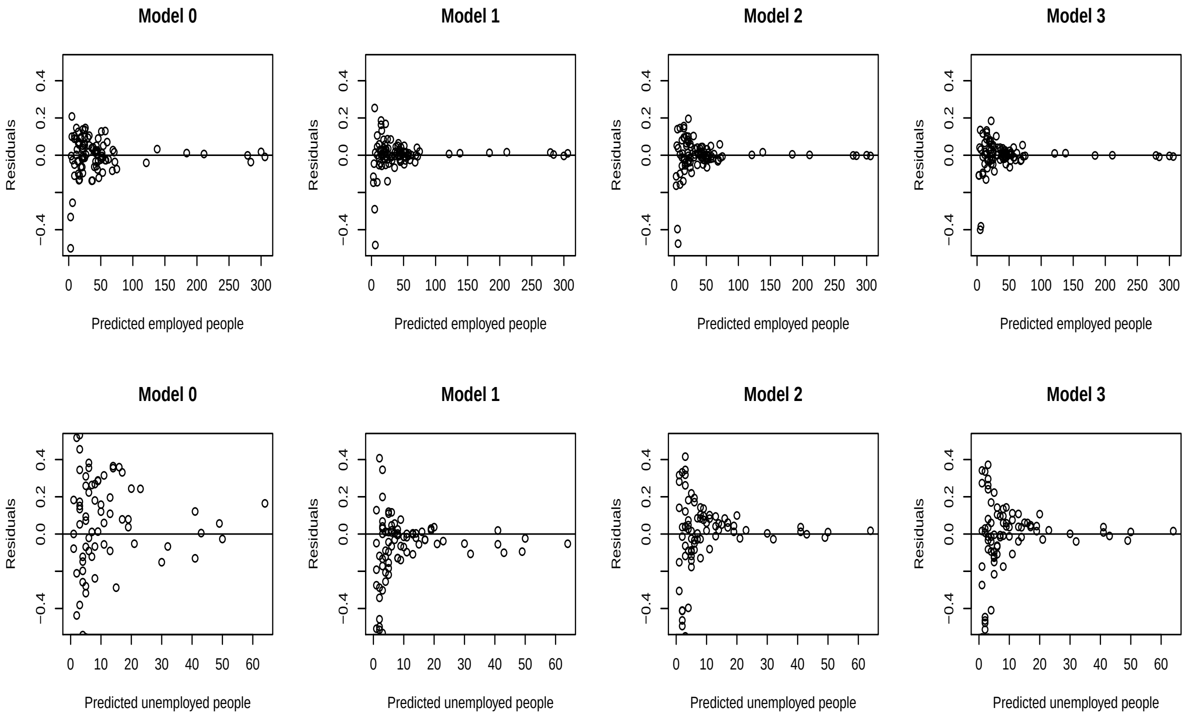


Figure 6.5.4: Domain residuals versus predicted sample totals for the three models

Figures 6.5.8 and 6.5.9 plot the parametric bootstrap estimates of the relative mean squared errors (RMSE) of the model-based estimators of totals of employed and of unemployment rates for all the models. The RMSEs of the corresponding direct estimates are much higher than their model-based counterparts. This is the reason why they have not been plotted in Figures 6.5.8 and 6.5.9. We can see that the RMSE for Model 3 are, in general terms, the best.

Tables 6.5.6 and 6.5.7 gives some condensed numerical results for the fourth quarter of 2011, for all the models and for unemployment rates. For each sex we firstly sort the domains by province and after that, in each province, we sort the domains by sample size, starting by the domain with smallest sample size, and then we chose five domains in each province. We show the results of the direct and model-based estimates (labeled by "dir" and "mod" respectively) and the corresponding RMSE estimates. The sample sizes are labeled by n. This tables present blank spaces in domains where $y_{dk10} = 0, k = 1, 2$. By observing the columns of RMSEs we conclude that model-based estimators that include time effects are preferred to the direct ones.

Table 6.5.8 gives the direct and model-based estimator for Model 3 and for employed and unemployed people at the province level for men (top) and women (bottom) for the fourth quarter of 2011 in the SLFS. In this table we can see also the consistency factors λ_p, defined a in Chapter 5. We observe that the deviation from the SLFS estimation at province level are at most of 1%.

Another way of measuring the benefits of using Model 2-3 is to check the stability of the estimates along the time period. Figure 6.5.10 presents the unemployment rates for

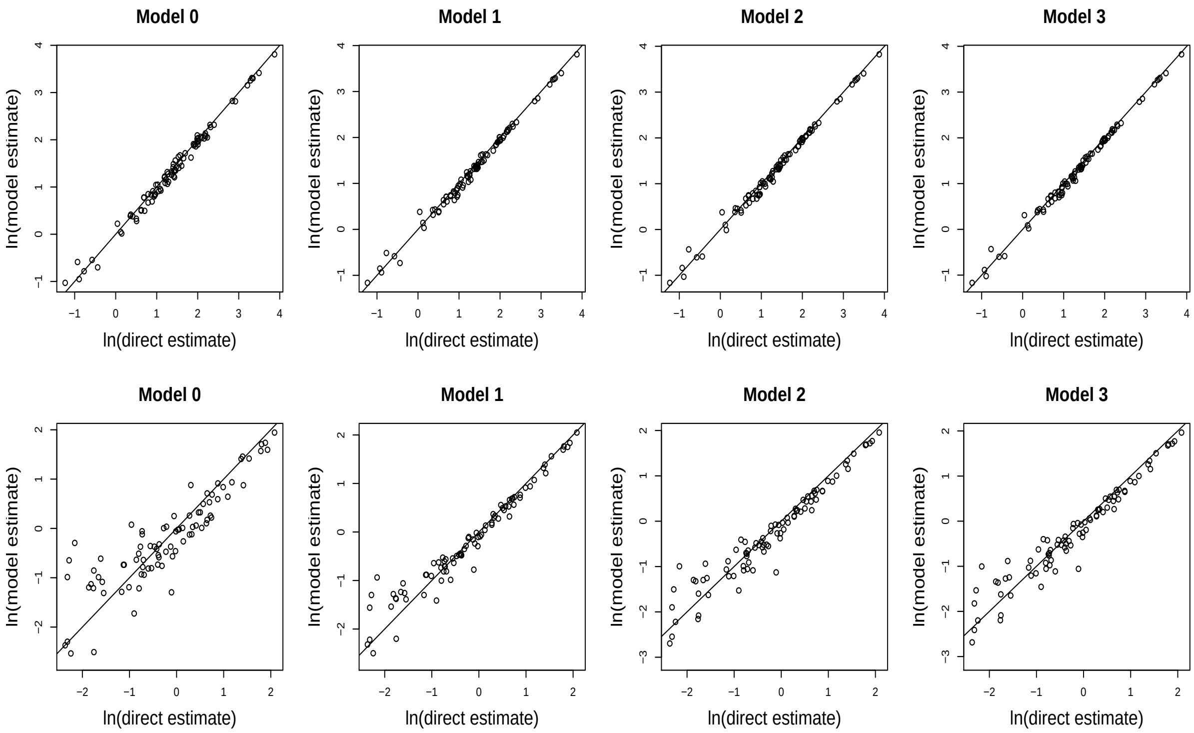


Figure 6.5.5: Direct versus model-based estimates. Total of employed in top line and total of unemployed in bottom line.

women for six counties that divide the sample size distribution below its median into six equal parts. In all the considered counties, we observe that the results of Model 2-3 are much more stable than the results of the direct and Model 0-1. Stability is a property highly valued by the Statistical Offices when publishing the survey results. For Statistical Offices it is hard to justify that there are more than three points (in %) of difference between the unemployment rates of two consecutive quarters in a given county.

Now, the best two models (model 2 and model 3) can be compared with the usual measures for model comparison such as the loglikelihood or the BIC. The resulting values of these measures are listed in Table 6.5.9. We can see that for model 3 including the ϕ parameter has the better Loglikelihood and BIC. Therefore we recommend this model.

Figures 6.5.11 and 6.5.12 map the estimates of unemployment rates per sex in each county of Galicia for the fourth quarter of 2011 and the variation of the unemployment rates between the fourh quarter of 2009 and the fourth quarter of 2011. The colors are more intense in areas with higher unemployment rates an higher variation. We observe that the counties of the west coast are those that, in general terms, have higher unemployment rates. In this figures we can also see that in that area was where most unemployment rates increased between 2009 and 2011, this increase was much higher in men than in women.

Figures 6.5.13 shows the model-based estimators of unemployment rates for two quarters, the fourth quarter of 2009 and the fourth quarter of 2011. Like in the maps, in these figures we can observe that the unemployment rates for women are higher than for men.

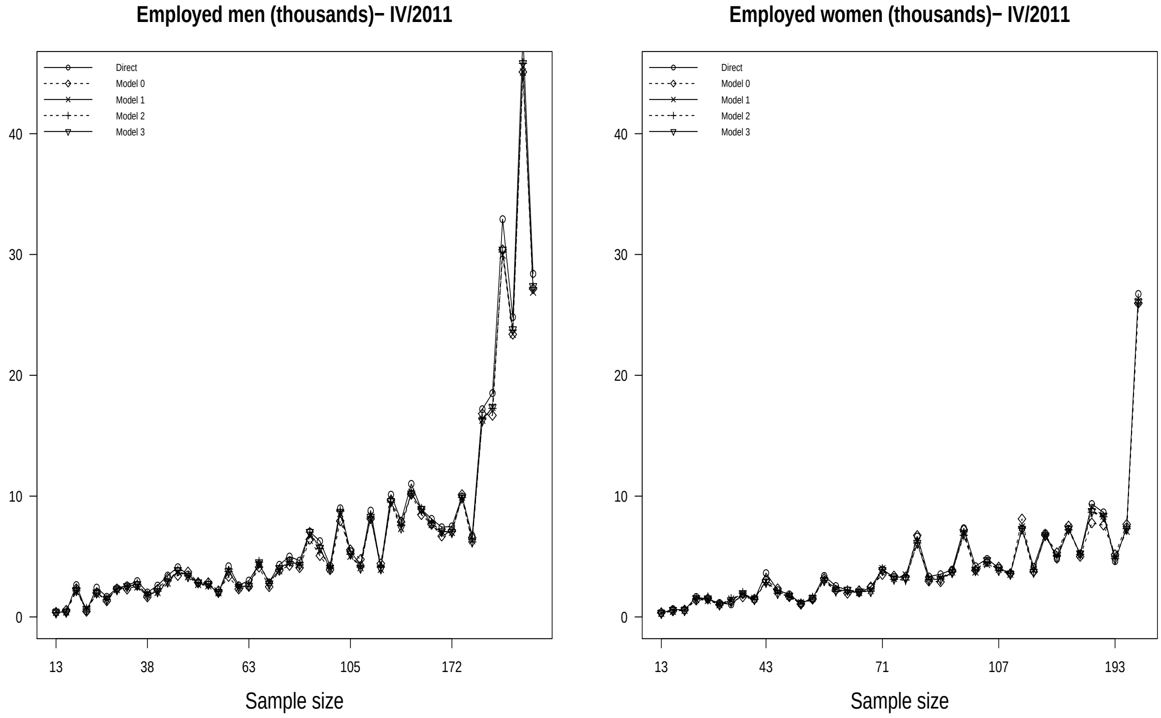


Figure 6.5.6: Direct and model-based estimates of totals of employed men and women in the fourth quarter of 2011.

We can also confirm that in this period there has been a general increase in unemployment rates, but this increase was much higher in men.

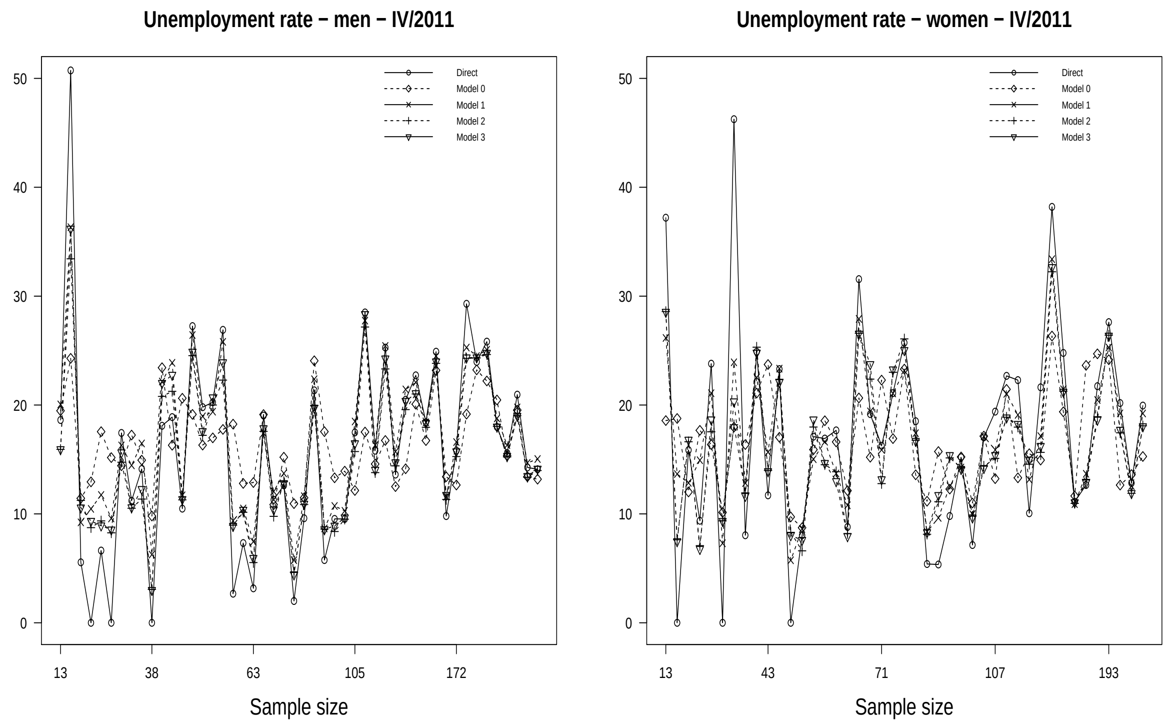


Figure 6.5.7: Direct and model-based estimates of unemployment rates for men and women in the fourth quarter of 2011.

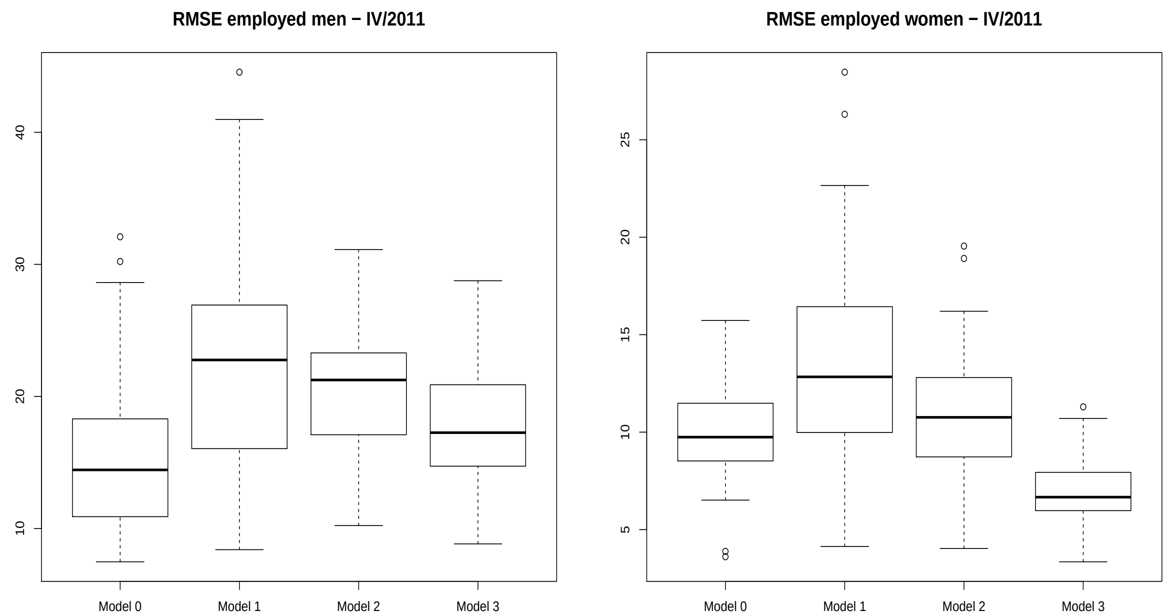


Figure 6.5.8: RMSEs of direct and model-based estimator of employment totals per sex in the fourth quarter of 2011.

