Catalogue no. 12-001-X
ISSN 1492-0921

Survey Methodology

Small area estimation for unemployment using latent Markov models

by Gaia Bertarelli, M. Giovanna Ranalli, Francesco Bartolucci, Michele D'Alò and Fabrizio Solari

Release date: December 20, 2018

Statistics Canada Statistique Canada

Canada

How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, www.statcan.gc.ca.

You can also contact us by

Email at STATCAN.infostats-infostats.STATCAN@canada.ca

Telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following numbers:

- Statistical Information Service 1-800-263-1136
- National telecommunications device for the hearing impaired 1-800-363-7629
- Fax line 1-514-283-9350

Depository Services Program

- Inquiries line 1-800-635-7943
- Fax line 1-800-565-7757

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under "Contact us" > "Standards of service to the public."

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Standard table symbols

The following symbols are used in Statistics Canada publications:

. not available for any reference period
.. not available for a specific reference period
... not applicable
0 true zero or a value rounded to zero
0^s value rounded to 0 (zero) where there is a meaningful distinction between true zero and the value that was rounded
^{p} preliminary
^{r} revised
x suppressed to meet the confidentiality requirements of the *Statistics Act*
^{E} use with caution
F too unreliable to be published
* significantly different from reference category ($p < 0.05$)

Published by authority of the Minister responsible for Statistics Canada

An HTML version is also available.

Cette publication est aussi disponible en français.

Survey Methodology, December 2018
Vol. 44, No. 2, pp. 167-192
Statistics Canada, Catalogue No. 12-001-X

Small area estimation for unemployment using latent Markov models

Gaia Bertarelli, M. Giovanna Ranalli, Francesco Bartolucci, Michele D'Alò and Fabrizio Solari[1]

Abstract

In Italy, the Labor Force Survey (LFS) is conducted quarterly by the National Statistical Institute (ISTAT) to produce estimates of the labor force status of the population at different geographical levels. In particular, ISTAT provides LFS estimates of employed and unemployed counts for local Labor Market Areas (LMAs). LMAs are 611 sub-regional clusters of municipalities and are unplanned domains for which direct estimates have overly large sampling errors. This implies the need of Small Area Estimation (SAE) methods. In this paper we develop a new area level SAE method that uses a Latent Markov Model (LMM) as linking model. In LMMs, the characteristic of interest, and its evolution in time, is represented by a latent process that follows a Markov chain, usually of first order. Therefore, areas are allowed to change their latent state across time. The proposed model is applied to quarterly data from the LFS for the period 2004 to 2014 and fitted within a hierarchical Bayesian framework using a data augmentation Gibbs sampler. Estimates are compared with those obtained by the classical Fay-Herriot model, by a time-series area level SAE model, and on the basis of data coming from the 2011 Population Census.

Key Words: Area level model; Hierarchical Bayes; Time-series data; Labor Force Survey; Augmented data.

1 Introduction

In Italy, the Labor Force Survey (LFS) is conducted quarterly by ISTAT, the National Statistical Institute, to produce estimates of the labor force status of the population at a national, regional (NUTS2), and provincial (LAU1) level, with monthly, quarterly, and yearly frequency, respectively. Since 1996, ISTAT also disseminates yearly LFS estimates of employed and unemployed counts for local Labor Market Areas (LMAs). LMAs are sub-regional geographical areas where the bulk of the labor force lives and works, and where establishments can find the largest amount of the labor force necessary to occupy the offered jobs. These are 611 distinct and functional areas defined as clusters of municipalities through an allocation process based on commuting patterns collected by the 2011 Population Census (Istat, 2014). Unlike NUTS2 and LAU1 areas, LMAs are unplanned domains that cut across sampling strata and LAU1 areas. In addition, direct estimators have overly large sampling errors particularly for areas with small sample sizes. This makes it necessary to borrow strength from data on auxiliary variables from other areas through appropriate models, leading to indirect or model-based estimates.

Small Area Estimation (SAE) methods are used in inference for finite populations to obtain estimates of parameters of interest when domain sample sizes are too small to provide adequate precision for direct domain estimators. Statistical models for SAE can be formulated at the individual or area (i.e., aggregate) levels. In this paper we focus on the latter. The Fay-Herriot model (Fay and Herriot, 1979, FH) is the basic area level SAE model: it uses cross-sectional information for predicting small area parameters of interest by combining direct estimates and population level auxiliary information with a linear mixed model. When

1. Gaia Bertarelli, Dept. of Economics and Management, University of Pisa, Italy; M. Giovanna Ranalli, Dept. of Political Science, University of Perugia, Italy. E-mail: giovanna.ranalli@unipg.it; Francesco Bartolucci, Dept. of Economics, University of Perugia, Italy; Michele D'Alò and Fabrizio Solari, Italian National Statistical Institute.

longitudinal data are also available, it is possible to borrow strength over time. Among others, Rao and Yu (1994) propose a model involving autocorrelated random effects and use both time-series and cross-sectional data, while Marhuenda, Molina and Morales (2013) develop a spatio-temporal FH model using an autoregressive model in space together with a first-order autoregressive covariance structure in time.

Several papers deal with SAE using time-series models and the Kalman filter after expressing them in a state-space form. Pfeffermann and Burck (1990) introduce state-space models to estimate the Canadian unemployment rates and Pfeffermann and Rubin-Bleuer (1993) use this approach to model the correlation between the trends of domain series in a multivariate structural time-series model. Pfeffermann and Tiller (2006) add monthly benchmark constraints to the time-series state-space model, while Harvey and Chung (2000) consider a bivariate state-space model to obtain more stable and precise estimates of change in unemployment. Krieg and Van der Brakel (2012) model domain series in a multivariate time-series model and apply the cointegration idea to construct more parsimonious common trend models. Level break estimation within the structural time-series framework is illustrated in Van den Brakel and Krieg (2015). More recently, Van der Brakel and Krieg (2016) and Boonstra and Van den Brakel (2016) apply these models to data from the Dutch LFS.

Proposals for area level time-series data have also been developed following a Hierarchical Bayesian (HB) approach. In particular, Ghosh, Nangia and Kim (1996) apply a fully HB analysis using a time-series model to the estimation of median income of four-person families. Datta, Lahiri, Maiti and Lu (1999) apply this approach to a longer time-series from the U.S. Current Population Survey and use a random walk model for the area random effects. You, Rao and Gambino (2003) apply the same model to unemployment rate estimation for the Canadian LFS. Recently, Boonstra (2014) uses a time-series HB multilevel model to estimate unemployment at the municipality level using data from the Dutch LFS. In particular, estimates are obtained for each quarter and include random municipality effects and random municipality by quarter effects.

In this work we develop a new area level SAE method based on Latent Markov Models (LMMs, see Bartolucci, Farcomeni and Pennoni, 2013, for a thorough description) to estimate unemployment incidences in LMAs using quarterly data from 2004 to 2014 within an HB framework. Area level SAE models consist of two parts, a sampling model formalizing the assumptions on direct estimators and their relationship with underlying area parameters, and a linking model that relates these parameters to area specific auxiliary information. In this work, an LMM is used as linking model and the sampling model is introduced as the highest level of the hierarchy. The resulting model is fitted within a Bayesian framework using a Gibbs sampler with augmented data (corresponding to the latent variables) that allows for a more efficient sampling of the model parameters (Tanner and Wong, 1987).

LMMs, introduced by Wiggins (1973), allow for the analysis of longitudinal data when the response variables measure common characteristics of interest that are not directly observable. The basic LMM formulation is similar to that of hidden Markov models for time-series data (MacDonald and Zucchini, 1997). In these models, the characteristics of interest and their evolution in time are represented by a latent process that follows a Markov chain, typically of first order, so that single areas are allowed to move between latent states across time. LMMs may be seen as an extension of Markov chain models to control for measurement errors. Moreover, LMMs can be seen as an extension of latent class models (Lazarsfeld,

Henry and Anderson, 1968) to longitudinal data. Latent class models have been considered in a SAE framework in Fabrizi, Montanari and Ranalli (2016), where a latent class unit level model for predicting disability small area counts from survey data is introduced for cross sectional data.

The remainder of this paper is organized as follows. Section 2 provides a more detailed description of the available LFS data, while Section 3 introduces notation and reviews some relevant time-series area level SAE methods available in the literature. In Section 4, the model and the procedure for its estimation are presented in detail. Section 5 is devoted to the discussion of the results of the application to the LFS data. Conclusions and possible future developments are outlined in Section 6.