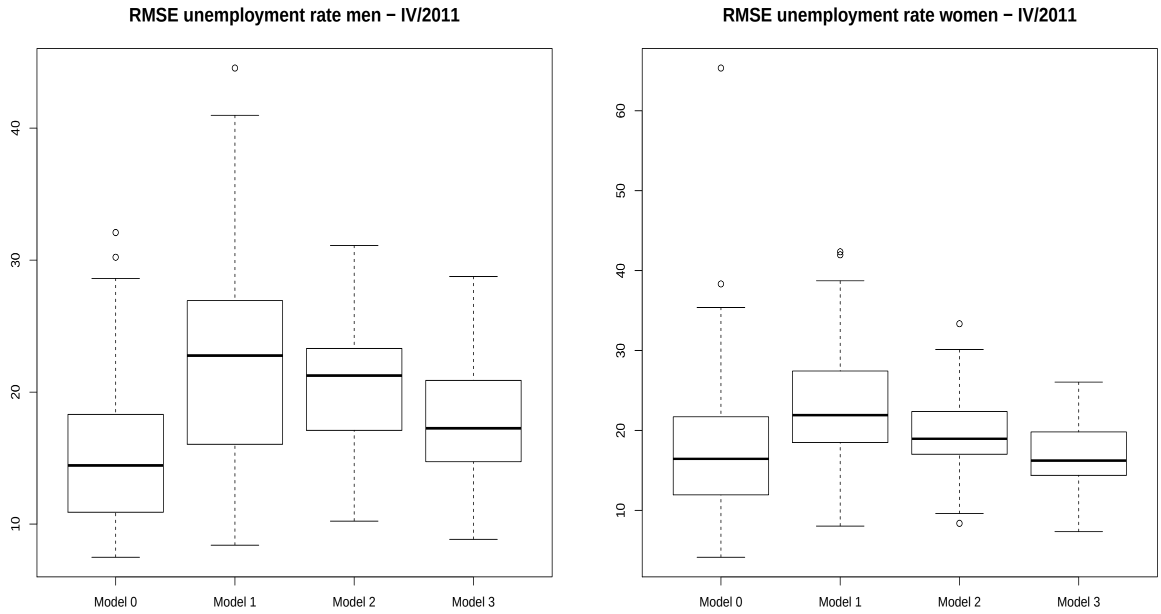


Figure 6.5.9: RMSEs of model-based estimator of unemployment rates in the fourth quarter of 2011.

p	n	dir	mod_0	mod_1	mod_2	mod_3	dir	mod_0	mod_1	mod_2	mod_3
1	21	5.55	11.43	9.22	11.24	10.57	201.87	10.58	39.14	30.49	23.98
1	37	11.21	17.24	14.46	10.52	10.59	129.74	16.49	28.24	24.95	20.30
1	46	10.48	20.62	11.71	11.29	11.34	106.40	16.40	22.79	21.33	20.99
1	65	19.03	19.13	17.42	17.58	17.85	51.53	28.62	17.79	21.29	20.49
1	384	15.59	15.34	16.32	15.23	15.30	24.94	9.28	10.07	11.23	9.43
2	13	18.62	19.47	20.05	15.93	15.93	146.41	23.54	39.66	31.12	24.90
2	62	7.32	12.80	10.50	10.15	10.36	174.27	14.88	26.69	25.97	18.20
2	79	12.63	15.21	13.72	12.97	12.81	70.06	9.70	24.29	21.28	16.66
2	95	9.54	13.32	10.73	8.38	8.90	96.01	14.84	22.73	19.90	16.72
2	592	14.10	13.19	15.04	13.98	14.13	21.18	8.29	8.68	10.47	8.84
3	15	50.73	24.29	36.33	33.43	36.18	31.04	27.60	40.97	25.25	25.84
3	58	19.80	16.32	18.96	17.22	17.58	56.39	13.96	25.18	22.45	17.42
3	58	26.90	17.76	25.80	22.29	23.90	44.93	16.78	25.21	21.99	18.81
3	107	28.51	17.52	27.75	27.17	28.33	26.48	15.13	21.10	18.46	16.88
3	184	23.97	23.23	24.50	24.33	24.36	25.11	12.08	14.47	12.49	10.80
4	40	18.09	23.43	22.16	20.79	22.01	71.93	24.59	27.17	22.03	19.60
4	94	5.76	17.56	8.67	8.61	8.57	137.13	11.81	19.31	17.18	16.13
4	138	24.90	23.18	24.51	23.83	23.99	24.44	13.80	13.20	15.33	12.42
4	262	25.82	22.21	25.08	24.64	24.78	15.68	10.59	9.81	11.99	9.83
4	413	20.95	19.53	19.78	19.02	18.98	16.08	7.67	8.40	10.26	9.13

Table 6.5.6: Men unemployment rates (left) and their estimated RMSEs (right).

p	n	dir	mod_0	mod_1	mod_2	mod_3	dir	mod_0	mod_1	mod_2	mod_3
1	20	9.37	17.67	14.93	7.09	6.76	208.55	4.14	30.28	25.11	23.35
1	32	8.04	16.37	12.83	11.93	11.63	241.60	8.07	26.85	24.88	19.75
1	56	16.92	18.56	16.63	14.52	14.66	71.25	20.95	23.59	19.79	16.78
1	79	25.72	23.30	23.14	26.07	25.03	41.00	16.42	20.40	17.21	15.51
1	114	22.29	13.31	19.09	17.97	18.24	33.81	9.29	17.16	16.91	15.71
2	13	37.20	18.57	26.16	28.58	28.54	58.99	16.65	38.57	30.12	24.41
2	54	8.03	8.73	8.61	6.61	7.59	267.95	65.35	38.72	24.01	26.07
2	97	5.40	11.16	8.35	8.13	8.29	234.42	22.22	25.01	19.36	18.14
2	107	19.39	13.24	15.87	15.10	15.39	42.52	12.31	19.35	17.61	14.14
2	643	12.84	13.61	12.56	11.90	11.89	25.65	6.17	8.04	9.61	8.35
3	14		18.76	13.66	7.63	7.48		17.35	34.99	33.35	22.85
3	54	17.09	15.90	15.03	17.98	18.66	101.72	38.34	28.77	22.15	19.08
3	69	19.17	15.21	19.42	22.38	23.74	67.89	34.77	27.27	18.96	19.90
3	108	22.68	21.44	21.03	18.88	18.84	34.85	7.03	15.80	16.68	14.36
3	193	27.61	24.16	25.29	26.41	26.35	25.20	18.40	13.11	12.68	10.58
4	40	24.90	21.10	22.53	25.31	24.79	68.63	29.75	25.92	21.23	18.76
4	103	17.03	17.14	17.13	14.39	14.20	50.11	13.16	19.21	17.62	14.51
4	139	21.63	14.94	17.15	15.65	16.22	34.94	12.52	18.73	17.17	14.92
4	144	38.20	26.34	33.40	32.24	32.65	16.72	10.39	13.84	12.14	10.73
4	189	21.73	24.70	20.46	18.93	18.63	32.13	8.49	11.58	13.22	9.85

Table 6.5.7: Women unemployment rates (left) and their estimated RMSEs (right).

			Employed people			Unemployed people		
sex	Province	n	dir	mod3	λ_p	dir	mod3	λ_p
Men	1	2874	248516	238993	1.04	45969	43689	1.05
	2	1541	72917	69103	1.06	9307	9439	0.99
	3	1336	59824	60034	1.00	13883	13481	1.03
	4	2988	192842	187530	1.03	53477	50195	1.07
Women	1	3258	225489	222590	1.01	44696	42160	1.06
	2	1659	63163	60697	1.04	9304	8548	1.09
	3	1479	51726	50753	1.02	12418	11825	1.05
	4	3427	158415	160645	0.99	51303	49675	1.03

Table 6.5.8: Estimated men and women province totals for IV/2011.

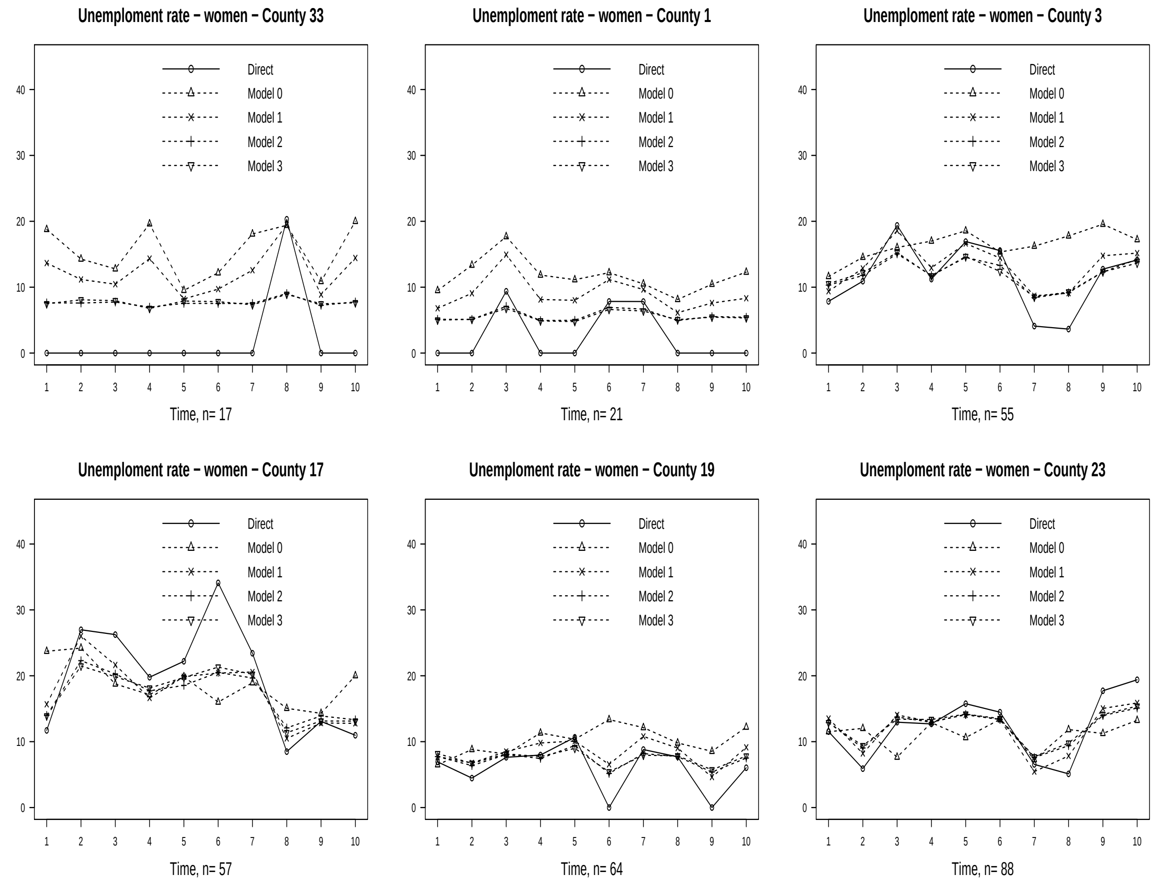
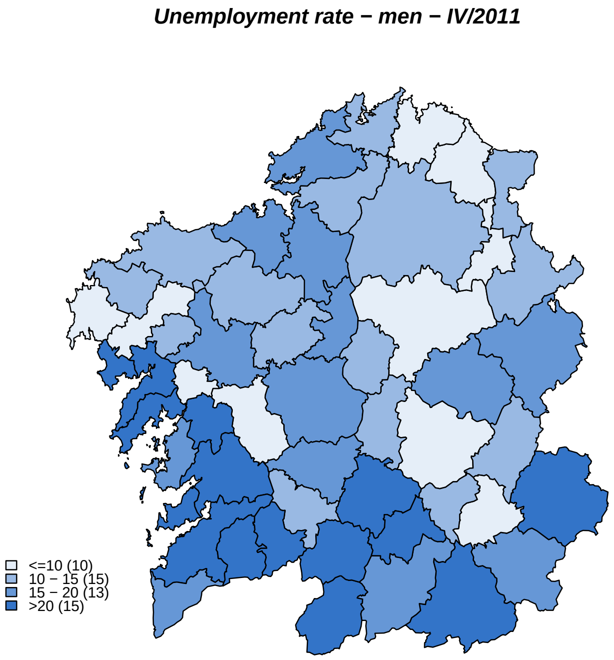


Figure 6.5.10: Unemployment rates in some counties and all the periods.

	Model 2	Model 3
Loglikelihood	-5269.5	-5037.6
BIC	10593.1	10135.4

Table 6.5.9: Model comparison.



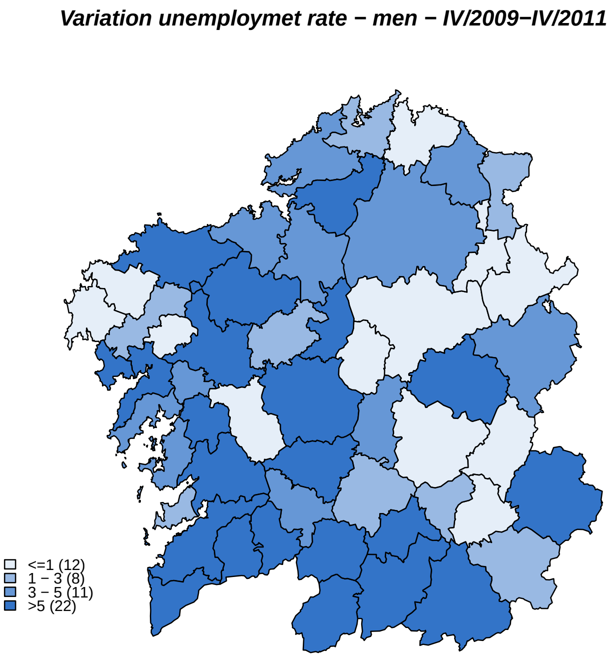
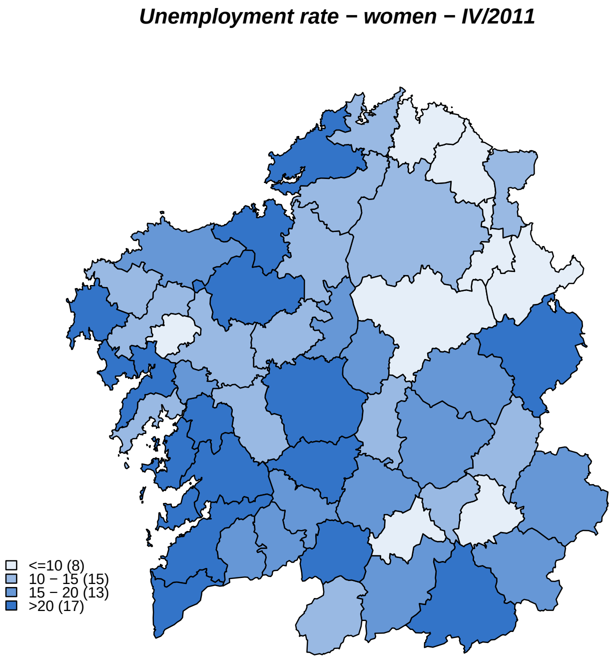


Figure 6.5.11: Estimates of men unemployment rates in Galician counties in the last quarter of 2011 (top) and estimates of the variation between men unemployment rates in the last quarter of 2011 in front of the last quarter of 2009 (bottom).



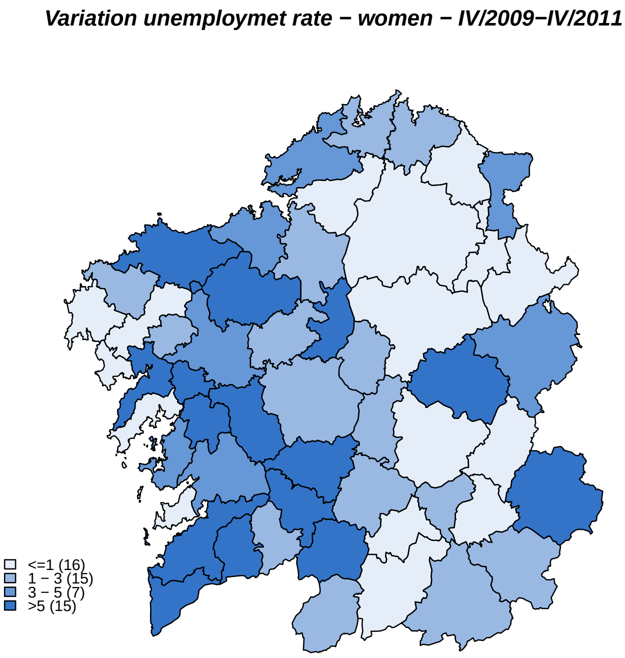


Figure 6.5.12: Estimates of women unemployment rates in Galician counties in the last quarter of 2011 (top) and estimates of the variation between women unemployment rates in the last quarter of 2011 in front of the last quarter of 2009 (bottom).

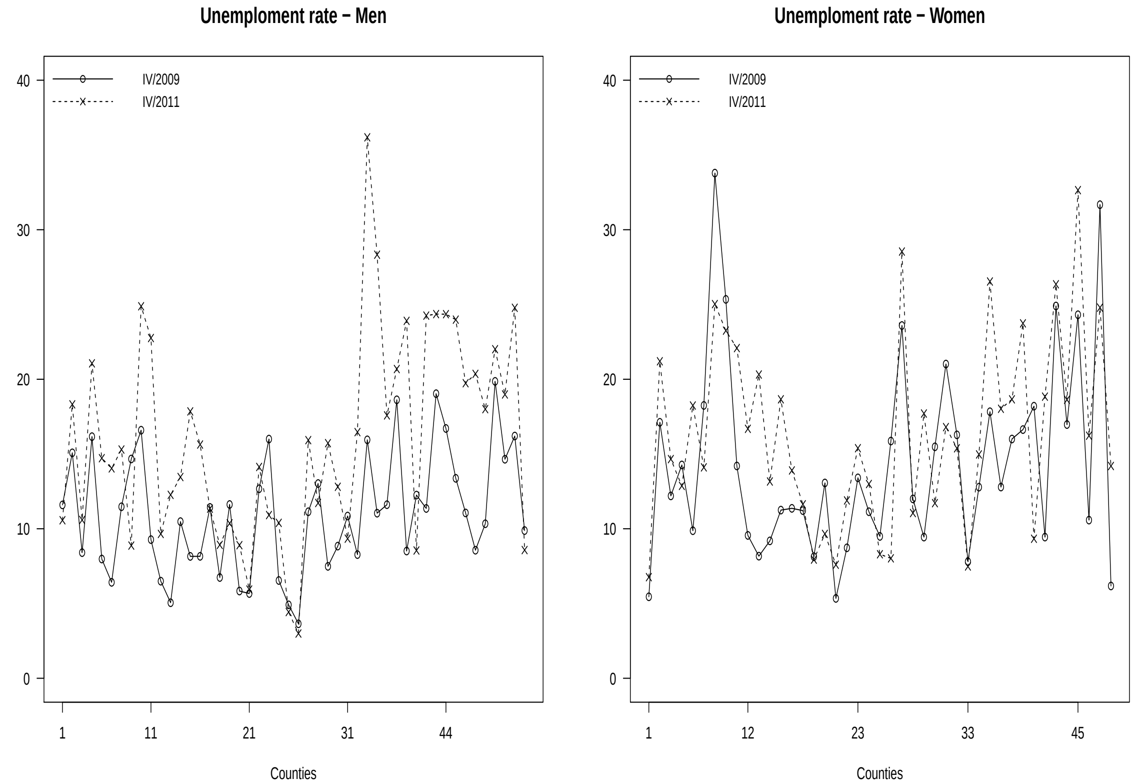


Figure 6.5.13: Model-based estimator of unemployment rates for fourth quarter of 2009 and fourth quarter of 2011.

Chapter 7

Conclusions and future lines

In this thesis we study the problem of estimating totals of employed and unemployed people and unemployment rates in the counties of Galicia. The problem is how to produce reliable estimates of this characteristics for areas (counties) where only small samples or no samples are avaliable, and how to assess their precision. We propose three area level multinomial mixed models with time and area effects and with independent random effects on the two categories of the target vector. Unlike the work by Molina et al. (2007), we employ multinomial models with two random effects, one associated with the category of employed people and the other associated with the category of unemployed people. This is due to the different modeling requirements for each labour category in the Galician data. We use area-level models because they have wider scope than the unit level models because area-level auxiliary information is more readily available than unit-level auxiliary data. The use of auxiliary information for SAE is vital because with the small sample sizes often encountered in practice, even the most elaborated model can be of little help if it does not involve a set of covariates with good predictive power for the small area quantities of interest (Pfeffermann (2012)).

The obtained model-based estimates for all the models are compared with the direct ones. They have lower mean squared errors, especially for counties with small sample size. Another advantage of the proposed model-based estimators is their property of being consistent in the sense that estimates of domain totals of employed, unemployed and inactive people sum up to the size of the domain. This is a desirable benchmarking property.

In two of the models we propose the use of time effects. The inclusion of time effects allows to obtain estimates of employed, unemployed totals and of unemployment rates in a more accurate and stable form than if separate models were fitted for each time period. Because of these properties, the proposed methodology is very suitable for being used in official statistical offices. Further, as estimates follow the pattern of direct estimators for large counties and behave stably for small counties, the smoothing effect of using past time periods seems reasonable.

We have presented an application where the temporal models are fitted to the 10 most recent quarters. Another important issue is how the proposed models may be applied for repeated surveys and more concretely in the SLFS. The use of 10 quarters is the consequence of using all our available data and therefore is arbitrary for this purpose. The simulations suggest using at least the 5 and the 9 most recent quarters when using Model 2 and Model 3, respectively.

By using all the data up to the last quarter, the introduced temporal models give estimates for any of the considered quarters. Nevertheless, we suggest using these models for obtaining only estimates in the last quarter. Because of the practical difficulties that Statistical Offices have with revising published data and with running time consuming computational procedures, we propose using a dynamic fixed length time "window" by adding each time the current quarter and discarding the earliest one. If this methodology were ever put in production, we recommend windows of 5 and 9 quarters for Models 2 and 3, respectively.

We would like to emphasize that the introduced approach to estimating labour force indicators, and also the ones by Molina et al. (2007) and López-Vizcaíno et al. (2013), are not adapted to the particularities of a complex sampling design. They are derived for simple random sampling and do not take into account for potential LFS sampling-design effects. The sampling weights are only used through $\hat{N}_{dt}^{dir}$ when calculating the model-based estimates of domain totals of employed and unemployed people. They are also used in the calibration to the province totals. Many area-level small area estimation methods introduce the complex sampling design information through the direct estimator of the domain total or mean ($\hat{Y}_{dkt}^{dir}$ or $\hat{\bar{Y}}_{dkt}^{dir}$) and their moments. As a multinomially distributed vector is the sum of i.i.d. multi-Bernoulli vectors, the survey domain totals of employed and unemployed people can be modeled by means of multinomial distributions. The modelization of $(\hat{Y}_{d1t}^{dir}, \dots, \hat{Y}_{dq-1t}^{dir})$ as multinomial seems to be unrealistic.

An issue that should be taken into account is that true domain size N_{dt} is a known quantity and that it is assumed to be equal to $\hat{N}_{dt}^{dir}$. In practice, this is not true. Therefore, when estimating the MSE of the estimators of domain totals of employed and unemployed people we are ignoring the uncertainty in $\hat{N}_{dt}^{dir}$ and its correlation with $\hat{p}_{dkt}$. To proceed in a rigorous way, the estimation of the extra variability and of the ignored correlation is needed. This might require implementing some bootstrap or Jackknife method at the unit-level, something that might have a high computational cost. Nevertheless, the direct domain size estimates, $\hat{N}_{dt}^{dir}$, are constructed from the official calibrated sampling weights, that are obtained after correcting the non response and after calibrating to the "known" populations sizes. They can be considered as the official best estimates of our domain sizes. In this sense, we may admit the simplification of assuming that the true domain sizes are equal to the direct domain size estimates.

As calibrated weights enter in the estimation process through the direct estimators $\hat{N}_{dt}^{dir}$, it is worthwhile to investigate how large is the correlation between the $\hat{N}_{dt}^{dir}$ and the $\hat{p}_{dkt}$ values. The estimated correlations for the target period IV/2011 are 0.27 and 0.29 for the categories of employed ($k = 1$) and unemployed ($k = 2$) people respectively, which are rather low. From the socioeconomic perspective, this positive correlation shows something that it is well known. The Galician rural counties tend to have lower population size and higher proportion of inactive people.

It is interesting to analyze if the estimation procedure uses direct estimators $\hat{N}_{dt}^{dir}$ based on weights that are calibrated for the model covariates. In the presented application to real data, the multinomial models use the auxiliary variable SEXAGE at the county level. On the one hand, SEXAGE contains the combinations of sex and age groups with 3 age groups. On the other hand, the weights are calibrated to 11 age groups at the whole population (Galicia), which is in strict sense a different auxiliary variable. Therefore, we have not applied the proposed methodology by using twice the same auxiliary variables, first in the calibration of weights and second in the multinomial models.

This work produces estimates for sex by county domains, but it does not deal with the fact that the multinomial outcomes for males and females from the same county might be correlated. The proposed Model 2 and Model 3 have not separate fixed effects for males and females and so essentially handle this problem via the correlation between males and female values of the model covariates. Further generalizations could be done by adding a new sex index $s = 1, 2$ to the set of data indexes (d, k, t) and by considering possible correlation structures between sexes.

The estimator defined in (6.3.1) is a plug-in estimator of the expected value of the small area proportion given the small area distribution of the covariates and the random effects in the model. It can be calculated with a low computational cost. It is a parsimonious estimator. The optimal estimator is the so-called Empirical Best Predictor (EBP) of the population proportion, which is a plug-in estimator of the conditional expectation of the small area proportion given the small area distribution of the covariates and the sample data. The EBP can be obtained by approximating two $2(T + 1)$-variate integrals by Monte Carlo Integration and its corresponding mean squared error can be calculated by parametric bootstrap. The high computational cost is thus the main drawback for using the EBP under Model 3. This is the main reason why we prefer employing the plug-in estimator in this work.

Indeed model-based predictors permit predictions of non sampled area for which no design-based theory exists. This is not to say that design-based estimators have no role in model-based prediction. The design-based estimators are the input data for the models and we use the design-based estimators to assessing the model-based predictors and for calibrating them via benchmarking. The benchmarking has the advantage of guaranteeing consistency of publication between the model-based small area predictors and the design-based estimator for the aggregated area. This is often required by statistical offices. In this case we use the ratio adjustment, but an extension of this work may be use quarterly benchmark constraints in the model, like in the paper of Pfeffermann and Tiller (2006).

As stated in the introduction, an important aspect of SAE is the assessment of the accuracy of the predictors. Three procedures for estimating the mean square error (MSE) have been considered. The first one has a explicit form of the type $\mathcal{G}_1 + \mathcal{G}_2 + 2\mathcal{G}_3$ based on Prasad and Rao (1990). It is based on a Taylor linearisation. The remaining procedures rely on bootstrap methods and they also allows the MSE estimation of non linear predictors. In the simulations we have observed that the bootstrap methods present a better behavior.

We have carried out an application to real data from the SLFS of Galicia and we have noticed that the introduced model-based procedure gives estimates with lower RMSE than the direct ones, in similar sense to other small area estimation studies in this context (Esteban et al., 2012; Herrador et al., 2009; Molina et al., 2007; Ugarte et al., 2009a). In this case, the best model in terms of RMSE is the model that has correlated time and area effects. This is the best model in terms of AIC. Nevertheless we would like to point out that the Akaike information criterion (AIC) is a popular, but not well established, model selection criterion for generalized linear mixed-effects models. The conditional AIC (cAIC) was derived for linear mixed-effects models by Vaida and Blanchard (2005). It is an appropriate information criteria for linear mixed model selection. However, the application of cAIC to generalized linear mixed models is not straightforward and requires further research to obtain specific derivations. This is pointed out by Lian (2012) when deriving the cAIC for Poisson regression models with random effects. The derivation of the cAIC for multinomial mixed model is also not straightforward. As far as we know, this is still a research problem that has not been treated in the literature and that deserves an specific future research. Therefore the obtained estimates of labour indicators are useful for policy making at county level.

An extension of these models can incorporate spatial correlations between domains. In practice it is often reasonable to assume that the effects associated with neighboring areas (defined, for example, by a contiguity criterion) are proportionally correlated to a measure of distance (not necessarily geographical), with correlations decreasing to zero as the distance increases. That is, small area models should allow for spatial correlation of area random effects. Such models are common in spatial statistics (Cressie, 1993) and in SAE when linear mixed models are used (Chandra and Chambers, 2009).

As for the labour market results in Galicia we can conclude that there has been a general increase in unemployment rates in the considered period and in almost all the counties, although this increase was greater for men. This can be conditioned, inter alia, by the sharp fall in employment in the construction sector, which employs mainly men. Due to the economic situation in Spain, with a fall in gross domestic product in 2012 of 1.9%, this population has not been able to find work in another field of activity. We also conclude that the counties of the west coast of Galicia are those that, in general terms, have higher unemployment rates. In this part was, also, where most unemployment rates increased between 2009 and 2011. That area is the most dynamic part of Galicia and the least aged. In that area live the 75% of the Galician population and the unemployment rates are high because companies can not absorb as many workers.

Chapter 8

Software: The mme package

The mme package is the result of the implementation of the models discussed throughout this dissertation. Over the last years, the R computing environment has become a powerful scientific tool that offers a rich collection of classical and modern statistical modeling techniques. Motivated by its flexibility and its widely acceptance among the scientific community, we have chosen R as programming language to develop this library of functions.

Package 'mme'

February 5, 2014

Type Package

Title Multinomial Mixed Effects Models

Version 0.1-4

Date 2013-06-10

Author E. Lopez-Vizcaino, M.J. Lombardia and D. Morales

Maintainer E. Lopez-Vizcaino <mestherlv32@gmail.com>

Depends R (>= 1.8.0),mixstock,MASS,Matrix

Description mme fit Gaussian Multinomial mixed-effects models for small area estimation: Model 1, with one random effect in each category of the response variable; Model 2, introducing independent time effect; Model 3, introducing correlated time effect. mme calculates analytical and parametric bootstrap MSE estimators.

License GPL (>= 2)

LazyData yes

R topics documented:

`mme-package` *Multinomial Mixed Effects Models*

Description

The mme package implements three multinomial area level mixed effects models for small area estimation. The first model (Model 1) is based on the area level multinomial mixed model with independent random effects for each category of the response variable (Lopez-Vizcaino et al, 2013). The second model (Model 2) takes advantage from the availability of survey data from different time periods and uses a multinomial mixed model with independent random effects for each category of the response variable and with independent time and domain random effects. The third model (Model 3) is similar to the second one, but with correlated time random effects. To fit the models, we combine the penalized quasi-likelihood (PQL) method, introduced by Breslow and Clayton (1993) for estimating and predicting th fixed and random effects, with the residual maximum likelihood (REML) method for estimating the variance components. In all models the package use two approaches to estimate the mean square error (MSE), first through an analytical expression and second by bootstrap techniques.

References

Lopez-Vizcaino, ME, Lombardia, MJ and Morales, D (2013). Multinomial-based small area estimation of labour force indicators. Statistical Modelling, 13 ,153-178.

Lopez-Vizcaino, ME, Lombardia, MJ and Morales, D (2013). Small area estimation of labour force indicator under a multinomial mixed model with correlated time and area effects. Submitted for review.

Breslow, N, Clayton, D (1993). Aproximate inference in generalized linear mixed models. Journal of the American Statistical Association, 88, 9-25.

`addtolist` *Add items from a list*

Description

This function adds items from a list of dimension d*t, where d is the number of areas and t is the number of times periods.

Usage

```
addtolist(B_d, t, d)
```

Arguments

`B_d`	a list in each area.
`t`	number of time periods.
`d`	number of areas.

Value

B_d a list of dimension d.

See Also

`Fbetaf.it`, `Fbetaf.ct`, `modelfit2`, `modelfit3`

Examples

```
k=3  #number of categories for the response variable
pp=c(1,1) #vector with the number of auxiliary variables in each category
data(simdata2) # data
mod=2
datar=data.mme(simdata2,k,pp,mod)

##Add the time periods
l=addtolist(datar$X,datar$t,datar$d)
```

`addtomatrix` *Add rows from a matrix*

Description

This function adds rows from a matrix of dimension d*t*(k-1) times d*(k-1).

Usage

```
addtomatrix(C2, d, t, k)
```

Arguments

`C2`	a matrix of dimension d*t*(k-1) times d*(k-1).
`d`	number of areas.
`t`	number of time periods.
`k`	number of categories of the response variable.

Value

C22 a matrix of dimension d*(k-1) times d*(k-1).

See Also

`Fbetaf.it`, `Fbetaf.ct`, `modelfit2`,`modelfit3`

Examples

```
k=3 #number of categories of the response variable
d=15 # number of areas
t=2  # number of time periods
mat=matrix(1,d*t*(k-1),d*(k-1)) # a matrix

##Add items in the matrix
mat2=addtomatrix(mat,d,t,k)
```

`ci` *Standard deviation and p-values of the estimated model parameters*

Description

This function calculates the standard deviations and the p-values of the estimated model parameters. The standard deviations are obtained from the asymptotic Fisher information matrix in the fitting algorithms `modelfit1`, `modelfit2`, `modelfit3`, depending of the current multinomial mixed model.

Usage

```
ci(a, F)
```

Arguments

`a`	vector with the estimated parameters obtained from `modelfit1`, `modelfit2` or `modelfit3`.
`F`	inverse of the Fisher Information Matrix obtained from `modelfit1`, `modelfit2` or `modelfit3`.

Value

A list containing the following components.

`Std.dev`	vector with the standard deviations of the parameters. The parameters are sorted per category.
`p.value`	vector with the p-values of the parameters for testing H0:a=0.

References

Lopez-Vizcaino, ME, Lombardia, MJ and Morales, D (2013). Multinomial-based small area estimation of labour force indicators. Statistical Modelling, 13, 153-178.

See Also

`modelfit1`, `modelfit2`, `modelfit3`.

Examples

```
library(Matrix)

k=3 #number of categories of the response variable
pp=c(1,1) #vector with the number of auxiliary variables in each category
data(simdata) #data
mod=1  #Type of model
datar=data.mme(simdata,k,pp,mod)

#Model fit
result=modelfit1(pp,datar$Xk,datar$X,datar$Z,datar$initial,datar$y[,1:(k-1)],
       datar$n,datar$N)
beta=result[[8]][,1] #fixed effects
Fisher=result[[3]] #Fisher information matrix

##Standard deviation and p-values
res=ci(beta,Fisher)
```

`data.mme` *Function to generate matrices and the initial values*

Description

Based on the input data, this function generates some matrices that are required in subsequent calculations and the initial values obtained from the function `initial.values`.

Usage

```
data.mme(fi, k, pp, mod)
```

Arguments

`fi` input data set with (d x t) rows and 4+k+sum(pp) columns. The first four columns of the data set are, in this order: the area indicator (integer), the time indicator (integer), the sample size of each domain and the population size of each domain. The following k columns are the categories of the response variable. Then, the auxiliary variables: first, the auxiliary variables of the first category, second, the auxiliary variables of the second category, and so on. Examples of input data sets are in `simdata`, `simdata2` and `simdata3`.

`k` number of categories of the response variable.

`pp` vector with the number of auxiliary variables per category.

`mod` a number specifying the type of models: 1=multinomial mixed model with one independent random effect in each category of the response variable (Model 1), 2=multinomial mixed model with two independent random effects in each category of the response variable: one domain random effect and another independent time and domain random effect (Model 2) and 3= multinomial model with two independent random effects in each category of the response variable: one domain random effect and another correlated time and domain random effect (Model 3).

Value

A list containing the following components.

`n` vector with the area sample sizes.

`N` vector with the area population sizes.

`Z` design matrix of random effects.

`Xk` list of matrices with the auxiliary variables per category. The dimension of the list is the number of domains

`X` list of matrices with the auxiliary variables. The dimension of the list is the number of categories of the response variable minus one.

`y` matrix with the response variable. The rows are the domains and the columns are the categories of the response variable.

`initial` a list with the initial values for the fixed and the random effects obtained from `initial.values` .

References

Lopez-Vizcaino, ME, Lombardia, MJ and Morales, D (2013). Multinomial-based small area estimation of labour force indicators. Statistical Modelling, 13, 153-178.

See Also

`initial.values`, `wmatrix`, `phi.mult`, `prmu`, `Fbetaf`, `phi.direct`, `sPhikf`, `ci`, `modelfit1`, `msef`, `mseb`

Examples

```
k=3 #number of categories of the response variable
pp=c(1,1) #vector with the number of auxiliary variables in each category
data(simdata2) #Data
mod=2

##Needed matrix and initial values
datar=data.mme(simdata2,k,pp,mod)
```

`Fbetaf` *Inverse of the Fisher information matrix of the fixed and random effects in Model 1*

Description

This function calculates the inverse of the Fisher information matrix of the fixed effects (beta) and the random effects (u) and the score vectors S.beta and S.u, for the model with one independent random effect in each category of the response variable (Model 1). `modelfit1` uses the output of this function to estimate the fixed and random effects by the PQL method.

Usage

```
Fbetaf(sigmap, X, Z, phi, y, mu, u)
```

Arguments

`sigmap`	a list with the model variance-covariance matrices for each domain.
`X`	list of matrices with the auxiliary variables obtained from `data.mme`. The dimension of the list is the number of categories of the response variable.
`Z`	design matrix of random effects.
`phi`	vector with the values of the variance components obtained from `modelfit1`.
`y`	matrix with the response variable except the reference category. The rows are the domains and the columns are the categories of the response variable minus one.
`mu`	matrix with the estimated mean of the response variable obtained from `prmu`.
`u`	matrix with the values of random effects obtained from `modelfit1`.

Value

A list containing the following components.

- `F.beta.beta` the first diagonal element of the inverse of the Fisher information matrix.
- `F.beta.u` the element out of the diagonal of the inverse of the Fisher information matrix.
- `F.u.u` the second diagonal element of the inverse of the Fisher information matrix.
- `S.beta` beta scores.
- `S.u` u scores.

References

Lopez-Vizcaino, ME, Lombardia, MJ and Morales, D (2013). Multinomial-based small area estimation of labour force indicators. Statistical Modelling, 13 ,153-178.

See Also

`data.mme`, `initial.values`, `wmatrix`, `phi.mult`, `prmu`, `phi.direct`, `sPhikf`, `ci`, `modelfit1`, `msef`, `mseb`.

Examples

```
require(Matrix)

k=3 #number of categories of the response variable
pp=c(1,1) #vector with the number of auxiliary variables in each category
data(simdata) #data
mod=1 #type of model
datar=data.mme(simdata,k,pp,mod)
initial=datar$initial
mean=prmu(datar$n,datar$Xk,initial$beta.0,initial$u.0)
sigmap=wmatrix(datar$n,mean$estimated.probabilities)

#Inverse of the Fisher information matrix
Fisher=Fbetaf(sigmap,datar$X,datar$Z,initial$phi.0,datar$y[,1:(k-1)],
       mean$mean,initial$u.0)
```

`Fbetaf.ct`	*Inverse of the Fisher information matrix of fixed and random effects in Model 3*

Description

This function calculates the score vector S and the inverse of the Fisher information matrix for the fixed (beta) and the random effects (u1, u2) in Model 3. This model has two independet sets of random effects. The first one contains independent random effects u1dk associated to each category and domain. The second set contains random effects u2dkt associated to each category, domain and time period. Model 3 assumes that the u2dk are AR(1) correlated across time. `modelfit3` uses the output of this function to estimate the fixed and random effect by the PQL method.

Usage

```
Fbetaf.ct(sigmap, X, Z, phi1, phi2, y, mu, u1, u2, rho)
```

Arguments

`sigmap`	a list with the model variance-covariance matrices for each domain obtained from `wmatrix`.
`X`	list of matrices with the auxiliary variables obtained from `data.mme`. The dimension of the list is the number of categories of the response variable minus one.
`Z`	design matrix of random effects.
`phi1`	vector with the values of the variance components for the first random effects obtained from `modelfit3`.
`phi2`	vector with the values of the variance components for the second random effects obtained from `modelfit3`.
`y`	matrix with the response variable, except the reference category. The rows are the domains and the columns are the categories of the response variable minus one.
`mu`	matrix with the estimated mean of the response variable.
`u1`	matrix with the values of the first random effect obtained from `modelfit3`.
`u2`	matrix with the values of the second random effect obtained from `modelfit3`.
`rho`	vector with the values of the correlation parameter obtained from `modelfit3`.

Value

A list containing the following components.

`F`	the inverse of the Fisher information matrix of (beta, u1, u2).
`S`	(beta, u1, u2) score vectors

References

Lopez-Vizcaino, ME, Lombardia, MJ and Morales, D (2013). Small area estimation of labour force indicators under a multinomial mixed model with correlated time and area effects. Submitted for review.