2 Data and preliminary analysis

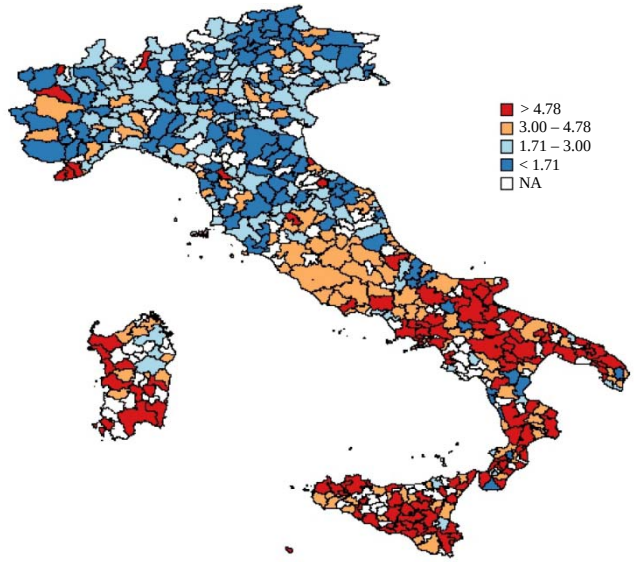
As already mentioned, LMAs are unplanned domains for the LFS. In fact, the sampling design is as follows. Within a given LAU1, municipalities are classified as Self-Representing Areas (larger municipalities) and Non-Self-Representing Areas (smaller municipalities). In Self-Representing Areas, a stratified cluster sampling design is applied: each municipality is a single stratum and households are selected by means of systematic sampling. In Non-Self-Representing Areas, the sample is based on a stratified two stage sampling design: municipalities are Primary Sampling Units, while households are Secondary Sampling Units. Primary Sampling Units are divided into strata of the same dimension in terms of population size. One Primary Sampling Unit is drawn from each stratum without replacement and with probability proportional to the Primary Sampling Unit population size. Secondary Sampling Units are selected by means of systematic sampling in each Primary Sampling Unit. All members of each sample household, both in Self-Representing Areas and in Non-Self-Representing Areas are interviewed. In each quarter, about 70,000 households and 1,350 municipalities are included in the sample. Note that some LMAs (usually the smallest ones) have a very small sample size. Furthermore, usually about one third of the LMAs is not included in the sample at all (i.e., they have a zero sample size).

The LFS follows a rotating panel sampling design, according to a 2-(2)-2 scheme: households are interviewed in two consecutive quarters and, after a two-quarter break, they are interviewed for two additional consecutive quarters. Although the LFS panel design induces correlation among quarterly estimates, due to partial overlap of the sample units, we do not account for it in our model specification in the application illustrated in Section 5. In any case, we expect that this does not affect the comparison among different methods.

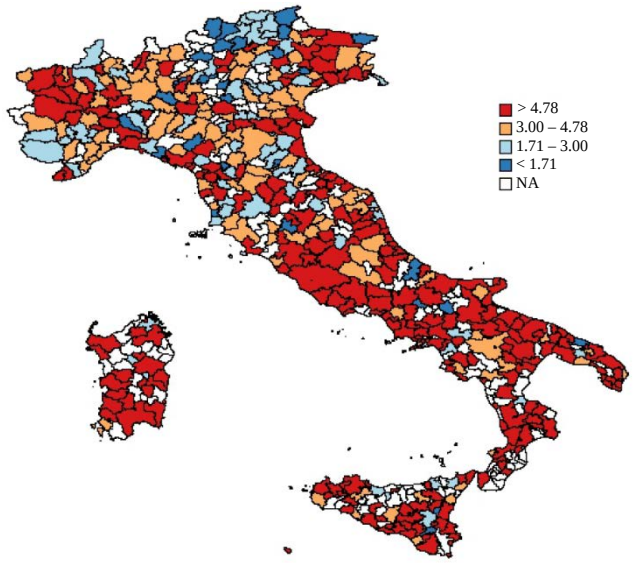
In this work we model quarterly unemployment incidences for 611 LMAs for the period 2004-Q1 to 2014-Q4 (44 quarters). Figure 2.1 shows the map of direct estimates in the first and in the last time occasion of the observation time span. Figure 2.2, on the other hand, shows all the direct estimates for each small area in two NUTS2 areas: Lombardy (left panel) is a rich region in the North of Italy, while Sicily (right panel) is the southern Island and is much less wealthy. We observe, in general, that direct estimates are extremely variable and that unemployment has decreased over the first three years, and then started to increase considerably.

Direct estimates in unplanned domains are characterized by a high Coefficient of Variation (CV), which is used as a measure of uncertainty associated with the estimates. In addition, 6,762 out of 26,884 direct estimates

cannot be computed because the sample dimension is zero. Usually, in Official Statistics, an estimate for a Labor Force parameter with a CV greater than 33.3% is considered too unreliable and is not recommended for release. Estimates with a CV between 16.6% and 33.3% must be released with caveats because their sampling variability is quite high, while estimates with a CV smaller than 16.6% are of sufficient accuracy and have no release restrictions; see Statistics Canada (2016, page 35). In our data, the vast majority of direct estimates has a very large CV and cannot be considered reliable, as it is shown in Figure 2.3.



(a) 2004-Q1



(b) 2014-Q4

Figure 2.1 Direct estimates of unemployment incidences (%) for the first and the last time occasion: first quarter of 2004 (a) and the last quarter of 2014 (b).

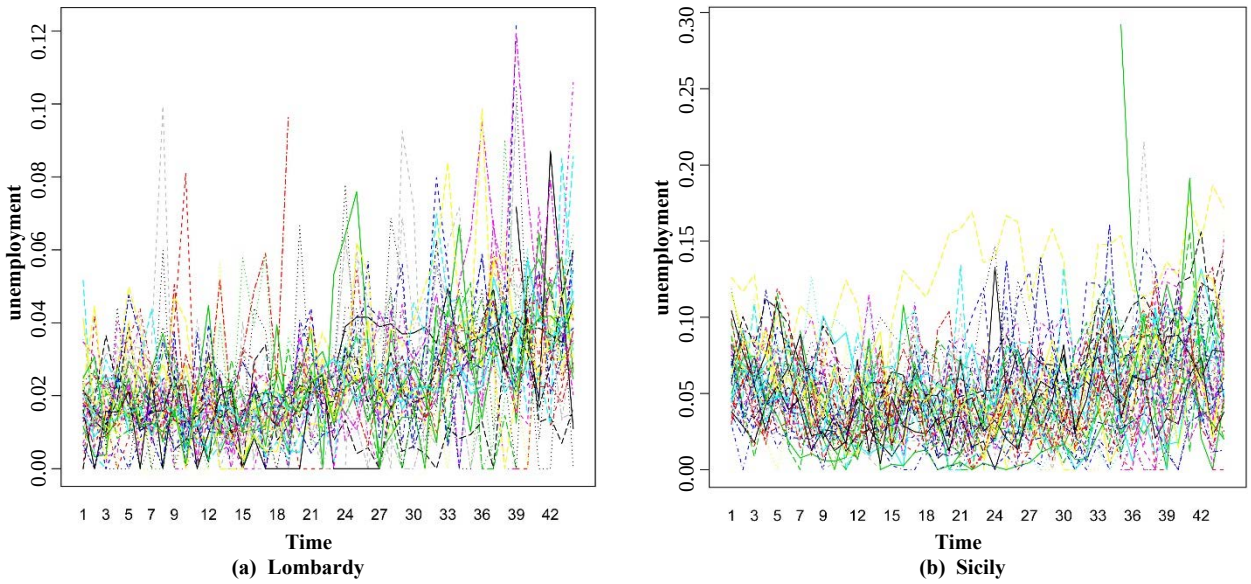


Figure 2.2 Quarterly direct estimates of unemployment incidences in two NUTS2 Regions: Lombardy (a) and Sicily (b), from 2004-Q1 to 2014-Q4.