See Also

`data.mme`, `initial.values`, `wmatrix`, `phi.mult.ct`, `prmu.time`, `phi.direct.ct`, `sPhikf.ct`, `ci`, `modelfit3`, `msef.ct`, `omega`, `mseb`

Examples

```
require(Matrix)

k=3 #number of categories of the response variable
pp=c(1,1) #vector with the number of auxiliary variables in each category
mod=3 #type of model
data(simdata3)
datar=data.mme(simdata3,k,pp,mod)
initial=datar$initial
mean=prmu.time(datar$n,datar$Xk,initial$beta.0,initial$u1.0,initial$u2.0)
sigmap=wmatrix(datar$n,mean$estimated.probabilities) #variance-covariance

##The inverse of the Fisher information matrix and the score matrix
Fisher.beta=Fbetaf.ct(sigmap,datar$X,datar$Z,initial$phi1.0,initial$phi2.0,
         datar$y[,1:(k-1)],mean$mean,initial$u1.0,initial$u2.0,initial$rho.0)
```

`Fbetaf.it` *The inverse of the Fisher information matrix of the fixed and random effects for Model 2*

Description

This function calculates the score vector S and the inverse of the Fisher information matrix for the fixed (beta) and the random effects (u1, u2) in Model 2. This model has two independet sets of random effects. The first one contains independent random effects u1dk associated to each category and domain. The second set contains random effects u2dkt associated to each category, domain and time period. Model 2 assumes that the u2dk are independent across time. `modelfit2` uses the output of this function to estimate the fixed and random effect by the PQL method.

Usage

```
Fbetaf.it(sigmap, X, Z, phi1, phi2, y, mu, u1, u2)
```

Arguments

`sigmap`	a list with the model variance-covariance matrices for each domain.
`X`	list of matrices with the auxiliary variables obtained from `data.mme`. The dimension of the list is the number of categories of the response variable minus one.
`Z`	design matrix of random effects obtained from `data.mme`.
`phi1`	vector with the first variance component obtained from `modelfit2`.
`phi2`	vector with the second variance component obtained from `modelfit2`.
`y`	matrix with the response variable, except the reference category obtained from `data.mme`. The rows are the domains and the columns are the categories of the response variable minus one.
`mu`	matrix with the estimated mean of the response variable obtained from `prmu.time`.
`u1`	matrix with the values of the first random effect obtained from `modelfit2`.
`u2`	matrix with the values of the second random effect obtained from `modelfit2`.

Value

A list containing the following components.

`F`	the inverse of the Fisher information matrix.
`S`	(beta, u1, u2) scores

References

Lopez-Vizcaino, ME, Lombardia, MJ and Morales, D (2013). Small area estimation of labour force indicator under a multinomial mixed model with correlated time and area effects. Submitted for review.

See Also

`data.mme`, `initial.values`, `wmatrix`, `phi.mult.it`, `prmu.time`, `phi.direct.it`, `sPhikf.it`, `ci`, `modelfit2`, `msef.it`, `mseb`.

Examples

```
require(Matrix)

k=3 #number of categories of the response variable
pp=c(1,1) #vector with the number of auxiliary variables in each category
mod=2  #Type of model
data(simdata2) #data
datar=data.mme(simdata2,k,pp,mod)
initial=datar$initial
mean=prmu.time(datar$n,datar$Xk,initial$beta.0,initial$u1.0,initial$u2.0)
sigmap=wmatrix(datar$n,mean$estimated.probabilities)

##The inverse of the Fisher information matrix of the fixed effects
Fisher=Fbetaf.it(sigmap,datar$X,datar$Z,initial$phi1.0,initial$phi2.0,
       datar$y[,1:(k-1)],mean$mean,initial$u1.0,initial$u2.0)
```

`initial.values`	*Initial values for fitting algorithm to estimate the fixed and random effects and the variance components*

Description

This function sets the initial values. An iterative algorithm fits the multinomial mixed models that requires initial values for the fixed effects, the random effects and the variance components. This initial values are used in `modelfit1`, `modelfit2` and `modelfit3`.

Usage

```
initial.values(d, pp, datar, mod)
```

Arguments

`d`	number of areas.
`pp`	vector with the number of auxiliary variables per category.
`datar`	output of function `data.mme`.
`mod`	a number specifying the type of model: 1=multinomial mixed model with one independent random effect for each category of the response variable (Model 1), 2=multinomial mixed model with two independent random effects for each category of the response variable: one domain random effect and another independent time and domain random effect (Model 2) and 3= multinomial mixed model with two independent random effects for each category of the response variable: one domain random effect and another correlated time and domain random effect (Model 3).

Value

A list containing the following components, depending on the chosen model.

- `beta.0` a list with the initial values for the fixed effects beta per category.
- `phi.0` vector with the initial values for the variance components phi of Model 1.
- `phi1.0` vector with the initial values for the variance components phi1 of Model 2 or 3.
- `phi2.0` vector with the initial values for the variance components phi2 of Model 2 or 3.
- `u` matrix with the initial values for the random effect for Model 1.
- `u1.0` matrix with the initial values for the first random effect for Model 2 or 3.
- `u2.0` matrix with the initial values for the second random effect for Model 2 or 3.
- `rho.0` vector with the initial values for the correlation parameter for Model 3.

References

Lopez-Vizcaino, ME, Lombardia, MJ and Morales, D (2013). Multinomial-based small area estimation of labour force indicators. Statistical Modelling, 13, 153-178.

Lopez-Vizcaino, ME, Lombardia, MJ and Morales, D (2013). Small area estimation of labour force indicators under a multinomial mixed model with correlated time and area effects. Submitted for review.

See Also

`data.mme`, `wmatrix`, `phi.mult.it`, `prmu.time`, `Fbetaf.it`, `phi.direct.it`, `sPhikf.it`, `ci`, `modelfit2`, `msef.it`, `mseb`

Examples

```
library(Matrix)

k=3 #number of categories of the response variable
pp=c(1,1) #vector with the number of auxiliary variables in each category
data(simdata)
D=nrow(simdata)
mod=1 #Type of model
datar=data.mme(simdata,k,pp,mod)

## Initial values for fixed, random effects and variance components
initial=datar$initial
```

`mmedata` *Create objects of class mmedata*

Description

This function creates objects of class mmedata.

Usage

```
mmedata(fi, k, pp)
```

Arguments

`fi`	input data set with (d X t) rows and 4+k+sum(pp) columns. The first four columns of the data set are, in this order: the area indicator (integer), the time indicator (integer), the sample size of each domain and the population size of each domain. The following k columns are the categories of the response variable. Then, the auxiliary variables: first, the auxiliary variables of the first category, second, the auxiliary variables of the second category, and so on. Examples of input data set are in `simdata`, `simdata2` and `simdata3`.
`k`	number of categories of the response variable.
`pp`	vector with the number of auxiliary variables per category.

See Also

`modelfit1`, `modelfit2`, `modelfit3`

Examples

```
k=3 #number of categories of the response variable
pp=c(1,1) #vector with the number of auxiliary variables in each category
data(simdata)
r=mmedata(simdata,k,pp)
```

`model`	*Choose between the three models*

Description

This function chooses one of the three models.

Usage

```
model(d, t, pp, Xk, X, Z, initial, y, M, MM, mod)
```

Arguments

`d`	number of areas.
`t`	number of time periods.
`pp`	vector with the number of the auxiliary variables per category.
`Xk`	list of matrices with the auxiliary variables per category obtained from `data.mme`. The dimension of the list is the number of domains.
`X`	list of matrices with the auxiliary variables obtained from `data.mme`. The dimension of the list is the number of categories of the response variable minus one.
`Z`	design matrix of random effects obtained from `data.mme`.
`initial`	output of the function `initial.values`.
`y`	matrix with the response variable obtained from `data.mme`. The rows are the domains and the columns are the categories of the response variable.
`M`	vector with the area sample sizes.
`MM`	vector with the population sample sizes.
`mod`	a number specifying the type of models: 1=multinomial mixed model with one independent random effect in each category of the response variable (Model 1), 2=multinomial mixed model with two independent random effects in each category of the response variable: one domain random effect and another independent time and domain random effect (Model 2) and 3= multinomial model with two independent random effects in each category of the response variable: one domain random effect and another correlated time and domain random effect (Model 3).

Value

the output of the function `modelfit1`, `modelfit2` or `modelfit3`.

References

Lopez-Vizcaino, ME, Lombardia, MJ and Morales, D (2013). Multinomial-based small area estimation of labour force indicators. Statistical Modelling, 13 ,153-178.

Lopez-Vizcaino, ME, Lombardia, MJ and Morales, D (2013). Small area estimation of labour force indicator under a multinomial mixed model with correlated time and area effects. Submitted for review.

Examples

```
k=3 #number of categories of the response variable
pp=c(1,1) #vector with the number of auxiliary variables in each category
data(simdata)  #data
mod=1 #Model 1
datar=data.mme(simdata,k,pp,mod)
result=model(datar$d,datar$t,pp,datar$Xk,datar$X,datar$Z,datar$initial,
datar$y[,1:(k-1)],datar$n,datar$N, mod)
```

`modelfit1` *Function used to fit Model 1*

Description

This function fits the multinomial mixed model with one independent random effect per category of the response variable (Model 1), like in the formulation described in Lopez-Vizcaino et al. (2013). The fitting algorithm combines the penalized quasi-likelihood method (PQL) for estimating and predicting the fixed and random effects with the residual maximum likelihood method (REML) for estimating the variance components. This function uses as initial values the output of the function `initial.values`

Usage

```
modelfit1(pp, Xk, X, Z, initial, y, M, MM)
```

Arguments

`pp`	vector with the number of the auxiliary variables per category.
`Xk`	list of matrices with the auxiliary variables per category obtained from `data.mme`. The dimension of the list is the number of domains.
`X`	list of matrices with the auxiliary variables obtained from `data.mme`. The dimension of the list is the number of categories of the response variable minus one.
`Z`	design matrix of random effects obtained from `data.mme`.

`initial` output of the function `initial.values`.

`y` matrix with the response variable except the reference category obtained from `data.mme`. The rows are the domains and the columns are the categories of the response variable minus 1.

`M` vector with the area sample sizes.

`MM` vector with the population sample sizes.

Value

A list containing the following components.

`Estimated.probabilities`
matrix with the estimated probabilities for the categories of response variable.

`Fisher.information.matrix.phi`
Fisher information matrix of the random effect.

`Fisher.information.matrix.beta`
Fisher information matrix of the fixed effect.

`u` matrix with the estimated random effects.

`mean` matrix with the estimated mean of the response variable.

`warning1` 0=OK,1=The model could not be fitted.

`warning2` 0=OK,1=The value of the variance component is negative: the initial value is taken.

`beta.Stddev.p.value`
matrix with the estimated fixed effects, its standard deviations and its p-values.

`phi.Stddev.p.value`
matrix with the estimated variance components, its standard deviations and its p-values.

References

Lopez-Vizcaino, ME, Lombardia, MJ and Morales, D (2013). Multinomial-based small area estimation of labour force indicators. Statistical Modelling, 13, 153-178.

See Also

`data.mme`, `initial.values`, `wmatrix`, `phi.mult`, `prmu`, `phi.direct`, `sPhikf`, `ci`, `Fbetaf`, `msef`, `mseb`.

Examples

```
library(Matrix)

k=3 #number of categories of the response variable
pp=c(1,1) #vector with the number of auxiliary variables in each category
data(simdata)  #data
mod=1 #type of model
datar=data.mme(simdata,k,pp,mod)

#Model fit
result=modelfit1(pp,datar$Xk,datar$X,datar$Z,datar$initial,datar$y[,1:(k-1)],
      datar$n,datar$N)
```

`modelfit2` *Function to fit Model 2*

Description

This function fits the multinomial mixed model with two independent random effects for each category of the response variable: one domain random effect and another independent time and domain random effect (Model 2). The formulation is described in Lopez-Vizcaino et al. (2013). The fitting algorithm combines the penalized quasi-likelihood method (PQL) for estimating and predicting the fixed and random effects, respectively, with the residual maximum likelihood method (REML) for estimating the variance components. This function uses as initial values the output of the function `initial.values`.

Usage

```
modelfit2(d, t, pp, Xk, X, Z, initial, y, M, MM)
```

Arguments

`d`	number of areas.
`t`	number of time periods.
`pp`	vector with the number of the auxiliary variables per category.
`Xk`	list of matrices with the auxiliary variables per category obtained from `data.mme`. The dimension of the list is the number of domains.
`X`	list of matrices with the auxiliary variables obtained from `data.mme`. The dimension of the list is the number of categories of the response variable minus one.
`Z`	design matrix of random effects `data.mme`.

`initial`	output of the function `initial.values`.
`y`	matrix with the response variable obtained from `data.mme`, except the reference category. The rows are the domains and the columns are the categories of the response variable minus one.
`M`	vector with the area sample sizes.
`MM`	vector with the population sample sizes.

Value

A list containing the following components.

`Estimated.probabilities`
: matrix with the estimated probabilities for the categories of response variable.

`Fisher.information.matrix.phi`
: Fisher information matrix of the variance components.

`Fisher.information.matrix.beta`
: Fisher information matrix of the fixed effects.

`u1`
: matrix with the estimated first random effect.

`u2`
: matrix with the estimated second random effect.

`mean`
: matrix with the estimated mean of response variable.

`warning1`
: 0=OK,1=The model could not be fitted.

`warning2`
: 0=OK,1=The value of the variance component is negative: the initial value is taken.

`beta.Stddev.p.value`
: matrix with the estimated fixed effects, its standard deviations and its p-values.

`phi.Stddev.p.value`
: matrix with the estimated variance components, its standard deviations and its p-values.

References

Lopez-Vizcaino, ME, Lombardia, MJ and Morales, D (2013). Small area estimation of labour force indicators under a multinomial mixed model with correlated time and area effects. Submitted for review.

See Also

`data.mme`, `initial.values`, `wmatrix`, `phi.mult.it`, `prmu.time`, `phi.direct.it`, `sPhikf.it`, `ci`, `Fbetaf.it`, `msef.it`, `mseb`

Examples

```
library(mixstock)
library(Matrix)

k=3 #number of categories of the response variable
pp=c(1,1) #vector with the number of auxiliary variables in each category
mod=2 #type of model
data(simdata2)  #data
datar=data.mme(simdata2,k,pp,mod)

##Model fit
result=modelfit2(datar$d,datar$t,pp,datar$Xk,datar$X,datar$Z,datar$initial,
       datar$y[,1:(k-1)],datar$n,datar$N)
```

`modelfit3` *Function used to fit Model 3*

Description

This function fits the multinomial mixed model with two independent random effects for each category of the response variable: one domain random effect and another correlated time and domain random effect (Model 3). The formulation is described in Lopez-Vizcaino et al. (2013). The fitting algorithm combine the penalized quasi-likelihood method (PQL) for estimating and predicting the fixed and random effects, respectively, with the residual maximun likelihood method (REML) for estimating the variance components. This function uses as initial values the output of the function `initial.values`.

Usage

```
modelfit3(d, t, pp, Xk, X, Z, initial, y, M, MM, b)
```

Arguments

`d`	number of areas.
`t`	number of time periods.
`pp`	vector with the number of the auxiliary variables per category.
`Xk`	list of matrices with the auxiliary variables per category obtained from `data.mme`. The dimension of the list is the number of domains.
`X`	list of matrices with the auxiliary variables obtained from `data.mme`. The dimension of the list is the number of categories of the response variable minus one.

`Z`	design matrix of random effects obtained from `data.mme`.
`initial`	output of the function `initial.values`.
`y`	matrix with the response variable obtained from `data.mme`, except the reference category. The rows are the domains and the columns are the categories of the response variable minus one.
`M`	vector with the area sample sizes.
`MM`	vector with the population sample sizes.
`b`	parameter that indicates the bootstrap.

Value

A list containing the following components.

`Estimated.probabilities`	matrix with the estimated probabilities for the categories of response variable.
`Fisher.information.matrix.phi`	Fisher information matrix of phi.
`Fisher.information.matrix.beta`	Fisher information matrix of beta.
`u1`	matrix with the estimated first random effect.
`u2`	matrix with the estimated second random effect.
`mean`	matrix with the estimated mean of the response variable.
`warning1`	0=OK,1=The model could not be fitted.
`warning2`	0=OK,1=The value of the variance component is negative: the initial value is taken.
`beta.Stddev.p.value`	matrix with the estimated fixed effects, its standard deviations and its p-values.
`phi.Stddev.p.value`	matrix with the estimated variance components, its standard deviations and its p-values.
`rho`	estimated correlation parameter.
`rho.Stddev.p.value`	matrix with the estimated correlation parameter, its standard deviations and its p-values.

References

Lopez-Vizcaino, ME, Lombardia, MJ and Morales, D (2013). Small area estimation of labour force indicators under a multinomial mixed model with correlated time and area effects. Submitted for review.

See Also

`data.mme`, `initial.values`, `wmatrix`, `phi.mult.ct`, `prmu.time`, `phi.direct.ct`, `sPhikf.ct`, `omega`, `ci`, `Fbetaf.ct`, `msef.ct`, `mseb`

Examples

```
## Not run:
library(mixstock)
library(Matrix)
library(MASS)

k=3 #number of categories of the response variable
pp=c(1,1) #vector with the number of auxiliary variables in each category
mod=3 #type of model
data(simdata3)   #data
datar=data.mme(simdata3,k,pp,mod)

##Model fit
result=modelfit3(datar$d,datar$t,pp,datar$Xk,datar$X,datar$Z,datar$initial,
       datar$y[,1:(k-1)],datar$n,datar$N,0)
## End(Not run)
```

`mseb` *Bias and MSE using parametric bootstrap*

Description

This function calculates the bias and the mse for the multinomial mixed effects models using parametric bootstrap. Three types of multinomial mixed models are considered, with one independent domain random effect in each category of the response variable (Model 1), with two random effects: the first, with a domain random effect and with independent time and domain random effect (Model 2) and the second, with a domain random effect and with correlated time and domain random effect (Model 3). See details of the parametric bootstrap procedure in Gonzalez-Manteiga et al. (2008) and in Lopez-Vizcaino et al. (2013) for the adaptation to these three models. This function uses the output of `modelfit1`, `modelfit2` or `modelfit3`, depending of the current multinomial mixed model.

Usage

```
mseb(pp, Xk, X, Z, M, MM, resul, B, mod)
```

Arguments

`pp`	vector with the number of the auxiliary variables per category.
`Xk`	list of matrices with the auxiliary variables per category obtained from `data.mme`. The dimension of the list is the number of domains.
`X`	list of matrices with the auxiliary variables obtained from `data.mme`. The dimension of the list is the number of categories of the response variable minus one.
`Z`	design matrix of random effects obtained from `data.mme`.
`M`	vector with the area sample sizes.
`MM`	vector with the population sample sizes.
`resul`	output of the function `modelfit1`, `modelfit2` or `modelfit3`.
`B`	number of bootstrap replications.
`mod`	a number specifying the type of models: 1=multinomial mixed model with one independent random effect in each category of the response variable (Model 1), 2=multinomial mixed model with two independent random effects in each category of the response variable: one domain random effect and another independent time and domain random effect (Model 2) and 3= multinomial model with two independent random effects in each category of the response variable: one domain random effect and another correlated time and domain random effect (Model 3).

Value

a list containing the following components.

`bias.pboot`	BIAS of the parametric bootstrap estimator of the mean of the response variable
`mse.pboot`	MSE of the parametric bootstrap estimator of the mean of the response variable
`rmse.pboot`	RMSE of the parametric bootstrap estimator of the mean of the response variable

References

Lopez-Vizcaino, ME, Lombardia, MJ and Morales, D (2013). Multinomial-based small area estimation of labour force indicators. Statistical Modelling, 13 ,153-178.

Lopez-Vizcaino, ME, Lombardia, MJ and Morales, D (2013). Small area estimation of labour force indicator under a multinomial mixed model with correlated time and area effects. Submitted for review.

Gonzalez-Manteiga, W, Lombardia, MJ, Molina, I, Morales, D, Santamaria, L (2008). Estimation of the mean squared error of predictors of small area linear parameters under a logistic mixed model, Computational Statistics and Data Analysis, 51, 2720-2733.

See Also

`data.mme`, `initial.values`, `wmatrix`, `phi.mult`, `phi.mult.it`, `phi.mult.ct`, `prmu`, `prmu.time`, `phi.direct`, `phi.direct.it`, `phi.direct.ct`, `sPhikf`, `sPhikf.it`, `sPhikf.ct`, `modelfit1`, `modelfit2`, `modelfit3`, `omega`, `Fbetaf`, `Fbetaf.it`, `Fbetaf.ct`, `ci`.

Examples

```
require(Matrix)

k=3 #number of categories of the response variable
pp=c(1,1) #vector with the number of auxiliary variables in each category
data(simdata)
mod=1  # Type of model
datar=data.mme(simdata,k,pp,mod)
##Model fit
result=modelfit1(pp,datar$Xk,datar$X,datar$Z,datar$initial,
datar$y[,1:(k-1)],datar$n,datar$N)

B=1    #Bootstrap iterations
ss=12345 #SEED
set.seed(ss)

##Bootstrap parametric BIAS and MSE
mse.pboot=mseb(pp,datar$Xk,datar$X,datar$Z,datar$n,datar$N,result,B,mod)
```

---

msef — *Analytic MSE for Model 1*

Description

This function calculates the analytic MSE for the multinomial mixed model with one independent random effect per category of the response variable (Model 1). See Lopez-Vizcaino et al. (2013), section 4, for details. The formulas of Prasad and Rao (1990) are adapted to Model 1. This function uses the output of `modelfit1`.

Usage

```
msef(pp, X, Z, resul, MM, M)
```

Arguments

`resul`	the output of the function `modelfit1`.
`X`	list of matrices with the auxiliary variables obtained from `data.mme`. The dimension of the list is the number of categories of the response variable minus one.
`Z`	design matrix of random effects obtained from `data.mme`.
`pp`	vector with the number of the auxiliary variables per category.
`M`	vector with the area sample sizes.
`MM`	vector with the population sample sizes.

Value

mse is a matrix with the MSE estimator calculated by adapting the explicit formulas of Prasad and Rao (1990).

References

Lopez-Vizcaino, ME, Lombardia, MJ and Morales, D (2013). Multinomial-based small area estimation of labour force indicators. Statistical Modelling, 13, 153-178.

Prasad, NGN, Rao, JNK (1990).The estimation of the mean squared error of small area estimators. Journal of the American Statistical Association, 85, 163-171.

See Also

`data.mme`, `initial.values`, `wmatrix`, `phi.mult`, `prmu`, `phi.direct`, `sPhikf`, `modelfit1`, `Fbetaf`, `ci`, `mseb`.

Examples

```
require(Matrix)

k=3 #number of categories of the response variable
pp=c(1,1) #vector with the number of auxiliary variables in each category
data(simdata)  #data
mod=1 # type of model
datar=data.mme(simdata,k,pp,mod)
# Model fit
result=modelfit1(pp,datar$Xk,datar$X,datar$Z,datar$initial,datar$y[,1:(k-1)],
       datar$n,datar$N)

#Analytic MSE
mse=msef(pp,datar$X,datar$Z,result,datar$N,datar$n)
```

`msef.ct` *Analytic MSE for Model 3*

Description

This function calculates the analytic MSE for the multinomial mixed model with two independent random effects for each category of the response variable: one random effect associated with the domain and another correlated random effect associated with time and domain (Model 3). See details of the model and the expresion of mse in Lopez-Vizcaino et al. (2013). The formulas of Prasad and Rao (1990) are adapted to Model 3. This function uses the output of `modelfit3`.

Usage

```
msef.ct(p, X, result, M, MM)
```

Arguments

`p`	vector with the number of the auxiliary variables per category.
`X`	list of matrices with the auxiliary variables obtained from `data.mme`. The dimension of the list is the number of categories of the response variable minus one.
`result`	the output of the function `modelfit3`.
`M`	vector with the area sample sizes.
`MM`	vector with the population sample sizes.

Value

mse.analitic is a matrix with the MSE estimator calculated by adapting the explicit formulas of Prasad and Rao (1990).

References

Lopez-Vizcaino, ME, Lombardia, MJ and Morales, D (2013). Small area estimation of labour force indicators under a multinomial mixed model with correlated time and area effects. Submitted for review.

Prasad, NGN, Rao, JNK (1990).The estimation of the mean squared error of small area estimators. Journal of the American Statistical Association, 85, 163-171.

See Also

`data.mme`, `initial.values`, `wmatrix`, `phi.mult.ct`, `prmu.time`, `phi.direct.ct`, `sPhikf.ct`, `modelfit3`, `Fbetaf.ct`, `ci`, `omega`, `mseb`.

Examples

```
## Not run:
library(mixstock)
library(Matrix)
library(MASS)

k=3 #number of categories of the response variable
pp=c(1,1) #vector with the number of auxiliary variables in each category
mod=3 #type of model
data(simdata3) #data
datar=data.mme(simdata3,k,pp,mod)
##Model fit
result=modelfit3(d,t,pp,datar$Xk,datar$X,datar$Z,datar$initial,
       datar$y[,1:(k-1)],datar$n,datar$N,0)

##Analytic MSE
msef=msef.ct(pp,datar$X,result,datar$n,datar$N)
## End(Not run)
```

`msef.it` *Analytic MSE for Model 2*

Description

This function calculates the analytic MSE for the multinomial mixed model with two independent random effects for each category of the response variable: one random effect associated with the domain and another independent random effect associated with time and domain (Model 2). See details of the model and the expresion of mse in Lopez-Vizcaino et al. (2013). The formulas of Prasad and Rao (1990) are adapted to Model 2. This function uses the output of `modelfit2`.

Usage

```
msef.it(p, X, result, M, MM)
```

Arguments

`p`	vector with the number of the auxiliary variables per category.
`X`	list of matrices with the auxiliary variables obtained from `data.mme`. The dimension of the list is the number of categories of the response variable minus one.
`result`	the output of the function `modelfit2`.
`M`	vector with the area sample sizes.
`MM`	vector with the population sample sizes.

Value

mse.analitic is a matrix with the MSE estimator calculated by adapting the explicit formulas of Prasad and Rao (1990). The matrix dimension is the number of domains multiplied by the number of categories minus 1.

References

Lopez-Vizcaino, ME, Lombardia, MJ and Morales, D (2013). Small area estimation of labour force indicator under a multinomial mixed model with correlated time and area effects. Submitted for review.

Prasad, NGN, Rao, JNK (1990).The estimation of the mean squared error of small area estimators. Journal of the American Statistical Association, 85, 163-171.

See Also

`data.mme`, `initial.values`, `wmatrix`, `phi.mult.it`, `prmu.time`, `phi.direct.it`, `sPhikf.it`, `modelfit2`, `Fbetaf.it`, `ci`, `mseb`

Examples

```
library(mixstock)
library(Matrix)

k=3 #number of categories of the response variable
pp=c(1,1) #vector with the number of auxiliary variables in each category
mod=2 #type of model
data(simdata2)
datar=data.mme(simdata2,k,pp,mod)
##Model fit
result=modelfit2(datar$d,datar$t,pp,datar$Xk,datar$X,datar$Z,datar$initial,
       datar$y[,1:(k-1)],datar$n,datar$N)

##Analytic MSE
msef=msef.it(pp,datar$X,result,datar$n,datar$N)
```

`omega` *Model correlation matrix for Model 3*

Description

This function calculates the model correlation matrix and the first derivative of the model correlation matrix for Model 3. Model 3 is the multinomial mixed model with two independent random effects for each category of the response variable: one domain random effect and another correlated time and domain random effect.

Usage

```
omega(t, k, rho, phi2)
```

Arguments

`t`	number of time periods.
`k`	number of categories of the response variable.
`rho`	vector with the correlation parameter obtained from `modelfit3`.
`phi2`	vector with the values of the second variance component obtained from `modelfit3`.

Value

A list containing the following components.

`Omega.d` correlation matrix.

`First.derivative.Omegad`
Fisher derivative of the model correlation matrix.

References

Lopez-Vizcaino, ME, Lombardia, MJ and Morales, D (2013). Small area estimation of labour force indicator under a multinomial mixed model with correlated time and area effects. Submitted for review.

See Also

`data.mme`, `initial.values`, `wmatrix` ,`phi.mult.ct`, `prmu.time`, `phi.direct.ct`, `Fbetaf.ct`, `sPhikf.ct`, `ci`, `modelfit3`, `msef.ct`, `mseb`

Examples

```
require(Matrix)

k=3 #number of categories of the response variable
pp=c(1,1) #vector with the number of auxiliary variables in each category
mod=3 #type of model
data(simdata3)   #data
datar=data.mme(simdata3,k,pp,mod)
initial=datar$initial
mean=prmu.time(datar$n,datar$Xk,initial$beta.0,initial$u1.0,initial$u2.0)

##The model correlation matrix
matrix.corr=omega(datar$t,k,initial$rho.0,initial$phi2.0)
```

`phi.direct`	*Variance components for Model 1*

Description

This function calculates the variance components for the multinomial mixed model with one independent random effect in each category of the response variable (Model 1). These values are used in the second part of the fitting algorithm implemented in `modelfit1`. The algorithm adapts the ideas of Schall (1991) to a multivariate model and the variance components are estimated by the REML method.

Usage

```
phi.direct(sigmap, phi, X, u)
```

Arguments

`sigmap`	a list with the model variance-covariance matrices for each domain obtained from `wmatrix`.
`X`	list of matrices with the auxiliary variables obtained from `data.mme`. The dimension of the list is the number of categories of the response variable minus one.
`phi`	vector with the initial values of the variance components obtained from `modelfit1`.
`u`	matrix with the values of the random effects obtained from `modelfit1`.

Value

a list containing the following components.

`phi.new`	vector with the variance components.

References

Lopez-Vizcaino, ME, Lombardia, MJ and Morales, D (2013). Multinomial-based small area estimation of labour force indicators. Statistical Modelling, 13 ,153-178.

Schall, R (1991). Estimation in generalized linear models with random effects. Biometrika, 78,719-727.

See Also

`data.mme`, `initial.values`, `wmatrix`, `phi.mult`, `prmu`, `Fbetaf`, `sPhikf`, `ci`, `modelfit1`, `msef`, `mseb`.

Examples

```
require(Matrix)

k=3 #number of categories of the response variable
pp=c(1,1) #vector with the number of auxiliary variables in each category
data(simdata)   #data
mod=1 #type of model
datar=data.mme(simdata,k,pp,mod)
initial=datar$initial
mean=prmu(datar$n,datar$Xk,initial$beta.0,initial$u.0)
#model variance-covariance matrix
sigmap=wmatrix(datar$n,mean$estimated.probabilities)

##Variance components
phi=phi.direct(sigmap,initial$phi.0,datar$X,initial$u.0)
```

`phi.direct.ct` *Variance components for Model 3*

Description

This function calculates the variance components for the multinomial mixed model with two independent random effects for each category of the response variable: one domain random effect and another correlated time and domain random effect (Model 3). This variance components are used in the second part of the fitting algorithm implemented in `modelfit3`. The algorithm adapts the ideas of Schall (1991) to a multivariate model. The variance components are estimated by the REML method.

Usage

```
phi.direct.ct(p, sigmap, X, theta, phi1, phi2, u1, u2,
  rho)
```

Arguments

`p`	vector with the number of auxiliary variables per category.
`sigmap`	a list with the model variance-covariance matrices for each domain obtained from `wmatrix`.
`X`	list of matrices with the auxiliary variables obtained from `data.mme`. The dimension of the list is the number of categories of the response variable minus one.
`theta`	matrix with the estimated log-probabilites of each category in front of the reference category obtained from `prmu.time`.

`phi1` vector with the initial values of the first variance component obtained from `modelfit3`.

`phi2` vector with the initial values of the second variance component obtained from `modelfit3`.

`u1` matrix with the values of the first random effect obtained from `modelfit3`.

`u2` matrix with the values of the second random effect obtained from `modelfit3`.

`rho` vector with the initial values of the correlation parameter obtained from `modelfit3`.

Value

a list containing the following components.

`phi1.new` vector with the values of the variance component for the first random effect.

`phi2.new` vector with the values of the variance component for the second random effect.

`rho.new` vector with the correlation parameter.

References

Lopez-Vizcaino, ME, Lombardia, MJ and Morales, D (2013). Small area estimation of labour force indicator under a multinomial mixed model with correlated time and area effects. Submitted for review.

Schall, R (1991). Estimation in generalized linear models with random effects. Biometrika, 78,719-727.

See Also

`data.mme`, `initial.values`, `wmatrix`, `phi.mult.ct`, `prmu.time`, `Fbetaf.ct` `sPhikf.ct`, `ci`, `modelfit3`, `msef.ct`, `mseb`, `omega`

Examples

```
require(Matrix)

k=3 #number of categories of the response variable
pp=c(1,1) #vector with the number of auxiliary variables in each category
mod=3 #type of model
data(simdata3) #data
datar=data.mme(simdata3,k,pp,mod)
initial=datar$initial
mean=prmu.time(datar$n,datar$Xk,initial$beta.0,initial$u1.0,initial$u2.0)
sigmap=wmatrix(datar$n,mean$estimated.probabilities)
```

```
##The variance components
phi.ct=phi.direct.ct(pp,sigmap,datar$X,mean$eta,initial$phi1.0,
       initial$phi2.0,initial$u1.0,initial$u2.0,initial$rho.0)
```

`phi.direct.it` *Variance components for Model 2*

Description

This function calculates the variance components for the multinomial mixed model with two independent random effects for each category of the response variable: one domain random effect and another independent time and domain random effect (Model 2). This variance components are used in the second part of the fitting algorithm implemented in `modelfit2`. The algorithm adapts the ideas of Schall (1991) to a multivariate model. The variance components are estimated by the REML method.

Usage

```
phi.direct.it(pp, sigmap, X, phi1, phi2, u1, u2)
```

Arguments

`pp`	vector with the number of auxiliary variables per category.
`sigmap`	a list with the model variance-covariance matrices for each domain obtained from `wmatrix`.
`X`	list of matrices with the auxiliary variables obtained from `data.mme`. The dimension of the list is the number of categories of the response variable.
`phi1`	vector with the initial values of the first variance component obtained from `modelfit2`.
`phi2`	vector with the initial values of the second variance component obtained from `modelfit2`.
`u1`	matrix with the values of the first random effect obtained from `modelfit2`.
`u2`	matrix with the values of the second random effect obtained from `modelfit2`.

Value

a list containing the following components.

`phi1.new`	vector with the values of the variance component for the first random effect.
`phi2.new`	vector with the values of the variance component for the second random effect.

References

Lopez-Vizcaino, ME, Lombardia, MJ and Morales, D (2013). Small area estimation of labour force indicators under a multinomial mixed model with correlated time and area effects. Submitted for review.

Schall, R (1991). Estimation in generalized linear models with random effects. Biometrika, 78,719-727.

See Also

`data.mme`, `initial.values`, `wmatrix`, `phi.mult.it`, `prmu.time`, `Fbetaf.it` `sPhikf.it`, `ci`, `modelfit2`, `msef.it`, `mseb`

Examples

```
require(Matrix)

k=3 #number of categories of the response variable
pp=c(1,1) #vector with the number of auxiliary variables in each category
d=10 #number of areas
mod=2  #Type of model
data(simdata2)   #data
datar=data.mme(simdata2,k,pp,mod)
initial=datar$initial
mean=prmu.time(datar$n,datar$Xk,initial$beta.0,initial$u1.0,initial$u2.0)
sigmap=wmatrix(datar$n,mean$estimated.probabilities) #variance-covariance

## The variance components
phi.it=phi.direct.it(pp,sigmap,datar$X,initial$phi1.0,initial$phi2.0,
       initial$u1.0,initial$u2.0)
```

`phi.mult` *Initial values for the variance components for Model 1*

Description

This function is used in `initial.values` to calculate the initial values for the variance components in the multinomial mixed model with one independent random effect in each category of the response variable (Model 1).

Usage

```
phi.mult(beta.0, y, Xk, M)
```

Arguments

`beta.0`	initial values for the fixed effects obtained in `initial.values`.
`y`	matrix with the response variable obtained from `data.mme`. The rows are the domains and the columns are the categories of the response variable.
`Xk`	list of matrices with the auxiliary variables per category obtained from `data.mme`. The dimension of the list is the number of domains.
`M`	vector with the sample size of the areas.

Value

phi.0 vector of inicial values for the variance components

References

Lopez-Vizcaino, ME, Lombardia, MJ and Morales, D (2013). Multinomial-based small area estimation of labour force indicators. Statistical Modelling, 13, 153-178.

See Also

`data.mme`, `initial.values`, `wmatrix`, `prmu`, `Fbetaf`, `phi.direct`, `sPhikf`, `ci`, `modelfit1`, `msef`, `mseb`.

Examples

```
require(Matrix)

k=3 #number of categories of the response variable
pp=c(1,1) #vector with the number of auxiliary variables in each category
data(simdata) #data
mod=1 #type of model
datar=data.mme(simdata,k,pp,mod)
###beta values
beta.new=list()
beta.new[[1]]=matrix(c( 1.3,-1),2,1)
beta.new[[2]]=matrix(c( -1.6,1),2,1)

##Initial variance components
phi=phi.mult(beta.new,datar$y,datar$Xk,datar$n)
```

`phi.mult.ct`	*Initial values for the variance components in Model 3*

Description

This function is used in `initial.values` to calculate the initial values for the variance components in the multinomial mixed model with two independent random effects for each category of the response variable: one domain random effect and another correlated time and domain random effect (Model 3).

Usage

```
phi.mult.ct(beta.0, y, Xk, M, u1, u2)
```

Arguments

`beta.0`	a list with the initial values for the fixed effects per category obtained from `initial.values`.
`y`	matrix with the response variable obtained from `data.mme`. The rows are the domains and the columns are the categories of the response variable minus one.
`Xk`	list of matrices with the auxiliary variables per category obtained from `data.mme`. The dimension of the list is the number of domains.
`M`	vector with the sample size of the areas.
`u1`	matrix with the values for the first random effect obtained from `initial.values`.
`u2`	matrix with the values for the second random effect obtained from `initial.values`.

Value

A list containing the following components.