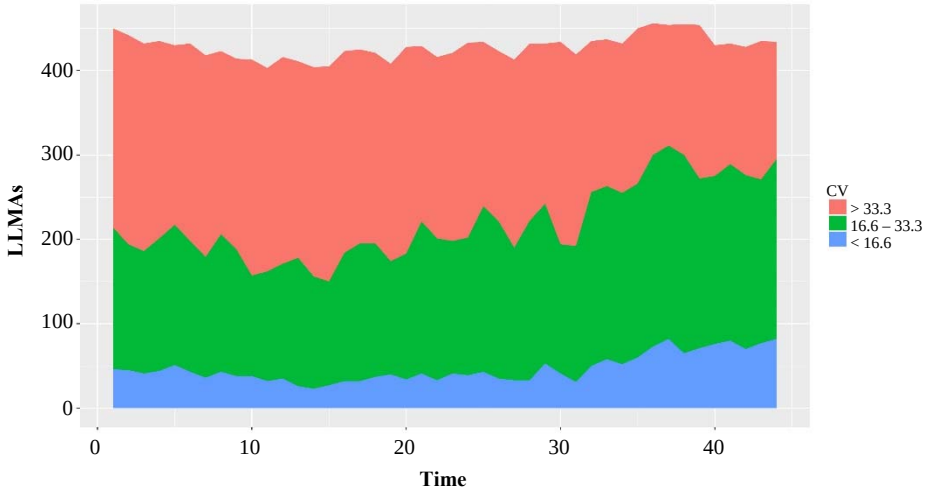


Figure 2.3 For each quarter, distribution of the sampled small areas according to classes of values of the CV of the direct estimates.

The basic idea of SAE is to introduce a statistical model to exploit the relationship between the variable of interest and some covariates for which population information is available. Auxiliary variables available for these data are the population rates in $sex \times 7\ age$ classes (15-19, 20-24, 25-29, 30-39, 40-49, 50-59, 60-74). Since LFS estimates are not seasonally adjusted, we take seasonality into account using year and quarter effects through 10 and 3 dummy variables, respectively.

The CVs of direct estimates are estimates themselves and their precision is a function of the sample size. Therefore, they are subject to a relevant sampling error that can affect small area modeling in different ways (Rao and Yu, 1994) and smoothing estimated Mean Squared Errors (MSEs) is necessary (see Rao, 2003, Chapter 5). In this work, we propose to use a regression model with a logarithmic transformation of the CV and of the MSE (see Wolter, 2007, Chapter 7). In particular, our approach is based on two steps: the first step consists in modeling the CV and then computing the smoothed MSE from this model. At the second step, we model the MSE directly for those small areas for which we do not have a valid CV (i.e., for those LMAs with a zero estimate).

Let $\hat{\theta}_{it}$ be the direct survey estimate for small area $i = 1, \ldots, m,$ with $m = 611,$ at time $t = 1, \ldots, T,$ with $T = 44$. Let CV_{it} denote the corresponding estimate of the CV. Note that Italy is divided into four geographic areas, namely broad-areas (e.g., North-West, North-East, Center, South and Islands), and that each LMA belongs to only one of these broad-areas. In order to smooth estimates of MSEs, we have the following auxiliary information:

- M_{it} is the population size at time t of the broad-area to which LMA i belongs;
- N_{it} is the population size of LMA i at time t;
- r_{it} is a 14-dimensional column-vector that contains population rates in $sex \times 7age$ classes, for LMA i at time t.

At the first step of the proposed procedure, we fit the following regression model for each broad-area:

$$\log(\mathrm{CV}_{it}) = \beta_0 + \log\left(\hat{\theta}_{it}\right)\beta_1 + \log\left(\frac{N_{it}}{M_{it}}\right)\beta_2 + \log(\mathbf{r}_{it})'\,\boldsymbol{\beta}_3 + \log(\mathbf{1}_{14} - \mathbf{r}_{it})'\,\boldsymbol{\beta}_4, \qquad (2.1)$$

where $\mathbf{1}_{14}$ is a 14-dimensional column vector of ones. The use of the log-transformation and the choice of the covariates has been assessed using standard model selection techniques, such as AIC and adjusted R^2. Using predictions denoted by $\widehat{\mathrm{CV}}_{it}$ from this model, smoothed MSEs are obtained as

$$\widehat{\mathrm{MSE}}_{it} = \widehat{\mathrm{CV}}_{it} \times \hat{\theta}_{it}.$$

In the second step of the proposed procedure, for all $\hat{\theta}_{it} = 0,$ CVs cannot be computed while MSEs are available since direct estimates are based on calibrated weights and MSE estimates are based on the residuals of a generalized regression that accounts for the auxiliary variables used in the calibration constraints. Then, MSEs are modeled directly and separately for each broad-area using the following model:

$$\log(\mathrm{MSE}_{it}) = \beta_0 + \log\left(\frac{N_{it}}{M_{it}}\right)\beta_1 + \log(\mathbf{r}_{it})'\,\boldsymbol{\beta}_2 + \log(\mathbf{1}_{14} - \mathbf{r}_{it})'\,\boldsymbol{\beta}_3.$$

Smoothed MSEs are obtained as predictions from this model. Note that, we have resorted to this two-step procedure because the former model, the one for CVs, fitted better than the latter for MSEs in our application. Figure 2.4 reports the final output of this two-step procedure and displays the original and the smoothed MSEs versus unemployment incidence for all sampled areas.

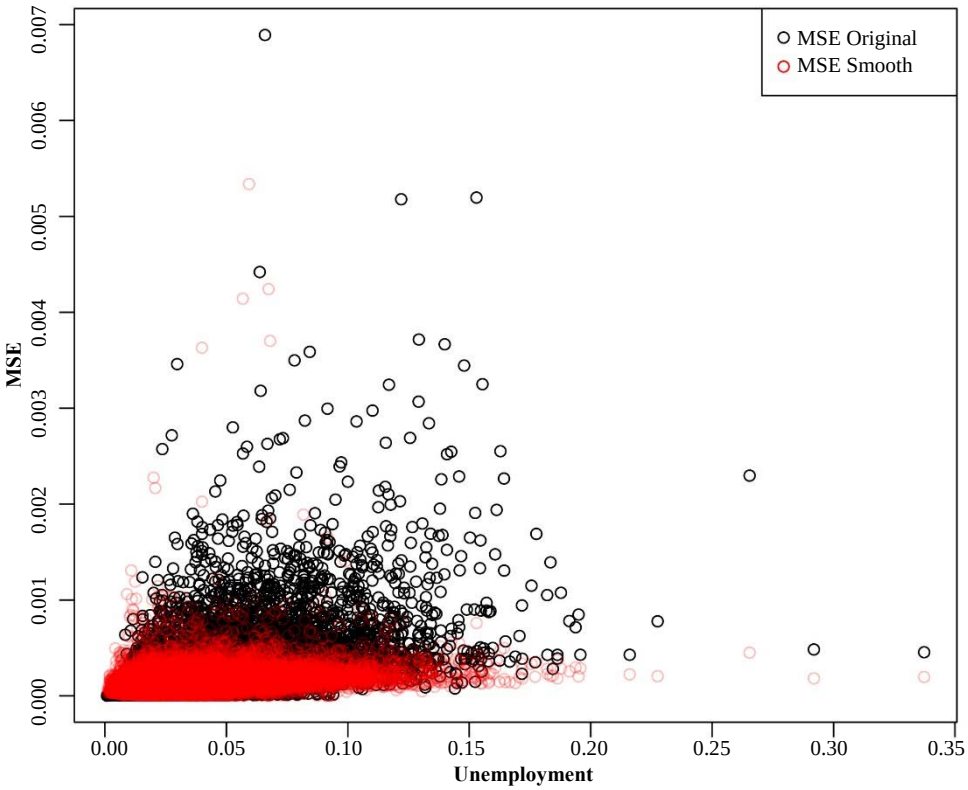


Figure 2.4 Original (black) and smoothed (red) MSEs vs unemployment incidence for all sampled areas.

3 Time series area level SAE models

Rao and Yu (1994) propose an area level model involving autocorrelated random effects and sampling errors using both time-series and cross sectional data. It consists of a sampling model

$$\hat{\theta}_{it} = \theta_{it} + e_{it}, \quad i = 1, \ldots, m, \quad t = 1, \ldots, T,$$

and an area-linking model

$$\theta_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + v_i + u_{it}, \quad i = 1, \ldots, m, \quad t = 1, \ldots, T,$$

where θ_{it} is the true value corresponding to the estimate $\hat{\theta}_{it}$ for the small area mean, $\mathbf{x}_{it}$ is a $p-$ dimensional column vector of fixed covariates, and e_{it} are normal sampling errors. Given the true value θ_{it}, each vector $\mathbf{e}_i = (e_{i1}, \ldots, e_{iT})'$ has multivariate normal distribution with zero mean and with known variance-covariance matrix $\boldsymbol{\Psi}_i$. Moreover, $v_i \sim N(0, \sigma_v^2)$ is the area effect and $u_{it} = \rho u_{i,t-1} + \epsilon_{it}$, with $|\rho| < 1$ and $\epsilon_{it} \sim N(0, \sigma_\epsilon^2)$ is the area-by-time effect. In this model, $\mathbf{e}_i$, v_i, and ϵ_{it} are assumed independent of each other. In our application $\boldsymbol{\Psi}_i$ is diagonal, with elements ψ_{it}, for $t = 1, \ldots, T$.

In the previous formulation, the area-linking model is basically a linear model with mixed coefficients. You et al. (2003, YRG) translate this model into an HB framework as follows. Let $\boldsymbol{\theta}_i = (\theta_{i1}, \ldots, \theta_{iT})'$ and $\hat{\boldsymbol{\theta}}_i = \left(\hat{\theta}_{i1}, \ldots, \hat{\theta}_{iT}\right)'$, then

$$\begin{aligned} \hat{\boldsymbol{\theta}}_i \,|\, \boldsymbol{\theta}_i &\sim N_T\left(\boldsymbol{\theta}_i, \boldsymbol{\Psi}_i\right), \\ \theta_{it} \,|\, \boldsymbol{\beta}, u_{it}, \sigma_v^2 &\sim N\left(\mathbf{x}_{it}'\boldsymbol{\beta} + u_{it}, \sigma_v^2\right), \\ u_{it} \,|\, u_{i,t-1}, \sigma_\epsilon^2 &\sim N\left(\rho u_{i,t-1}, \sigma_\epsilon^2\right), \end{aligned} \tag{3.1}$$

where $\boldsymbol{\beta}$, σ_v^2, and σ_ϵ^2 are mutually independent. The model is fully specified once priors are chosen for $\boldsymbol{\beta}$, σ_v^2, and σ_ϵ^2, namely as $f(\boldsymbol{\beta}) \propto 1$, $\sigma_v^2 \sim \text{IG}(a_1, b_1)$, and $\sigma_\epsilon^2 \sim \text{IG}(a_2, b_2)$, where a_1, a_2, b_1 and b_2 are known positive hyperparameters and, usually, set to be small and to reflect a vague knowledge about σ_v^2 and σ_ϵ^2.

Datta et al. (1999) follow this approach, but introduce a richer structure for the fixed part of the linking model by assuming

$$\theta_{it} = \mathbf{x}_{it}'\boldsymbol{\beta}_i + v_i + u_{it}, \tag{3.2}$$

where v_i and $\boldsymbol{\beta}_i$ are area-specific intercepts and regression coefficients, respectively, and u_{it} is an area-specific error term that follows the random-walk model

$$u_{it} \,|\, u_{i,t-1}, \sigma_\epsilon^2 \sim N\left(u_{i,t-1}, \sigma_\epsilon^2\right).$$

The column vector of auxiliary variables $\mathbf{x}_{it}$ may also include dummy variables for year and/or seasonality adjustments. Note that area-specific regression coefficients considerably increase the estimation complexity and the computational burden. For this reason, the hyperparameters are assumed to be m independent realizations from a common probability distribution specified by $v_i \sim N(0, \sigma_v^2)$ and $\boldsymbol{\beta}_i \sim N\left(\boldsymbol{\beta}, \mathbf{W}_\beta^{-1}\right)$, which, in turn, depend on appropriate parameters. See Datta et al. (1999) for further details.

4 The proposed model

In this section, the proposed SAE model based on LMMs is illustrated. It can be considered as a compromise between the YRG model based on (3.1), which leads to possible oversmoothing, and the computationally demanding alternative proposed in Datta et al. (1999), based on (3.2). We first outline a general description on LMMs and then move to the specification of the area level model and to its estimation.

4.1 Preliminaries

In LMMs, the existence of two types of process is assumed: an unobservable finite-state first-order Markov chain U_{it} with state space $\{1, \ldots, k\}$ and an observed process, which in our case corresponds to θ_{it}, with $i = 1, \ldots, m$ and $t = 1, \ldots, T$. It is assumed that the distribution of θ_{it} depends only on U_{it}; specifically, the θ_{it} are conditionally independent given the U_{it}. In addition, the latent state to which a small area belongs at a certain time point only depends on the latent state at the previous occasion.

The state-dependent distribution, namely the distribution of θ_{it} given U_{it}, can be a continuous or discrete. Such a distribution is typically taken from the exponential family. Thus, the overall vector of parameters of LMM, denoted by $\boldsymbol{\phi}$, includes parameters of the Markov chain, denoted by $\boldsymbol{\phi}_{\text{lat}}$, and the

vector of parameters $\boldsymbol{\phi}_{\text{obs}}$ of the state-dependent distribution. In fact, the model consists of two components, the measurement model and the latent model, which concern the conditional distribution of the response variables given the latent variables and the distribution of the latent variables, respectively. By jointly considering these components, the so-called manifest distribution is obtained: it is the marginal distribution of the response variables, once the latent variables have been integrated out.

The measurement model, based on parameters $\boldsymbol{\phi}_{\text{obs}}$, can be written as

$$\theta_{it} \,|\, U_{it} = u \sim p\left(\theta_{it} \,|\, u, \boldsymbol{\phi}_{\text{obs}}\right).$$

Moreover, the parameters $\boldsymbol{\phi}_{\text{lat}}$ of the Markov chain are:

- the vector of initial probabilities $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_k)'$ where

$$\pi_u = P\left(U_{i1} = u\right), \quad u = 1, \ldots, k;$$

- the transition probability matrix

$$\boldsymbol{\Pi} = \begin{pmatrix} \pi_{1|1} & \ldots & \pi_{1|k} \\ \vdots & \ddots & \vdots \\ \pi_{k|1} & \ldots & \pi_{k|k} \end{pmatrix},$$

  where

$$\pi_{u|\bar{u}} = P\left(U_{it} = u \,|\, U_{i,t-1} = \bar{u}\right), \quad \bar{u}, u = 1, \ldots, k,$$

  is the probability that area i visits state u at time t given that at time $t-1$ it was in state $\bar{u}$.

In this work we consider homogeneous LMMs, namely LMMs where, in agreement with the previous definition, the transition probability matrix is constant in time. Generalizations to non-homogeneous hidden Markov chains and time-varying transition probabilities could also be considered (Bartolucci and Farcomeni, 2009). Individual covariates could be included in the measurement or in the latent model. When the covariates are included in the measurement model (Bartolucci and Farcomeni, 2009), they affect the response variables directly and the latent process is conceived as a way to account for the unobserved heterogeneity between areas. Differently, when the covariates are in the latent model (Vermunt and Magidson, 2002; Bartolucci, Pennoni and Francis, 2007) they influence initial and transition probabilities of the latent process. In a SAE context, we will consider the former approach, so that auxiliary information can be used to improve predictions. Bayesian inference approaches to LMMs are already available in the literature (e.g., in Marin, Mengersen and Robert, 2005; Spezia, 2010). In the following section we illustrate how to incorporate an LMM into an area level SAE model.

4.2 Proposed approach to area level SAE

The proposed model is based on two levels in an HB framework: at the first level, a sampling error model is assumed, then an LMM is used as linking model. The latter is based on two equations, corresponding to the measurement model and to the latent component. In particular, we adopt the following structure:

- Sampling Model:

$$\hat{\boldsymbol{\theta}}_i \,|\, \boldsymbol{\theta}_i \sim N_T\left(\boldsymbol{\theta}_i, \boldsymbol{\Psi}_i\right), \qquad i = 1, \ldots, m;$$

- Linking Model:
  - Measurement Model:

$$\theta_{it} \,|\, U_{it} = u, \mathbf{x}_{it} \sim N\left(\mathbf{x}_{it}'\boldsymbol{\beta}_u, \sigma_u^2\right) \qquad i = 1, \ldots, m;\ t = 1, \ldots, T;$$

  - Latent Model, based on the initial probabilities π_u, $u = 1, \ldots, k$, and on the transition probabilities $\pi_{u|\bar{u}}$, $t = 2, \ldots, T$, $\bar{u}$, $u = 1, \ldots, k$, already defined.

Here $\boldsymbol{\beta}_u$ is the $p \times 1$ vector of the regression coefficients for the latent state to which area i at time t belongs, σ_u^2 is the corresponding error variance, and $\boldsymbol{\Psi}_i$ is the matrix of sampling variances, which is assumed to be known.