`phi.0`	vector of the initial values for the variance components.
`rho.0`	vector of the initial values for the correlation parameter.

References

Lopez-Vizcaino, ME, Lombardia, MJ and Morales, D (2013). Small area estimation of labour force indicators under a multinomial mixed model with correlated time and area effects. Submitted for review.

See Also

`data.mme`, `initial.values`, `wmatrix`, `prmu.time`, `Fbetaf.ct`, `phi.direct.ct`, `sPhikf.ct`, `ci`, `modelfit3`, `msef.ct`,`omega`, `mseb`.

Examples

```
require(Matrix)

k=3 #number of categories of the response variable
pp=c(1,1) #vector with the number of auxiliary variables in each category
mod=3  #type of model
data(simdata3) #data
D=nrow(simdata3)
datar=data.mme(simdata3,k,pp,mod)
###Fixed effects values
beta.new=list()
beta.new[[1]]=matrix(c( 1.3,-1),2,1)
beta.new[[2]]=matrix(c( -1.6,1),2,1)
## Random effects values
u1.new=rep(0.01,((k-1)*datar$d))
dim(u1.new)=c(datar$d,k-1)
u2.new=rep(0.01,((k-1)*D))
dim(u2.new)=c(D,k-1)

## Initial variance components
phi=phi.mult.ct(beta.new,datar$y,datar$Xk,datar$n,u1.new,u2.new)
```

`phi.mult.it` *Initial values for the variance components in Model 2*

Description

This function is used in `initial.values` to calculate the initial values for the variance components in the multinomial mixed model with two independent random effects for each category of the response variable: one domain random effect (u1) and another independent time and domain random effect (u2) (Model 2).

Usage

```
phi.mult.it(beta.0, y, Xk, M, u1, u2)
```

Arguments

beta.0	initial values for the fixed effects obtained from `initial.values`.
y	matrix with the response variable obtained from `data.mme`. The rows are the domains and the columns are the categories of the response variable.
Xk	list of matrices with the auxiliary variables per category obtained from `data.mme`. The dimension of the list is the number of domains.
M	vector with the sample size of the areas.
u1	vector with the initial values for the first random effect obtained from `initial.values`.
u2	vector with the initial values for the second random effect obtained from `initial.values`.

Value

phi.0 vector of the initial values for the variance components.

References

Lopez-Vizcaino, ME, Lombardia, MJ and Morales, D (2013). Multinomial-based small area estimation of labour force indicators. Statistical Modelling, 13, 153-178.

See Also

`data.mme`, `initial.values`, `wmatrix`, `prmu.time`, `Fbetaf.it`, `phi.direct.it`, `sPhikf.it`, `ci`, `modelfit2`, `msef.it`, `mseb`.

Examples

```
require(Matrix)

k=3 #number of categories of the response variable
pp=c(1,1) #vector with the number of auxiliary variables in each category
data(simdata2)  #data
mod=2 #Type of model
datar=data.mme(simdata2,k,pp,mod)
D=nrow(simdata2)
###fixed effects values
beta.new=list()
beta.new[[1]]=matrix(c( 1.3,-1),2,1)
beta.new[[2]]=matrix(c( -1.6,1),2,1)
## random effects values
u1.new=rep(0.01,((k-1)*datar$d))
dim(u1.new)=c(datar$d,k-1)
u2.new=rep(0.01,((k-1)*D))
```

```
dim(u2.new)=c(D,k-1)

##Initial variance components
phi=phi.mult.it(beta.new,datar$y,datar$Xk,datar$n,u1.new,u2.new)
```

`print.mme` *Print objects of class mme*

Description

This function prints objects of class mme.

Usage

```
 ## S3 method for class mme
print(x, ...)
```

Arguments

`x`	a list with the output of `modelfit1`, `modelfit2` or `modelfit3`.
`...`	further information.

See Also

`modelfit1`, `modelfit2`, `modelfit3`

Examples

```
k=3 #number of categories of the response variable
pp=c(1,1) #vector with the number of auxiliary variables in each category
mod=1  # Type of model
data(simdata)
datar=data.mme(simdata,k,pp,mod)
##Model fit
result=modelfit1(pp,datar$Xk,datar$X,datar$Z,datar$initial,
datar$y[,1:(k-1)],datar$n,datar$N)
result
```

Description

This function calculates the estimated probabilities and the estimated mean of the response variable, in the multinomial mixed model with one independent random effect in each category of the response variable (Model 1).

Usage

```
prmu(M, Xk, beta, u)
```

Arguments

`M`	vector with the area sample sizes.
`Xk`	list of matrices with the auxiliary variables per category obtained from `data.mme`. The dimension of the list is the number of domains.
`beta`	fixed effects obtained from `modelfit1`.
`u`	values of random effects obtained from `modelfit1`.

Value

A list containing the following components:

`Estimated.probabilities`	matrix with the estimated probabilities for the categories of response variable.
`mean`	matrix with the estimated mean of the response variable.
`eta`	matrix with the estimated log-rates of the probabilities of each category over the reference category.

References

Lopez-Vizcaino, ME, Lombardia, MJ and Morales, D (2013). Multinomial-based small area estimation of labour force indicators. Statistical Modelling, 13, 153-178.

See Also

`data.mme`, `initial.values`, `wmatrix`, `phi.mult`, `Fbetaf`, `phi.direct`, `sPhikf`, `ci`, `modelfit1`, `msef`, `mseb`.

Examples

```
require(Matrix)

k=3 #number of categories of the response variable
pp=c(1,1) #vector with the number of auxiliary variables in each category
data(simdata) #data
mod=1 #type of model
D=nrow(simdata)
datar=data.mme(simdata,k,pp,mod)
initial=datar$initial

##Estimated mean and probabilities
mean=prmu(datar$n,datar$Xk,initial$beta.0,initial$u.0)
```

`prmu.time` *Estimated mean and probabilities for Model 2 and 3*

Description

This function calculates the estimated probabilities and the estimated mean of the response variable, in the multinomial mixed models with two independent random effects, one random effect associated with the area and the other associated with the time, for each category of the response variable. The first model assumes independent time and domain random effect (Model 2) and the second model assumes correlated time and domain random effect (Model 3).

Usage

```
prmu.time(M, Xk, beta, u1, u2)
```

Arguments

`M`	vector with the area sample sizes.
`Xk`	list of matrices with the auxiliary variables per category obtained from `data.mme`. The dimension of the list is the number of domains.
`beta`	a list with the values for the fixed effects beta per category obtained from `modelfit2`.
`u1`	a vector with the values of the first random effect obtained from `modelfit2` or `modelfit3`.
`u2`	a vector with the values of the second random effect obtained from `modelfit2` or `modelfit3`.

Value

A list containing the following components:

`Estimated.probabilities`
: matrix with the estimated probabilities for the categories of response variable.

`mean`
: matrix with the estimated mean of the response variable.

`eta`
: matrix with the estimated log-rates of the probabilities of each category over the reference category.

References

Lopez-Vizcaino, ME, Lombardia, MJ and Morales, D (2013). Small area estimation of labour force indicator under a multinomial mixed model with correlated time and area effects. Submited for review.

See Also

`data.mme`, `initial.values`, `wmatrix`, `phi.mult.it`, `Fbetaf.it`, `phi.direct.it`, `sPhikf.it`, `ci`, `modelfit2`, `msef.it`, `mseb`

Examples

```
require(Matrix)

k=3 #number of categories of the response variable
pp=c(1,1) #vector with the number of auxiliary variables in each category
mod=2 #Type of model
data(simdata2) # data
datar=data.mme(simdata2,k,pp,mod)
initial=datar$initial

## Estimated mean and estimated probabilities
mean=prmu.time(datar$n,datar$Xk,initial$beta.0,initial$u1.0,initial$u2.0)
```

`simdata` *Dataset for Model 1*

Description

Dataset used by the multinomial mixed effects model with one independent random effect in each category of the response variable (Model 1). This dataset contains 15 small areas. The response variable has three categories. The last is the reference category. The variables are as follows:

Format

A data frame with 15 rows and 9 variables in columns

Details

- area: area indicator.
- Time: time indicator.
- sample: the sample size of each domain.
- Population: the population size of each domain.
- Y1: the first category of the response variable.
- Y2: the second category of the response variable.
- Y3: the third category of the response variable.
- X1: the covariate for the first category of the response variable.
- X2: the covariate for the second category of the response variable.

Examples

```
require(Matrix)

k=3 #number of categories of the response variable
pp=c(1,1) #vector with the number of auxiliary variables in each category
data(simdata)  #data
mod=1 # type of model
datar=data.mme(simdata,k,pp,mod)
# Model fit
result=model(datar$d,datar$t,pp,datar$Xk,datar$X,datar$Z,datar$initial,
       datar$y[,1:(k-1)],datar$n,datar$N,mod)

#Analytic MSE
mse=msef(pp,datar$X,datar$Z,result,datar$N,datar$n)

B=1    #Bootstrap iterations
ss=12345 #SEED
set.seed(ss)

##Bootstrap parametric BIAS and MSE
mse.pboot=mseb(pp,datar$Xk,datar$X,datar$Z,datar$n,datar$N,result,B,mod)
```

simdata2 *Dataset for Model 2*

Description

Dataset used by the multonomial mixed effects model with two independent random effects in each category of the response variable: one domain random effect and another independent time and domain random effect (Model 2). This dataset contains 10 small areas and two periods. The response variable has three categories. The last is the reference category. The variables are as follows:

Format

A data frame with 30 rows and 9 variables in columns

Details

- area: area indicator.
- Time: time indicator.
- sample: the sample size of each domain.
- Population: the population size of each domain.
- Y1: the first category of the response variable.
- Y2: the second category of the response variable.
- Y3: the third category of the response variable.
- X1: the covariate for the first category of the response variable.
- X2: the covariate for the second category of the response variable.

Examples

```
library(mixstock)
library(Matrix)

k=3 #number of categories of the response variable
pp=c(1,1) #vector with the number of auxiliary variables in each category
data(simdata2)
mod=2 #type of model
datar=data.mme(simdata2,k,pp,mod)

##Model fit
result=model(datar$d,datar$t,pp,datar$Xk,datar$X,datar$Z,datar$initial,
       datar$y[,1:(k-1)],datar$n,datar$N,mod)
```

```
##Analytic MSE
msef=msef.it(pp,datar$X,result,datar$n,datar$N)

B=1    #Bootstrap iterations
ss=12345 #SEED
set.seed(ss)

##Bootstrap parametric BIAS and MSE
mse.pboot=mseb(pp,datar$Xk,datar$X,datar$Z,datar$n,datar$N,result,B,mod)
```

`simdata3` *Dataset for Model 3*

Description

Dataset used by the multonomial mixed effects model with two independent random effects in each category of the response variable: one domain random effect and another correlated time and domain random effect (Model 3). This dataset contains ten small areas and four periods. The response variable has three categories. The last is the reference category. The variables are as follows:

Format

A data frame with 40 rows and 9 variables in columns

Details

- area: area indicator.
- Time: time indicator.
- sample: the sample size of each domain.
- Population: the population size of each domain.
- Y1: the first category of the response variable.
- Y2: the second category of the response variable.
- Y3: the third category of the response variable.
- X1: the covariate for the first category of the response variable.
- X2: the covariate for the second category of the response variable.

Examples

```
## Not run:
library(mixstock)
library(Matrix)
library(MASS)

k=3 #number of categories of the response variable
pp=c(1,1) #vector with the number of auxiliary variables in each category
mod=3 #type of model
data(simdata3) #data
datar=data.mme(simdata3,k,pp,mod)
##Model fit
result=model(datar$d,datar$t,pp,datar$Xk,datar$X,datar$Z,datar$initial,
       datar$y[,1:(k-1)],datar$n,datar$N,mod)

##Analytic MSE
msef=msef.ct(pp,datar$X,result,datar$n,datar$N)

B=1    #Bootstrap iterations
ss=12345 #SEED
set.seed(ss)

##Bootstrap parametric BIAS and MSE
mse.pboot=mseb(pp,datar$Xk,datar$X,datar$Z,datar$n,datar$N,result,B,mod)
## End(Not run)
```

`sPhikf` *Fisher information matrix and score vectors of the variance components for Model 1*

Description

This function computes the Fisher information matrix and the score vectors of the variance components, for the multinomial mixed model with one independent random effect in each category of the response variable (Model 1). These values are used in the fitting algorithm implemented in `modelfit1` to estimate the random effects. The algorithm adatps the ideas of Schall (1991) to a multivariate model. The variance components are estimated by the REML method.

Usage

```
sPhikf(pp, sigmap, X, eta, phi)
```

Arguments

`pp`	vector with the number of the auxiliary variables per category.
`sigmap`	a list with the model variance-covariance matrices for each domain obtained from `wmatrix`.
`X`	list of matrices with the auxiliary variables obtained from `data.mme`. The dimension of the list is the number of categories of the response variable minus one.
`eta`	matrix with the estimated log-rates of probabilities of each category over the reference category obtained from `prmu`.
`phi`	vector with the values of the variance components obtained from `modelfit1`.

Value

A list containing the following components.

`S.k`	phi score vector.
`F`	Fisher information matrix of the variance component phi.

References

Lopez-Vizcaino, ME, Lombardia, MJ and Morales, D (2013). Multinomial-based small area estimation of labour force indicators. Statistical Modelling, 13 ,153-178.

Schall, R (1991). Estimation in generalized linear models with random effects. Biometrika, 78,719-727.

See Also

`data.mme`, `initial.values`, `wmatrix`, `phi.mult`, `prmu`, `phi.direct`, `Fbetaf`, `ci`, `modelfit1`, `msef`, `mseb`.

Examples

```
require(Matrix)

k=3 #number of categories of the response variable
pp=c(1,1) #vector with the number of auxiliary variables in each category
data(simdata) #data
mod=1 #type of model
datar=data.mme(simdata,k,pp, mod)
initial=datar$initial
mean=prmu(datar$n,datar$Xk,initial$beta.0,initial$u.0)
sigmap=wmatrix(datar$n,mean$estimated.probabilities)

##Fisher information matrix and score vectors
Fisher.phi=sPhikf(pp,sigmap,datar$X,mean$eta,initial$phi.0)
```

`sPhikf.ct`	*Fisher information matrix and score vectors of the variance components for Model 3*

Description

This function computes the Fisher information matrix and the score vectors of the variance components, for the multinomial mixed model with two independent random effects for each category of the response variable: one domain random effect and another correlated time and domain random effect (Model 3). These values are used in the fitting algorithm implemented in `modelfit3` to estimate the random effects. The algorithm adatps the ideas of Schall (1991) to a multivariate model. The variance components are estimated by the REML method.

Usage

```
sPhikf.ct(d, t, pp, sigmap, X, eta, phi1, phi2, rho, pr,
  M)
```

Arguments

`d`	number of areas.
`t`	number of time periods.
`pp`	vector with the number of the auxiliary variables per category.
`sigmap`	a list with the model variance-covariance matrices for each domain obtained from `wmatrix`.
`X`	list of matrices with the auxiliary variables obtained from `data.mme`. The dimension of the list is the number of categories of the response variable minus one.
`eta`	matrix with the estimated log-rates of probabilites of each category over the reference category obtained from `prmu.time`.
`phi1`	vector with the values of the first variance component obtained from `modelfit3`.
`phi2`	vector with the values of the second variance component obtained from `modelfit3`.
`rho`	vector with the correlation parameter obtained from `modelfit3`.
`pr`	matrix with the estimated probabilities of the response variable obtained from `prmu.time`.
`M`	vector with the area sample sizes.

Value

A list containing the following components.

`S` (phi1, phi2, rho) score vector.

`F` Fisher information matrix of the variance components (phi1, phi2, rho).

References

Lopez-Vizcaino, ME, Lombardia, MJ and Morales, D (2013). Small area estimation of labour force indicators under a multinomial mixed model with correlated time and area effects. Submitted for review.

Schall, R (1991). Estimation in generalized linear models with random effects. Biometrika, 78,719-727.

See Also

`data.mme`, `initial.values`, `wmatrix`, `phi.mult.ct`, `prmu.time`, `phi.direct.ct`, `Fbetaf.ct`, `omega`, `ci`, `modelfit3`, `msef.ct`, `mseb`

Examples

```
require(Matrix)

k=3 #number of categories of the response variable
pp=c(1,1) #vector with the number of auxiliary variables in each category
mod=3 #type of model
data(simdata3) #data
datar=data.mme(simdata3,k,pp, mod)
initial=datar$initial
mean=prmu.time(datar$n,datar$Xk,initial$beta.0,initial$u1.0,initial$u2.0)
sigmap=wmatrix(datar$n,mean$estimated.probabilities)

## Fisher information matrix and the score vectors
Fisher.phi.ct=sPhikf.ct(datar$d,datar$t,pp,sigmap,datar$X,mean$eta,initial$phi1.0,
          initial$phi2.0,initial$rho.0,mean$estimated.probabilities,datar$n)
```

`sPhikf.it` *Fisher information matrix and score vectors of the variance components for Model 2*

Description

This function computes the Fisher information matrix and the score vectors of the variance components, for the multinomial mixed model with two independent random effects for each category of the response variable: one domain random effect and another independent time and domain random effect (Model 2). These values are used in the fitting algorithm implemented in `modelfit2` to estimate the random effects. The algorithm adatps the ideas of Schall (1991) to a multivariate model. The variance components are estimated by the REML method.

Usage

```
sPhikf.it(d, t, pp, sigmap, X, eta, phi1, phi2)
```

Arguments

`d`	number of areas.
`t`	number of time periods.
`pp`	vector with the number of the auxiliary variables per category.
`sigmap`	a list with the model variance-covariance matrices for each domain obtained from `wmatrix`.
`X`	list of matrices with the auxiliary variables obtained from `data.mme`. The dimension of the list is the number of categories of the response variable minus one.
`eta`	matrix with the estimated log-rates of probabilities of each category over the reference category obtained from `prmu.time`.
`phi1`	vector with the values of the first variance component obtained from `modelfit2`.
`phi2`	vector with the values of the second variance component obtained from `modelfit2`.

Value

A list containing the following components.

`S`	phi score vector.
`F`	Fisher information matrix of the variance components phi1 and phi2.

References

Lopez-Vizcaino, ME, Lombardia, MJ and Morales, D (2013). Small area estimation of labour force indicators under a multinomial mixed model with correlated time and area effects. Submitted for review.

Schall, R (1991). Estimation in generalized linear models with random effects. Biometrika, 78,719-727.

See Also

`data.mme`, `initial.values`, `wmatrix`, `phi.mult.it`, `prmu.time`, `phi.direct.it`, `Fbetaf.it`, `ci`, `modelfit2`, `msef.it`, `mseb`

Examples

```
require(Matrix)

k=3 #number of categories of the response variable
pp=c(1,1) #vector with the number of auxiliary variables in each category
mod=2 #Type of model
data(simdata2) #data
datar=data.mme(simdata2,k,pp,mod)
initial=datar$initial
mean=prmu.time(datar$n,datar$Xk,initial$beta.0,initial$u1.0,initial$u2.0)
sigmap=wmatrix(datar$n,mean$estimated.probabilities)

##Fisher information matrix and score vectors
Fisher.phi=sPhikf.it(datar$d,datar$t,pp,sigmap,datar$X,mean$eta,initial$phi1.0,
          initial$phi2.0)
```

`wmatrix` *Model variance-covariance matrix of the multinomial mixed models*

Description

This function calculates the variance-covariance matrix of the multinomial mixed models. Three types of multinomial mixed model are considered. The first model (Model 1), with one random effect in each category of the response variable; Model 2, introducing independent time effect; Model 3, introducing correlated time effect.