It must be noticed that, while in the classical area level SAE models heterogeneity is modeled using continuous (usually Normally distributed) random variables, here it is modeled with a discrete dynamic variable. As we can deduce from Figure 4.1, our data have a skewed distribution. However, the empirical distribution is not far from a Normal distribution. D'Alò, Di Consiglio, Falorsi, Ranalli and Solari (2012) show that the differences in estimates between adopting a Normal or a Binomial model are not as relevant as expected and Normal models are often used for estimation of unemployment rates (You et al., 2003; Boonstra, 2014). Finally adopting the Normal distribution has computational advantages which are clarified later in this section.

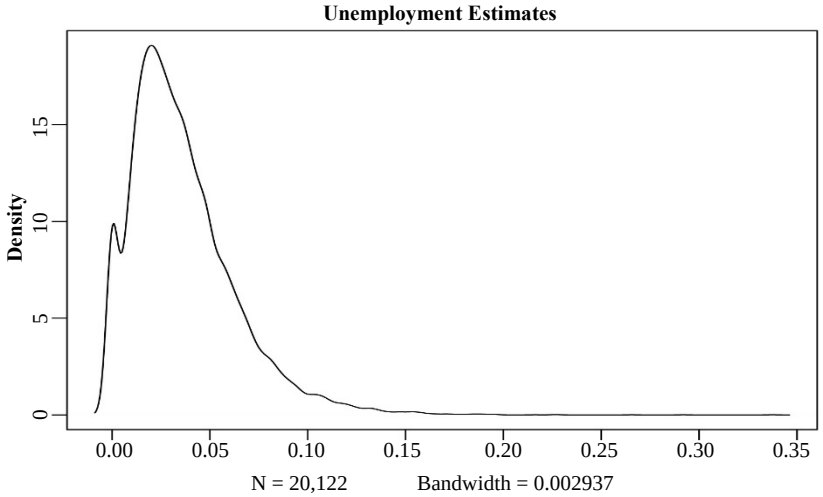


Figure 4.1 Density kernel plot of the direct estimates of unemployment incidences.

The model parameters of interest can be divided into three groups:

- the matrix of small area parameters:

$$\boldsymbol{\Theta} = \begin{pmatrix} \theta_{11} & \cdots & \theta_{1T} \\ \vdots & \ddots & \vdots \\ \theta_{m1} & \cdots & \theta_{mT} \end{pmatrix}; \tag{4.1}$$

- the vector of the measurement parameters:

$$\boldsymbol{\phi}_{\text{obs}} = \left(\boldsymbol{\beta}_1', \ldots, \boldsymbol{\beta}_k', \sigma_1^2, \ldots, \sigma_k^2\right)';$$

- the set of latent parameters:

$$\boldsymbol{\phi}_{\text{lat}} = \{\boldsymbol{\pi}, \boldsymbol{\Pi}\}.$$

To complete the Bayesian formulation of the proposed model, it is necessary to choose priors for the model parameters. Small area parameters do not need a specific prior because direct estimates based on observed data are available; therefore, a set of priors is chosen for the measurement and the latent parameters. Regarding $\boldsymbol{\phi}_{\text{obs}}$, diffuse normal priors are assumed for the regression coefficients. These priors are conjugate and computationally more convenient than the usually flat priors over the real line (see Rao, 2003, Chapter 10). In particular, we assume

$$\boldsymbol{\beta}_u \sim N_p\left(\boldsymbol{\eta}_0, \boldsymbol{\Sigma}_0\right), \quad u = 1, \ldots, k,$$

with $\boldsymbol{\Sigma}_0 = \sigma_u^2 \boldsymbol{\Lambda}_0^{-1}$ and $\boldsymbol{\Lambda}_0$ is a known diagonal matrix.

Variances σ_u^2, $u = 1, \ldots, k$, are unknown and, therefore, it is necessary to set a prior also on these parameters. The choice of the prior distribution for the variance components is critical as in Bayesian mixed models the posterior distributions of these parameters are known to be sensitive to this specification. The inverse Gamma distribution is a popular choice, see e.g., You et al. (2003) and Datta, Lahiri, Maiti and Lu (1999) among others. Gelman (2006), Gelman, Jakulin, Pittau and Su (2008), and Polson and Scott (2012) propose to assume a half-Cauchy distribution for the variance of the random effect. Alternatively, a Uniform distribution can also be considered. Fabrizi et al. (2016) conduct an exhaustive sensitivity analysis when using a latent class model in a multivariate setting and find no significant difference among these different alternatives. For this reason, we choose the same prior distribution considered in You et al. (2003) and use an inverse Gamma distribution with shape parameter a_0 and scale parameter b_0; then $\sigma_u^2 \sim \text{IG}\left(a_0, b_0\right)$, $u = 1, \ldots, k$, where $a_0, b_0 > 0$ are set to very small values. This choice makes it also easier to derive the full conditional distributions for the Gibbs sampler.

For $\boldsymbol{\phi}_{\text{lat}}$, a system of Dirichlet priors is set on the initial probabilities and on the transition probabilities. The Dirichlet distribution is a conjugate prior for the multinomial distribution. This means that if the prior distribution of the multinomial parameters is Dirichlet then the posterior distribution belongs to the same family. The benefit of this choice is that the posterior distribution is easy to compute and, in some sense, it is possible to quantify how much our beliefs have changed after collecting the data. Then, we assume

$$\boldsymbol{\pi} \sim \text{Dirichlet}(\mathbf{1}_k),$$

$$\boldsymbol{\pi}_{\bar{u}} = (\pi_{1|\bar{u}}, \ldots, \pi_{k|\bar{u}})' \sim \text{Dirichlet}(\mathbf{1}_k), \quad \bar{u} = 1, \ldots, k.$$

4.3 Estimation and model selection

In this work we make use of a data augmentation Markov Chain Monte Carlo (MCMC) method (Tanner and Wong, 1987; Liu, Wong and Kong, 1994; Van Dyk and Meng, 2001) based on the Gibbs sampler, in which the latent variables are treated as missing data (Marin et al., 2005; Germain, 2010). There are two main reasons for this choice. First of all, there is evidence that data augmentation has a better performance than other methods, as the marginal updating scheme (Boys and Henderson, 2003). Moreover, it simplifies the process of sampling from the posterior distribution. Details on this method and the full conditionals employed in the Gibbs sampler are given in Appendix A.1.

The choice of the number of latent states is a crucial step in applications. In the framework of LMMs, this requires a model selection procedure. From a Bayesian perspective, a fundamental goal is the computation of the marginal likelihood of the data for a given model. In this paper we use a model selection method based on the marginal likelihood and to estimate this quantity we use the method proposed by Carlin and Chib (1995), applied for each available model on the basis of the output of the MCMC algorithm. Technical details are provided in Appendix A.2.

A well-known problem occurring in Bayesian latent class and LMMs is the label switching. This implies that the component parameters are not identifiable as they are exchangeable. In a Bayesian context, if the prior distribution does not distinguish the component parameters between each other, then the resulting posterior distribution will be invariant with respect to permutations of the labels. Several solutions have been proposed; for a general review see Jasra, Holmes and Stephens (2005). The easiest approach is to use relabeling techniques retrospectively, by post-processing the MCMC output (Marin et al., 2005). However, in our case, we are interested in the prediction of the small area parameters, whose distribution depends on the number of areas in each latent state. Therefore, we do not use the post-processing approach and the MCMC output is permuted at every iteration according to the ordering of the mean of the response variables in each class.

5 Results

In this section we report the results of the application of the LMM area level SAE model to the LFS data presented in Section 2. We fit the model with $k = 2, \ldots, 6$ latent states. For each value of k, we run one Markov chain with 100,000 iterations and then we consider a burn-in period of 50,000 iterations. The posterior means are approximated by means of the retained MCMC samples. Similarly, the variance of the samples approximates the posterior variance of θ_{it}. We select $k = 4$ using the proposed model selection approach. In fact, using expression (A.4), we obtain the following values for the posterior density of the data: $p(\hat{\boldsymbol{\Theta}} \mid k = 2) = 59{,}152.41$, $p(\hat{\boldsymbol{\Theta}} \mid k = 3) = 64{,}405.11$, $p(\hat{\boldsymbol{\Theta}} \mid k = 4) = 68{,}816.06$, and $p(\hat{\boldsymbol{\Theta}} \mid k = 5) = 68{,}703.75$.

We validate our model selection procedure by comparing the final choice with that obtained using the Deviance Information Criterion (DIC). In particular, we focus on $k = 4, 5$ latent states for which the Bayes rule provides the largest values. The DIC confirms our results because we obtain 8,334.0 and 8,362.4 for $k = 4$ and $k = 5$, respectively.

Figure 5.1 compares the map of estimates for the first and the last quarter of the whole period. These can be compared with the maps of direct estimates reported in Figure 2.1. In particular, estimates on the first row of Figure 5.1 are obtained by the proposed LMM area level model. Those on the second row are obtained using a cross-sectional Fay-Herriot (FH) model computed with the R package `hbsae` (Boonstra, 2012), while those on the last row are obtained using the You et al. (2003, YRG) model, for which we have considered three possible choices for ρ, 0.50, 0.75, and 1.00, as in You et al. (2003). To measure the overall fit of the three alternative YRG models we have compared posterior predictive $p-$values (Meng, 1994). In particular, simulated values of a suitable discrepancy measure are generated from the posterior predictive distribution and, then, compared to the corresponding measure for the observed data. More specifically, if $d\left(\hat{\boldsymbol{\Theta}}, \boldsymbol{\Theta}\right)$ is a discrepancy measure that depends on the observed data, $\hat{\boldsymbol{\Theta}}$, and the parameter matrix $\boldsymbol{\Theta}$, then the posterior predictive $p-$value is defined as $P\left[d\left(\hat{\boldsymbol{\Theta}}^*, \boldsymbol{\Theta}\right) > d\left(\hat{\boldsymbol{\Theta}}, \boldsymbol{\Theta}\right) \middle| \hat{\boldsymbol{\Theta}}\right]$, where $\hat{\boldsymbol{\Theta}}^*$ is a sample from the posterior predictive distribution. If a model fits the observed data well, then the two values of the discrepancy measure are similar and, as a result, the value of the $p-$value is expected to be close to 0.5. On the other hand, $p-$values near 0 or 1 signal a model that is not well suited to the data. As in Datta et al. (1999) and in You et al. (2003), we use the following discrepancy measure

$$d\left(\hat{\boldsymbol{\Theta}}, \boldsymbol{\Theta}\right) = \sum_{i=1}^{m} \left(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i\right)' \boldsymbol{\Psi}_i^{-1} \left(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i\right)$$

for the overall fit. The posterior predictive measure suggests that the model with $\rho = 1$ provides a better fit to the data, in fact it takes value 0.188 for $\rho = 1.00$, 0.103 for $\rho = 0.75$, and 0.032 for $\rho = 0.50$. Note that for our model, we obtain a $p-$value equal to 0.311. We have also implemented the Datta et al. (1999) estimation approach. However, the number of areas and the overall number of observations made the estimation computationally prohibitive. For this reason, it is not considered further.

From Figure 5.1, we observe that all model-based estimates are smoother than the original direct estimates. Maps are color-coded according to the quartiles of the direct estimates for 2004-Q1. In general, estimates for 2004-Q1 show a quite distinct division between North, Center, and South of Italy, with relatively higher unemployment incidences in the South of the country. For 2014-Q4, unemployment incidences are all much higher all over the country, because of the economic crisis that hit the country in 2008. LMM and FH show similar patterns, and are in line with those of the direct estimator. YRG, on the other hand, provides more shrunk estimates and this is particularly evident for 2014-Q4 where a general and distinct underestimation is provided. This behavior is displayed for all time points. In fact, Figure 5.2 shows the absolute difference between the direct estimates and model-based estimates. Areas are ordered according to estimated variance of the direct estimates. All model-based estimators show a common general behavior: smaller differences for more precise estimates and increasingly larger differences for more variable direct

estimates. However, we can note that YRG provides systematically larger positive differences, by this casting some concerns on bias.

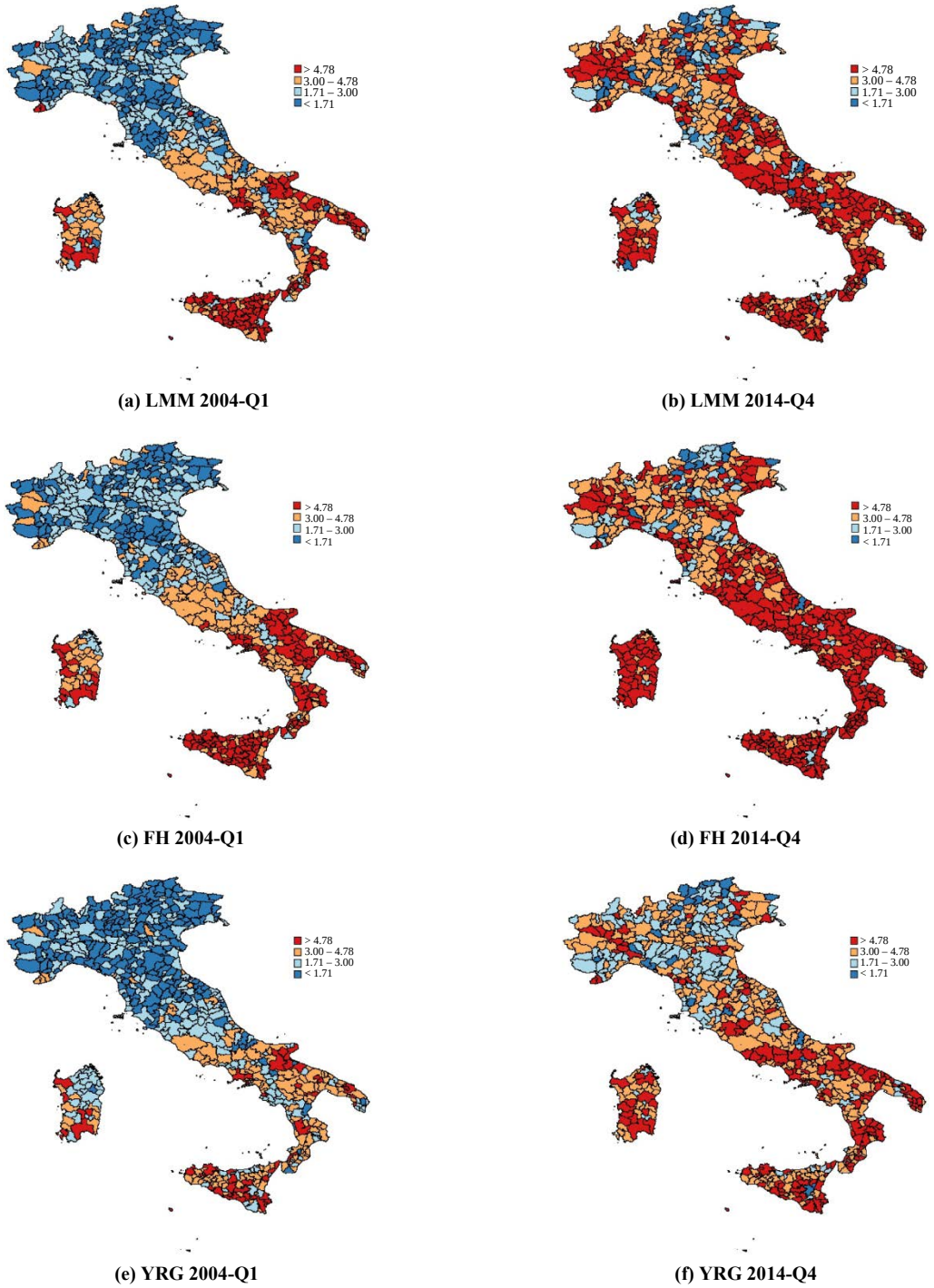


Figure 5.1 Unemployment incidences (%) estimated using LMM, FH and YRG for 2004-Q1 and 2014-Q4.