Usage

```
wmatrix(M, pr)
```

Arguments

`M`	vector with area sample sizes.
`pr`	matrix with the estimated probabilities for the categories of the response variable obtained from `prmu` or `prmu.time`.

Value

W a list with the model variance-covariance matrices for each domain.

References

Lopez-Vizcaino, ME, Lombardia, MJ and Morales, D (2013). Multinomial-based small area estimation of labour force indicators. Statistical Modelling,13,153-178.

See Also

data.mme, initial.values, phi.mult, prmu, prmu.time Fbetaf, phi.direct, sPhikf, ci, modelfit1, msef, mseb

Examples

```
require(Matrix)

k=3 #number of categories of the response variable
pp=c(1,1) #vector with the number of auxiliary variables in each category
mod=2 #type of model
data(simdata2)
datar=data.mme(simdata2,k,pp,mod)
initial=datar$initial
mean=prmu.time(datar$n,datar$Xk,initial$beta.0,initial$u1.0,initial$u2.0)
##The model variance-covariance matrix
varcov=wmatrix(datar$n,mean$estimated.probabilities)
```

Appendix A

Resumen en español

El paro es actualmente un problema de primer orden. Las encuestas sociológicas lo sitúan habitualmente como una de las preocupaciones principales de los ciudadanos, y la lucha contra el paro es prioritaria en la actuación política de todos los niveles de la Administración Pública. Es más, en el contexto de crisis en el que está inmersa la Unión Europea, en España el impacto de esta crisis en el mercado de trabajo ha sido mucho más intenso que en la mayoría de las economías avanzadas, la tasa de paro en el tercer trimestre del 2013 alcanzó el 25.98%, más de 14 puntos superior a la del 2008. En Galicia la situación del mercado de trabajo no es muy diferente, la tasa de paro está en el 21.6% y el número de parados alcanza ya las 277 mil personas. En esta situación los políticos de todos los niveles de la administración planifican y actuan con el objetivo de reducir el paro.

En general las medidas políticas de carácter global no suelen ser satisfactorias para las entidades locales, que también pueden desarrollar sus propias estrategias de empleo. Para ello necesitan algunas herramientas que les permitan determinar -con precisión, fiabilidad y puntualidad aceptables- las principales variables e indicadores del mercado laboral: empleo, paro, población activa, tasa de empleo y paro, y la ocupación en función de sexo, edad y sector de actividad, entre otros, para así poder desarrollar sus competencias. En España, como en otros países europeos, la estimación se realiza mediante la Encuesta de Población Activa (EPA), que utiliza un diseño estratificado con el criterio principal del tamaño del municipio (INE, 2009). La mayoría de los municipios no están representados en la muestra, y muchos de los que sí lo están tienen un tamaño muestral muy reducido, lo que hace que las estimaciones a nivel municipal tengan una precisión inaceptable. El problema que se plantea, entonces, es el de la insuficiencia de tamaño muestral, e incluso ausencia total en algunos casos, para llevar a cabo tales estimaciones. Ante esta situación se puede ampliar el tamaño muestral, pero esto, además de originar un evidente aumento en los costes y en la carga de respuesta a los informantes, puede conllevar otras pérdidas de calidad debidas a posibles retrasos en la obtención de los resultados y al impacto de los errores ajenos al muestreo. Por tanto, el aumento del tamaño muestral no es siempre aconsejable e incluso algunas veces inviable desde el punto de vista económico.

El interés por desarrollar técnicas de estimación para áreas pequeñas que permitan resolver razonablemente estos problemas es creciente entre los investigadores de la Estadística. El término ”área pequeña”, se utiliza frecuentemente para referirse a zonas geográficas, pero también se puede aplicar a otros dominios de interés como límites no geográficos, grupos de edad, sectores de actividad, etc. Es el tamaño muy reducido de la muestra en el dominio, y consecuentemente la gran varianza de los estimadores ”directos” lo que define el área pequeña, y no el tamaño del área en sí mismo. En el contexto de la estimación en áreas pequeñas, se dice que un estimador de un parámetro en un dominio dado es directo si está basado solamente en los datos muestrales del dominio específico. El problema de estos estimadores es que cuando no hay observaciones muestrales en alguna área de interés no se pueden calcular.

En este trabajo el objetivo es la estimación de las variables relacionadas con el mercado laboral (ocupados, parados y tasa de paro) en las comarcas gallegas utilizando técnicas de estimación en áreas pequeñas con modelos multinomiales mixtos de área. Los parados y ocupados se pueden estimar mediante dos modelos lineales mixtos separados, relacionando las estimaciones directas de las respectivas proporciones con otras variables auxiliares. En esta situación nadie nos asegura que las estimaciones caigan en el intervalo $[0, 1]$, lo cual es una desventaja importante. Otra desventaja es que estos modelos no tienen en cuenta la relación natural existente entre los parados, ocupados y la población inactiva, pues la suma de las tres categorías es el total poblacional de 16 y más años. Estas desventajas se pueden superar usando modelos logísticos. Estos modelos se usaron en los trabajos de Saei and Chambers (2003), Molina et al. (2007), Morales et al. (2007) y González-Manteiga et al. (2008b).

Por tanto, para llevar a cabo nuestro objetivo usaremos estimadores EBLUP basados en modelos multinomiales mixtos a nivel de área ya introducidos por Molina et al. (2007), en los que se asume una distribución multinomial logística conjunta con efectos aleatorios de área para las proporciones de parados y ocupados. En el modelo descrito en Molina et al. (2007), solo se consideraba un efecto aleatorio de área para cada una de las áreas y, por tanto, este efecto era el mismo para las dos clases multinomiales (ocupados y parados). En el problema que nos ocupa en este trabajo esta situación puede no ser apropiada, motivada por las características tan diferentes de estos dos colectivos en la comunidad gallega, lo cual llevó a que consideráramos dos efectos aleatorios, uno para cada una de las categorías multinomiales. Además, la disponibilidad de series de tiempo permite un aumento significativo de la muestra total en el dominio de estudio, lo cual nos llevó a introducir en el modelo efectos de tiempo independientes y correlacionados tal y como ya se ha hecho en otros trabajos de características similares (Pfeffermann and Burck, 1990; Rao and Yu, 1994; Saei and Chambers, 2003; Tiller, 1992; Ugarte et al., 2009a). En este caso, se aplican los modelos de área a los datos muestrales y para ilustrar el proceso se utilizan datos de la EPA de la Comunidad Autónoma de Galicia.

La finalidad principal de la EPA es conocer la actividad económica de la población en lo relativo a su componente humano. Está orientada a dar datos de las principales categorías poblacionales en relación con el mercado de trabajo (ocupados, parados, activos, inactivos) y a obtener clasificaciones de estas categorías según diversas características. La EPA está basada en definiciones y criterios internacionales y sus resultados permiten la comparabilidad con otros países europeos. Se trata de una encuesta trimestral que utiliza un muestreo estratificado bietápico con estratificación de las unidades muestrales primarias (UMP). Las UMP están constituídas por secciones censales que son áreas geográficas con un máximo de 500 viviendas o aproximadamente 3000 personas. Las unidades de segunda etapa están constituídas por las viviendas familiares principales y los alojamientos fijos. Dentro de las unidades de segunda etapa no se realiza submuestreo alguno, recogiéndose información de todas las personas que tengan su residencia en las mismas.

En este trabajo los dominios de interés son las comarcas cruzadas con el sexo. En Galicia hay 53 comarcas pero vamos a considerar $D = 96$ o $D = 102$ dominios, obtenidos de cruzar las 48 o 51 comarcas en las que hay muestra con los 2 sexos. Utilizaremos datos desde el tercer trimestre del 2009 hasta el cuarto trimestre del 2011, i.e., para $T = 10$ periodos. Como hay 48 o 51 comarcas en la EPA de Galicia, tenemos $D = 96$ o $D = 102$ dominios P_{dt} que denotan la población del dominio d en el periodo de tiempo t. Estos dominios los podemos particionar en los subconjuntos P_{d1t}, P_{d2t} y P_{d3t} de ocupados, parados e inactivos. Nuestro parámetro poblacional objetivo es el total de ocupados parados y las tasas de paro, esto es

$$Y_{dkt} = \sum_{j \in P_{dt}} y_{dktj},\ k = 1, 2,\ \ R_{dt} = \frac{Y_{d2t}}{Y_{d1t} + Y_{d2t}},$$

donde $y_{dktj} = 1$ si el individuo j del dominio d y el periodo de tiempo t está en P_{dtk} y $y_{dktj} = 0$ en otro caso.

La EPA no produce estimaciones oficiales a nivel de dominio, pero los estimadores directos análogos del total Y_{dkt}, de la media $\bar{Y}_{dkt} = Y_{dkt}/N_{dt}$, del tamaño N_{dt} y de la tasa R_{dt} son

$$\hat{Y}_{dkt}^{dir} = \sum_{j \in S_{dt}} w_{dtj}\, y_{dktj},\ \hat{\bar{Y}}_{dkt}^{dir} = \hat{Y}_{dkt}^{dir}/\hat{N}_{dt}^{dir},\ k = 1, 2,\ \ \hat{N}_{dt}^{dir} = \sum_{j \in S_{dt}} w_{dtj},\ \hat{R}_{dt}^{dir} = \frac{\hat{Y}_{d2t}^{dir}}{\hat{Y}_{d1t}^{dir} + \hat{Y}_{d2t}^{dir}}, \tag{A.0.1}$$

donde S_{dt} es el dominio muestral y los w_{dtj} son los pesos oficiales calibrados teniendo en cuenta la no respuesta.

En este punto describimos los modelos multinomiales mixtos empleados en este trabajo. Definimos los efectos aleatorios $u_{1,dk}$ y $u_{2,dkt}$ asociados al dominio d, la categoría k y el periodo de tiempo t, respectivamente. Consideramos dos conjuntos de efectos aleatorios. El primero es $\boldsymbol{u}_1 = (\boldsymbol{u}_{1,1}', \dots, \boldsymbol{u}_{1,D}')'$, con $\boldsymbol{u}_{1,d} = (u_{1,d1}, u_{1,d2})'$. El segundo es $\boldsymbol{u}_2 = (\boldsymbol{u}_{2,1}', \dots, \boldsymbol{u}_{2,D}')'$, con $\boldsymbol{u}_{2,d} = (\boldsymbol{u}_{2,d1}', \boldsymbol{u}_{2,d2}')'$, $\boldsymbol{u}_{2,dk} = (u_{2,dk1}, \dots, u_{2,dkT})'$, $k = 1, 2$, y $\boldsymbol{u}_{2,dt} = (u_{2,d1t}, u_{2,d2t})'$. La variable objetivo es $\boldsymbol{y} = (\boldsymbol{y}_1', \dots, \boldsymbol{y}_D')'$, donde $\boldsymbol{y}_d = (\boldsymbol{y}_{d1}', \dots, \boldsymbol{y}_{dT}')'$ and $\boldsymbol{y}_{dt} = (y_{d1t}, y_{d2t})'$, $d = 1, \dots, D$, $t = 1, \dots, T$. Entonces, el modelo principal (Modelo 3) asume que

1. $\boldsymbol{u}_1$ y $\boldsymbol{u}_2$ son independientes,

2. $\boldsymbol{u}_1 \sim N(\boldsymbol{0}, \boldsymbol{V}_{u_1})$, donde $\boldsymbol{V}_{u_1} = \underset{1 \le d \le D}{\text{diag}}\,(\text{diag}(\varphi_{11}, \varphi_{12}))$.

3. $\boldsymbol{u}_{2,dk} \sim N(\mathbf{0}, \boldsymbol{V}_{u_{2,dk}})$, $d = 1, \ldots, D$, $k = 1, 2$, son independientes con matriz de covarianzas AR(1), i.e. $\boldsymbol{V}_{u_{2,dk}} = \varphi_{2k}\Omega(\phi_k)$ y

$$\Omega(\phi_k) = \frac{1}{1-\phi_k^2}\begin{pmatrix} 1 & \phi_k & \ldots & \phi_k^{T-2} & \phi_k^{T-1} \\ \phi_k & 1 & \ddots & & \phi_k^{T-2} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \phi_k^{T-2} & & \ddots & 1 & \phi_k \\ \phi_k^{T-1} & \phi_k^{T-2} & \ldots & \phi_k & 1 \end{pmatrix}_{T\times T}.$$

El Modelo 3 también asume que la variable respuesta $\boldsymbol{y}_{dt}$, condicionada a $\boldsymbol{u}_{1,d}$ y $\boldsymbol{u}_{2,dt}$, es independiente con distribución multinomial

$$\boldsymbol{y}_{dt}|_{\boldsymbol{u}_{1,d},\boldsymbol{u}_{2,dt}} \sim \mathrm{M}(n_{dt}, p_{d1t}, p_{d2t}),\ d = 1, \ldots, D,\ t = 1, \ldots, T. \tag{A.0.2}$$

Para los parámetros naturales $\eta_{dkt} = \log(p_{dkt}/p_{d3t})$, el Model 3 asume

$$\eta_{dkt} = \boldsymbol{x}_{dkt}\boldsymbol{\beta}_k + u_{1,dk} + u_{2,dkt}, \quad d = 1, \ldots, D,\ k = 1, 2,\ t = 1, \ldots, T, \tag{A.0.3}$$

donde $\boldsymbol{x}_{dkt} = (x_{dkt1}, \ldots, x_{dktp_k})'$ and $\boldsymbol{\beta}_k = (\beta_{k1}, \ldots, \beta_{kp_k})'$. Equivalentemente, podemos escribir

$$p_{dkt} = \frac{\exp\{\eta_{dkt}\}}{1 + \exp\{\eta_{d1t}\} + \exp\{\eta_{d2t}\}}, \quad d = 1, \ldots, D,\ k = 1, 2,\ t = 1, \ldots, T.$$

Del Modelo 3, se pueden derivar los restantes dos modelos empleados en este trabajo. El Modelo 2 se obtiene haciendo en el Model 3 $\phi_1 = \phi_2 = 0$ y por tanto tiene efectos aleatorios independientes $u_{2,dkt}$. El Model 1 se obtiene considerando que el Modelo 2 es para un periodo de tiempo ($T = 1$) y considerando solo un efecto aleatorio $\boldsymbol{u}_1$. Este es el modelo estudiado en López-Vizcaíno et al. (2013a). El modelo de Molina et al. (2007) se obtiene haciendo $u_{1,d1} = u_{1,d2}$ en el Modelo 1.

El ajuste de los modelos se llevó a cabo usando una combinación del método de máxima verosimilitud penalizada (PQL), introducido por Breslow and Clayton (1996), para la estimación de β_{kr}'s, los $u_{1,dk}$'s y los $u_{2,dkt}$'s, con la máxima verosimilitud restringida (REML) para la estimación de las componentes de la varianza φ_{1k}, φ_{2k} and ϕ_k, $k = 1, \ldots, q-1$. El método se basa en una aproximación normal de la distribución de probabilidad conjunta del vector $(\boldsymbol{y}, \boldsymbol{u})$. Este algoritmo combinado lo introdujo Schall (1991), y posteriormente lo utilizó Saei and Chambers (2003) en el contexto de estimación de áreas pequeñas con modelos lineales generales mixtos. En este trabajo se adapta el método de ajuste al modelo multinomial mixto introducido.

En la práctica estamos interesados en estimar los totales de los dominios

$$Y_{dkt} = \sum_{j\in P_{dt}} y_{dktj}, \quad d = 1, \ldots, D,\ k = 1, \ldots, q-1,\ t = 1, \ldots, T,$$

donde P_{dt} es la población del dominio con tamaño N_{dt}. Un estimador sintético de Y_{dkt} es $\hat{Y}_{dkt} = \hat{m}_{dkt} = \hat{N}_{dt}\hat{p}_{dkt}$. Las estimaciones para las tasas se pueden obtener usando los correspondientes estimadores de los totales con el método plugging. El interés está en estimar $\boldsymbol{m}_{dt} = \hat{N}_{dt}\boldsymbol{p}_{dt}$, $d = 1, \dots, D, t = 1, \dots, T$, donde $\hat{N}_{dt}$ es un estimador del tamaño de la población que puede ser obtenido de los microdatos de la encuesta o de algún registro administrativo. Como estamos en un contexto de un modelo de área, $\hat{N}_{dt}$ se trata como una constante. En la aplicación a datos reales, se tomará $\hat{N}_{dt} = \hat{N}_{dt}^{dir}$ y se estimará $\boldsymbol{m}_{dt}$ por medio de $\hat{\boldsymbol{m}}_{dt} = \hat{N}_{dt}\hat{\boldsymbol{p}}_{dt}$, donde

$$\hat{\boldsymbol{p}}_{dt} = (\hat{p}_{d1t}, \hat{p}_{d2t})', \quad \hat{p}_{dkt} = \frac{\exp\{\hat{\eta}_{dkt}\}}{1 + \exp\{\hat{\eta}_{d1t}\} + \exp\{\hat{\eta}_{d2t}\}}, \quad \hat{\eta}_{dkt} = \boldsymbol{x}_{dkt}\hat{\boldsymbol{\beta}}_k + \hat{u}_{1,dk} + \hat{u}_{2,dkt}, \tag{A.0.4}$$

y $\hat{\boldsymbol{\beta}}_k$, $\hat{u}_{1,dk}$ and $\hat{u}_{2,dkt}$ se obtienen de la salida del algoritmo de ajuste PQL.

Estos estimadores suelen ser sesgados, por eso la acuracidad de las estimaciones es un tema fundamental en la SAE de la que se han publicado diversas aproximaciones en la literatura. El hecho de que los estimadores sean sesgados hay que complementarlo con ganancia en acuracidad. De ahí que en este trabajo se utilicen diferentes aproximaciones al error cuadrático medio (MSE), mediante una expresión analítica y mediante técnicas bootstrap. En este punto es deseable tener en cuenta que en las estadísticas de fuerza de trabajo la Office for National Statistics (ONS) del Reino Unido considera que una estimación es publicable, y por tanto oficial, si su coeficiente de variación es inferior al 20% (ONS, 2004).

Una vez decididos los modelos, los estimadores y la estimación del MSE, necesitamos obtener variables auxiliares para llevar a cabo el ajuste de los modelos. Las variables auxiliares que se utilizaron provienen de registros administrativos y de los datos que nos proporciona la EPA. Se utilizan las proporciones de personas en los dominios y dentro de las categorías que definen las siguientes agrupaciones.

- SEXAGE: Combinaciones de sexo y grupos de edad, con 6 valores. El sexo se codifica 1 para hombres y 2 para mujeres y la edad está categorizada en 3 grupos con códigos 1 para 16-24, 2 para 25-54 y 3 para $\geq$55. Lós códigos $1, 2, \dots, 6$ se usan para los pares *sexo-edad* $(1, 1), (1, 2), \dots, (2, 3)$.

- STUD: Esta variable describe el nivel de educación alcanzado, con valores 1-3 para los analfabetos y educación primaria, la secundaria y el nivel de educación superior, respectivamente.

- REG: Esta variable indica si un individuo está registrado o no como desempleado en las oficinas de empleo público. El desempleo se puede medir a través de los datos proporcionados por las oficinas de empleo público y mediante los datos proporcionados por la EPA (variable de estudio). Los datos de desempleo en estas dos fuentes son diferentes pero correlacionados. Nosotros estamos interesados en la definición de la EPA, pues esta es la que sigue las recomendaciones de la Organización Internacional del Trabajo (OIT) y EUROSTAT; es decir mide el desempleo como el número de personas sin empleo que quieren trabajar, están disponibles para trabajar y están buscando activamente empleo.

- SS: Esta variable indica si un individuo está afiliado a la Seguridad Social en alta laboral.

Ahora los modelos se ajustan a los datos y al conjunto de variables auxiliares significativas. La seleccíon de los modelos se hace utilizando herramientas descriptivas y de contraste. Finalmente, se obtienen las estimaciones basadas en el modelo de los totales de ocupados, parados y tasas de paro y sus errores cuadráticos medios. Si se comparan estas estimaciones con las que se obtienen con los estimadores directos de la EPA, se puede ver que se comportan mucho mejor en términos de MSE, sobre todo para las comarcas con tamaño de muestra pequeño, constituyendo esta metodología una alternativa para hacer estimaciones a nivel comarcal. Otra ventaja de estos estimadores basados en el modelo es que tiene la propiedad de que son consistentes en el sentido de que la suma de ocupados, parados e inactivos es la población total del dominio (población de 16 y más años).

Además en este trabajo se propone el uso de dos modelos con efectos de área y de tiempo combinados. El hecho de incluír un efecto de tiempo hace que obtengamos unas estimaciones de los ocupados, parados y tasas de paro con más precisión y más estables que si se ajustaran modelos independientes para cada período temporal. Estas características hacen que esta metodología sea atractiva para su uso en las oficinas de estadística oficial, pues las estimaciones siguen el patrón de los estimadores directos para las comarcas grandes y para las comarcas pequeñas se comportan con estabilidad, con lo cual el efecto del suavizado por el uso de varios periodos de tiempo parece razonable.

Además los predictores basados en el modelo permiten hacer predicciones para dominios sin muestra, para los cuales no existen los estimadores basados en el diseño. Esto no quiere decir que los estimadores basados en el diseño no tengan ningún papel en la predicción basada en modelos. Los estimadores basados en el diseño son los datos de entrada para los modelos y se utilizan para la evaluación de los estimadores basados en modelos y para la calibración de estos a los datos provinciales. La calibración tiene la ventaja de garantizar la consistencia en la publicación de los estimadores basados en el modelo y los estimadores basados en el diseño para áreas más grandes. Esto lo requieren a menudo las oficinas de estadística.

También se han propuesto diferentes aproximaciones para el cálculo del error cuadrático medio. Una de ellas se basa en la aproximación de Prasad and Rao (1990) y las restantes están basadas en técnicas bootstrap. En las simulaciones se observa un mejor comportamiento de las técnicas bootstrap.

En lo referente a los resultados del mercado de trabajo en Galicia se puede concluír que hubo un crecimiento general en las tasas de paro en el periodo considerado en casi todas las comarcas, aunque este incremento fue superior en los hombres. Esto puede estar condicionado, entre otras cosas, por la brusca caída del empleo en el sector de la construcción, que emplea principalmente a hombres. Debido a la situación económica en

España, con caídas del Producto Interior Bruto, este población no ha sido capaz de encontrar trabajo en otro sector de actividad. También se puede observar que las comarcas de la costa oeste de Galicia son aquellas que, en términos generales, tienen mayores tasas de desempleo. Esta área es la más dinámica de Galicia y con la población más joven. En esta área vive aproximadamente el 75% de la población gallega y las tasas de paro son tan altas porque las empresas no pueden absorver a tantos trabajadores.

Los modelos propuestos en esta tesis se han implementado en un paquete de R con el nombre mme. En los últimos años el software R se ha convertido en una poderosa herramienta científica que ofrece una colección rica de técnicas estadísticas modernas y clásicas. Motivado por su flexivilidad y su amplia aceptación en la comunidad científica hemos elegido R como lenguaje de programación para desarrollar la librería de funciones necesarias para el ajuste de los modelos propuestos en este trabajo.

En el paquete mme hemos introducido una serie de nuevas funciones que pueden se de interés para aquellos que están haciendo investigación aplicada. Las ocho funciones principales se resumen en la tabla siguiente:

Funcion	Descripcion	Referencia
`data.mme`	Basada en los datos introducidos esta función genera algunas matrices necesarias para cálculos posteriores y los valores iniciales de los efectos fijos y aleatorios para el algoritmo de ajuste.	López-Vizcaíno et al (2013a)
`fitmodel1`	Función empleada para el ajuste del modelo multinomial mixto con efectos aleatorios independientes para cada una de las categorías de la variable respuesta (Modelo 1)	López-Vizcaíno et al (2013a)
`fitmodel2`	Función empleada para el ajuste del modelo multinomial mixto con do efectos aleatorios independientes por cada categoría de la variable respuesta: un effecto aleatorio debido al dominio y otro efecto aleatorio independiente debido al tiempo y al dominio (Modelo 2)	López-Vizcaíno et al (2013b)
`fitmodel3`	Función empleada para el ajuste del modelo multinomial mixto con do efectos aleatorios independientes por cada categoría de la variable respuesta: un effecto aleatorio debido al dominio y otro efecto aleatorio correlado debido al tiempo y al dominio (Modelo 3)	López-Vizcaíno et al (2013b)
`msef`	Esta función se usa para calcular el MSE analítico para el Modelo 1	López-Vizcaíno et al (2013a)
`msef.it`	Esta función se usa para calcular el MSE analítico para el Modelo 2	López-Vizcaíno et al (2013b)
`msef.ct`	Esta función se usa para calcular el MSE analítico para el Modelo 3	López-Vizcaíno et al (2013b)
`mseb`	Función usada para calcular el sesgo y el MSE para los modelos multinomiales mixtos utilizando bootstrap paramétrico	López-Vizcaíno et al (2013a) and López-Vizcaíno et al (2013b)

Bibliography

Battese, G. E., Harter, R. M., and Fuller, W. A. (1988). An error component model for prediction of county crop areas using survey and satelite data. *Journal of the American Statistical Association*, 83:28–36.

Brackstone, G. J. (1987). Small area data: policy issues and technical challenges. In Platek, R., Rao, J. N. K., Sarndal, C. E., and Singh, M. P., editors, *Small Area Statistics*, pages 3–20. Wiley, New York.

Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24:123–140.

Breslow, N. and Clayton, D. (1996). Aproximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, 88:9–25.

Chandra, H. and Chambers, R. (2009). Multipurpose small area estimation. *Journal of Official Statistics*, 25:379395.

Chaudhuri, A. (1994). Small domain statistics: A review. *Statistica Neerlandica*, 48:215–236.

Council, N. R. (2000). *Small-area estimates of school-age children in poverty: Evaluation of current methodology. C.F. Citro and G. Kalton (Eds.).* Committee on National Statistics.

Cressie, N. A. C. (1993). *Statistics for spatial data.* Wiley, New York.

Cui, Y., Baldwind, S. B., Lightstone, A. S., Shih, M., Yu, H., and Teutsch, S. (2012). Small area estimates reveal high cigarette smoking prevalence in low-income cities of Los Angeles county. *J Urban Health*, 89(3):397–406.

Das, K., Jiang, J., and Rao, J. N. K. (2004). Mean squared error of empirical predictor. *The Annals of Statistics*, 32:818–840.

Datta, G. S. and Lahiri, P. (2000). A unified measure of uncertainty of estimated best linear unbiased predictors in small area estimation problems. *Statistica Sinica*, 10:613–627.

Datta, G. S., Lahiri, P., and Maiti, T. (2002). Empirical bayes estimatimation of median income of four-person families by state using time series and cross-sectional data. *Journal of Statistical Planning and Inference*, 102:83–97.

Datta, G. S., Lahiri, P., Maiti, T., and Lu, K. L. (1999). Hierarchical bayes estimatimation of unemployment rates for U.S. states. *Journal of the American Statistical Association*, 94:1074–1082.

Deville, J. and Sarndal, C. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Society*, 87:376–382.

Dick, P. (1995). Modelling net undercoverage in the 1991 canadian census. *Survey Methodology*, 21:45–54.

Ericksen, E. P. (1974). A regression method for estimatin population changes of local areas. *Journal of the American Statistical Association*, 69:867–875.

Ericksen, E. P. and Kadane, J. B. (1985). Estimating the population in census year: 1980 and beyond (with discussion). *Journal of the American Statistical Association*, 80:98–131.

Esteban, M. D., Morales, D., Pérez, A., and Santamaría, L. (2009). Small area estimation of poverty indicators under area-level time models. In *SAE 2009 Conference on small area estimation*.

Esteban, M. D., Morales, D., Pérez, A., and Santamaría, L. (2012). Small area estimation of poverty proportions under area-level time models. *Computational Statistics and Data Analysis*, DOI: 10.1016/j.csda.2011.10.015.

EUROSTAT (1996). *Sistema Europeo de Cuentas. SEC 1995.* Oficina de Publicaciones Oficiales de las Comunidades Europeas, Luxemburgo.

EUROSTAT (2010). Labour force survey in the EU, candidate and EFTA countries. Main characteristics of the national surveys, 2008. Technical report, Luxemburgo.

Eustat (2008). *Estimación de áreas pequeñas en la encuesta de población en relación con la actividad de la C.A. de Euskadi.* Eustat, Pais Vasco, Spain.

Faraway, J. J. (2006). *Extending the Linear Model with R. Generalized Linear, Mixed Effects and Nonparametric Regression Models.* Chapman & HallCRC, USA.

Fay, R. E. and 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.

Ghosh, M., Nangia, N., and Kim, D. (1996). Estimation of median income of four-person families: a bayesian time series approach. *Journal of the American Statistical Association*, 91:1423–1431.

Ghosh, M. and Rao, J. N. K. (1994). Small area estimation: An appraisal. *Statistical Science*, 9:55–93.

González-Manteiga, W., Lombardía, M. J., Molina, I., Morales, D., and Santamaría, L. (2008a). Analytic and bootstrap approximations of prediction errors under a multivariate Fay-Herriot model. *Computational Statistics and Data Analysis*, 52:5242–5252.

González-Manteiga, W., Lombardía, M. J., Molina, I., Morales, D., and Santamaría, L. (2008b). Estimacion of the mean squared error of predictors of small area linear parameters under a logistic mixed model. *Computational Statistics and Data Analysis*, 51:2720–2733.

Hall, P. and Maiti, T. (2006). On parametric bootstrap methods for small-area prediction. *Journal of the Royal Statistical Society, B*, 68:221–238.

Hansen, M. H., Hurtwiz, W. N., and Madow, W. G. (1953). *Sample Survey Methods and Theory I.* Wiley, NewYork.

Herrador, M., Morales, D., Esteban, M. D., Sánchez, A., Santamaría, L., Marhuenda, Y., Pérez, A., and Molina, I. (2009). Estimadores de áreas pequeñas basados en modelos para la encuesta de población activa. *Estadística Española*, 51(170):133–137.

Hosmer, D. W. and Lemeshow, S. (2000). *Applied Logistic Regression.* Wiley Series in Probability and Statistics, USA.

Hudson, C. G. (2009). Validation of a model for estimating state and local prevalence of serious mental illness. *Int J Methods Psychiatr Res.*, 18(4):251–64.

INE (2009). Encuesta de población activa. Diseño de la encuesta y evaluación de la calidad de los datos. Informe técnico. Technical report, Instituto Nacional de Estadística, Madrid, España.

Jiang, J. and Lahiri, P. (2006). Mixed model prediction and small area estimation. *Test*, 15(1):1–96.

Jiang, J., Lahiri, P., and Wan, S. (2002). A unified jackknife theory for empirical best prediction with m-estimation. *The Annals of Statistics*, 30:1782–1810.

Kackar, R. and Harville, D. A. (1981). Unbiasedness of two-stage estimation and prediction procedures for mixed linear models. *Communications in Statistics. Theory and Methods,Ser. A*, 10:1249–1261.

Kackar, R. and 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*, 79:853–862.

Lahiri, P. and Meza, J. L. (2002). Small-area estimation. In El-Shaarawi, A. H. and Piegorsch, W. W., editors, *Encyclopedia of Environmetric*, pages 2010–2014. John Wiley and Sons, Chichester.

Lian, H. (2012). A note on conditional Akaike information for Poisson regression with random effects. *Electronic Journal of Statistics*, 6:1–9.

López, R. (2000). Estimación para áreas pequeñas(1). *Estadística Española*, 42:291–338.

López-Vizcaíno, M., Lombardía, M., and Morales, D. (2013a). Multinomial-based small area estimation of labour force indicators. *Statistical Modelling*, 13:153–178.

López-Vizcaíno, M., Lombardía, M., and Morales, D. (2013b). Small area estimation of labour force indicator under a multinomial mixed model with correlated time and area effects. *Submited for review.*

Marker, D. A. (1999). Organization of small area estimators using generalized regression framework. *Journal of Official Statistics*, 15:1–24.

McCullath, P. and Nelder, J. (1999). *Generalized Linear Models.* Chapman and Hall.

MCGilchrist (1994). Estimation in generalized mixed models. *Journal of Royal Statistical Society, Series B*, 56:6169.

Molina, I., Saei, A., and Lombardía, M. J. (2007). Small area estimates of labour force participation under multinomial logit mixed model. *The Journal of the Royal Statistical Society, series A*, 170:975–1000.

Morales, D., Esteban, M. D., Sánchez, A., Santamaría, L., Marhuenda, Y., Pérez, A., Saralegui, J., and Herrador, M. (2007). Estimación en áreas pequeñas con datos de la encuesta de población activa en Canarias. *Estadística Española*, 49(165):301–332.

Morrison, P. (1971). Demographic information for cities: a manual for estimating and projecting local population characteristics. *RAND report R-618-HUD.*

Mukhopadhyay, P. (1998). *Small Area Estimation in Survey Sampling.* Narosa Publishing House, Delhi.

ONS (2004). Labour force survey user guide. Technical Report Vol 6, Office for National Statistics, United Kingdon.

Pérez, A. (2009). On the time stability of several small area estimators. In *SAE 2009 Conference on small area estimation.*

Pérez-Infante, J. I. (2006). *Las estadísticas del Mercado de trabajo en España.* Ministerio de Trabajo y Asuntos Sociales.

Pfeffermann, D. (2002). Small area estimation-new developments and directions. *International Statistical Review*, 70:125–143.

Pfeffermann, D. (2012). New important developments in small area estimation. *Statistical Science.*

Pfeffermann, D. and Burck, L. (1990). Robust small area estimation combinig time series an cross-sectional data. *Statistics Canada*, 16(2):217–237.

Pfeffermann, D. and Tiller, R. (2005). Bootstrap approximation to prediction mse for state-space models with estimated parameters. *Journal of Time Series Analysis*, 26:893–916.

Pfeffermann, D. and Tiller, R. (2006). Small-area estimation with state-space models subject to benchmark constraints. *Journal of the American Statistical Association*, 101(476):1387–1397.

Platek, R., Rao, J. N. K., Sarndal, C. E., and Singh, M. P. (1987). *Small Area Statistics*. Wiley, New York.

Prasad, N. G. N. and 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.

Purcell, N. J. and Kish, L. (1979). Estimates for small domain. *Biometrics*, 35:365–384.

Rao, J. N. K. (1986). Synthetic estimators, SPREE and Best Model Based Predictors. In *Proceedings of the Conference on Survey Research Methods in Agriculture*, pages 1–16.

Rao, J. N. K. (1999). Some recent advances in model-based small area estimation. *Survey Methodology*, 25:175–186.

Rao, J. N. K. (2001). Eb and eblup in small area estimation. In Ahmed, S. E. and Meid, N., editors, *Empirical Bayes and Likelihood Inferences, Lecture Notes in Statistics*, pages 33–43. Springer, New York.

Rao, J. N. K. (2003). *Small Area Estimation*. Wiley, New York.

Rao, J. N. K. and Yu, M. (1994). Small-area estimation by combining time-series and cross-sectional data. *The Canadian Journal of Statistics*, 22(4):511–528.

Saei, A. and Chambers, R. (2003). Small area estimation under linear an generalized linear mixed models with time and area effects. Technical report, Southampton, UK.

Salvati, N., Chandra, H., and Chambers, R. (2010). Small area estimation using a non-parametric model-based direct estimator. *Computational Statistics and Data Analysis*, 54:2159–2171.

Santiago-Pérez, M. I., López-Vizcaíno, M. E., Pérez-Ríos, M., Malvar, A., and Hervada, X. (2012). Cómo estimar prevalencias de consumo de tabaco en áreas pequeñas. Galicia: 2010-2011. In *XXX Reunión Científica de la Sociedad Española de Epidemiología: Epidemiología en tiempos de crisis. Haciendo sostenible el sistema de salud.*

Sarndal, C. E., Swensson, B., and Wretman, J. (1992). *Model Assisted Survey Sampling*. Springer, New York.

Scealy, J. (2009). Small area estimation using a multinomial logit mixed model with category specific random effects. In *SAE 2009 Conference on small area estimation.*

Schaible, W. L. (1996). *Indirect Estimation in U.S. Federal Programs*. Springer, NewYork.

Schall, R. (1991). Estimation in generalized linear models with random effects. *Biometrika*, 78:719–727.

Searle, S. R., Casella, G., and McCullogh, C. E. (1992). *Variance components*. John Wiley.

Srebotnjak, T., Mokdad, A. H., and Murray, C. J. (2011). A novel framework for validating and applying standardized small area measurement strategies. *Popul Health Metr.*, 29:8:26.

Tiller, R. B. (1992). Times series modeling of sample survey data from the U.S. Current population survey. *Statistics Sweden*, 8(2):149–166.

Ugarte, D., Militino, A. F., González, B., Goicoa, T., and López, M. (2005). Estimación del desempleo por comarcas en Navarra. In *Sexto Congreso de Economía de Navarra*.

Ugarte, M. D., Goicoa, T., Militino, A. F., and Durbán, M. (2009a). Spline smoothing in small area trend estimation and forecasting. *Computational Statistics and Data Analysis*, 53:3616–3629.

Ugarte, M. D., Goicoa, T., Militino, A. F., and Sagaseta-López, M. (2009b). Estimating unemployment in very small areas. *SORT*, 33:49–70.

Vaida, F. and Blanchard, S. (2005). Conditional akaike information for mixed-effects models. *Biometrika*, 92:351–370.

You, Y., Rao, J. N. K., and Gambino, J. (2001). Model-based unemployment rate estimation for the canadian labour force survey: a hierarchical approach. Technical report, Household Survey Method Division. Statistics Canada, Canada.

Zhang, Z., Zhang, L., Penman, A., and May, W. (2011). Using small-area estimation method to calculate county-level prevalence of obesity in Mississippi, 2007-2009. *Prev Chronic Dis.*, 8(4):A85.

Zidek, J. V. (1982). A review of methods for estimating population of local areas. Technical report, Vancouver, Canada.