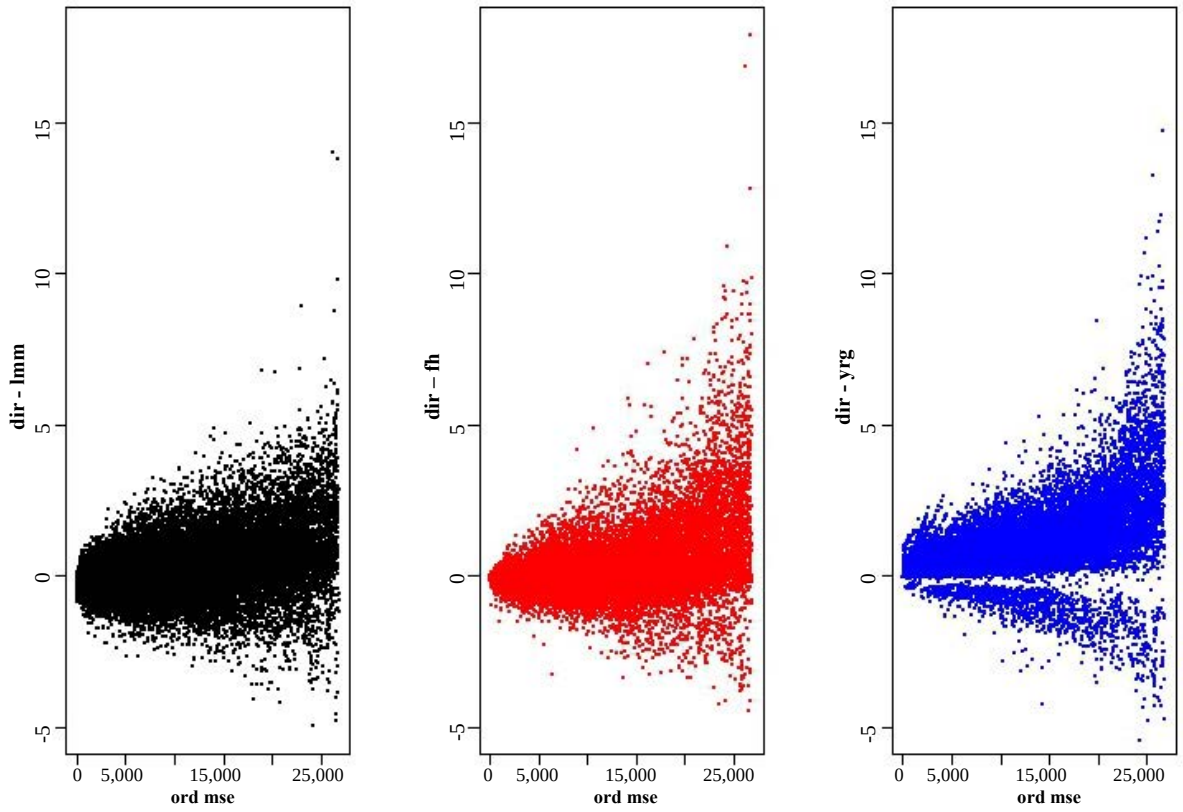


Figure 5.2 Difference between DIR and model-based small area estimates; LMM, FH, YRG, from left to right. Areas are arranged according to increasing estimated variance of the direct estimator.

As mentioned earlier, LMM uses a discrete random variable to model unobserved heterogeneity rather than the more common continuous (usually Gaussian) assumption. As a consequence, small areas can be clustered according to the latent state to which they belong at each time point. In this application, latent states are ordered and can be associated to the level of unemployment, conditionally on the covariates. Figure 5.3 shows the evolution of the latent states clustering for the small areas over the 44 time points. The fourth cluster is very small and comprises areas with a very high unemployment incidence. In addition, the pattern seems to be very stable over time, as the probability of changing latent state is very low. Note that, although there is a noticeable temporal trend in the data, this is captured by the dummy variables inserted to account for trend and seasonality. These finding are supported by the estimated initial and transition probabilities:

$$\hat{\boldsymbol{\pi}} = (0.505,\ 0.340,\ 0.144,\ 0.011)',$$

$$\hat{\boldsymbol{\Pi}} = \begin{pmatrix} 0.967 & 0.027 & 0.004 & 0.002 \\ 0.020 & 0.956 & 0.020 & 0.004 \\ 0.007 & 0.035 & 0.946 & 0.012 \\ 0.035 & 0.007 & 0.030 & 0.929 \end{pmatrix}.$$

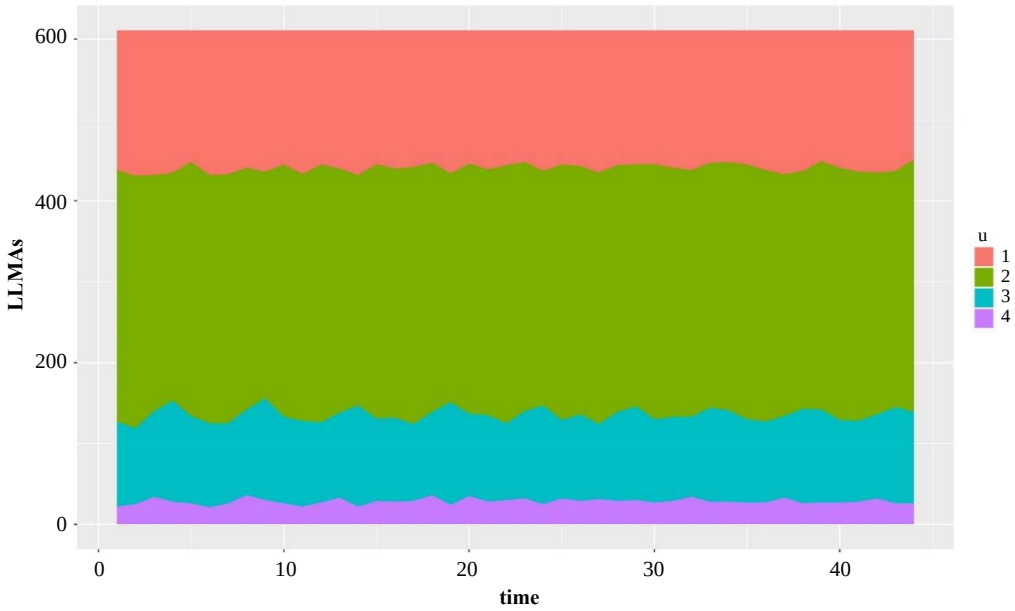


Figure 5.3 Latent states distribution from 2004-Q1 to 2014-Q4.

Figure 5.4 shows the time series of direct estimates and the corresponding model-based estimates for a selection of small areas. Aosta – panel (a) – is a small LMA in the very North of the country, with a small level of unemployment. LMM smooths the direct estimates more than the other methods, while YRG tracks the path of the direct estimates, but provides a noticeable negative bias. Milan – panel (b) – is a large city in the North of the country and the corresponding LMA has usually a very large sample size. As expected, FH and LMM track the values of DIR, while YRG exhibits a clear tendency to underestimation. Perugia and Brindisi are two mid-size towns in the Centre and in the South of Italy, respectively. The pattern of the model-based estimators is very clear: LMM provides a very good smoothing of the quite erratic trend of the direct estimates, better than FH, while YRG again displays a tendency to negative bias, particularly after the first few quarters.

It is expected that model-based estimates, besides providing estimates for the out-of-sample areas, provide gains in efficiency over direct estimates. In Figure 5.5 we report the distribution of the CV for comparing model-based small areas estimates for each time point, classified as in Figure 2.3 according to different relevant values of CV. FH provides estimates for out of sample areas, but it does not seem to provide a useful estimation option for these data since only few estimates have CV smaller than 16%. On the other hand, YRG provides a very good improvement in terms of estimated efficiency, with almost all estimates with a CV smaller than 33.3%. LMM provides a good improvement over FH with only approximately 15% of the small area estimates with a CV larger than 33.3%.

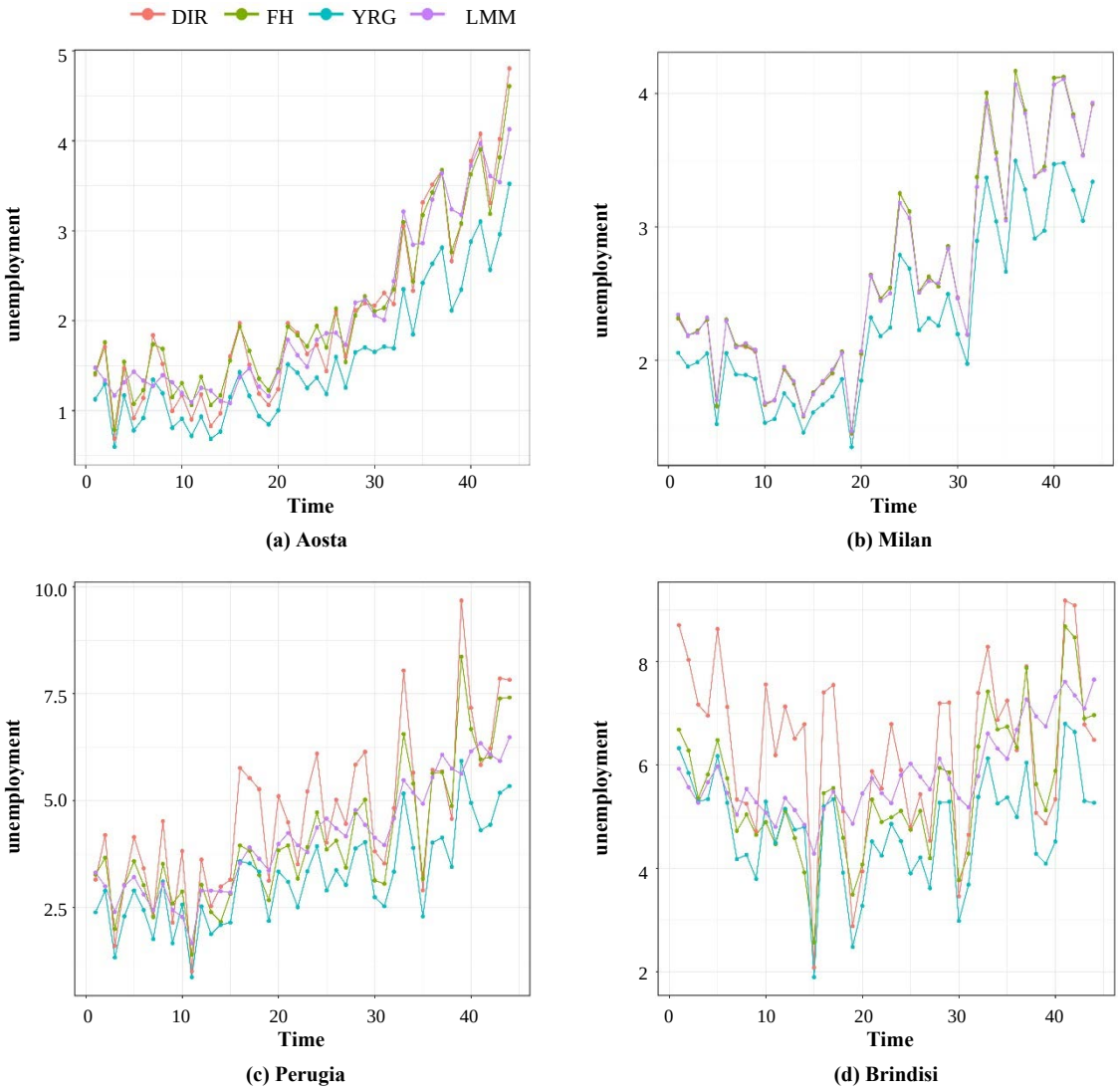


Figure 5.4 Time series of direct and model-based estimates for a selection of four small areas.

In addition, small area estimates should be close to population level quantities, when available. Here, we use data from the 2011 Italian Population Census and consider unemployment incidence for LMAs from the Census as a gold standard. In particular, we evaluate the distance between small area estimates for the closest time point, namely 2011-Q4, and the Census value, Cens_i, and compute the Absolute Relative Error for each area (ARE_i) as

$$\text{ARE}_i = \frac{\left|\hat{\theta}_i - \text{Cens}_i\right|}{\text{Cens}_i} \tag{5.1}$$

for each area i. The ARE_i also provides a sort of measure of relative bias and is important to evaluate and compare the performance in terms of overall error of the estimates. Note that the small area parameter of interest and the Census quantity do not have exactly the same definition. In fact, the LFS is a continuous survey and the corresponding unemployment incidence refers to a quarter, while that from the Census refers to a specific calendar day. In addition, order and wording of items in the two questionnaires used to evaluate

the unemployment status differ slightly. We compare the distribution of ARE_i for LMM and YRG in Figure 5.6. From the empirical distribution of ARE_i, we observe that LMM systematically provides smaller values than YRG. When looking at the subgroup of in-sample areas, we can compare this distribution with that of the direct estimator, and we conclude that LMM is in line with DIR for almost one half of the small areas, and then LMM provides estimates with a relatively smaller value of ARE_i. In conclusion, YRG estimates have a lower estimated variance, but exhibit higher estimated bias, in terms of the comparison with the Census and the direct estimates. This puts concern on coverage. On the other hand, LMM estimates are not as good as YRG estimates in terms of CV, but when looking at the bias, the overall behavior seems to be much more reliable.

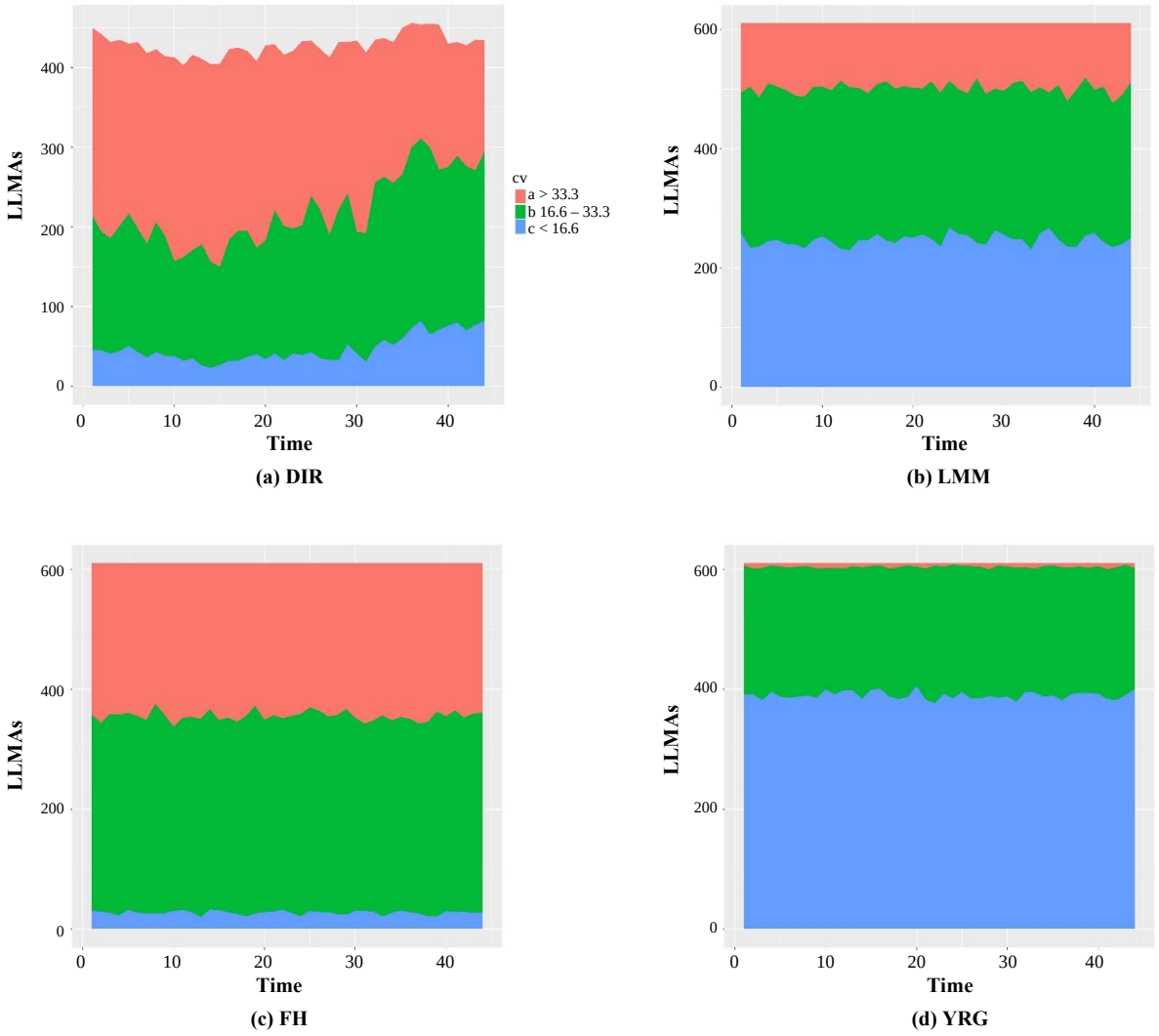


Figure 5.5 Distribution of the coefficients of variation for DIR, LMM, FH and YRG estimates for each quarter.

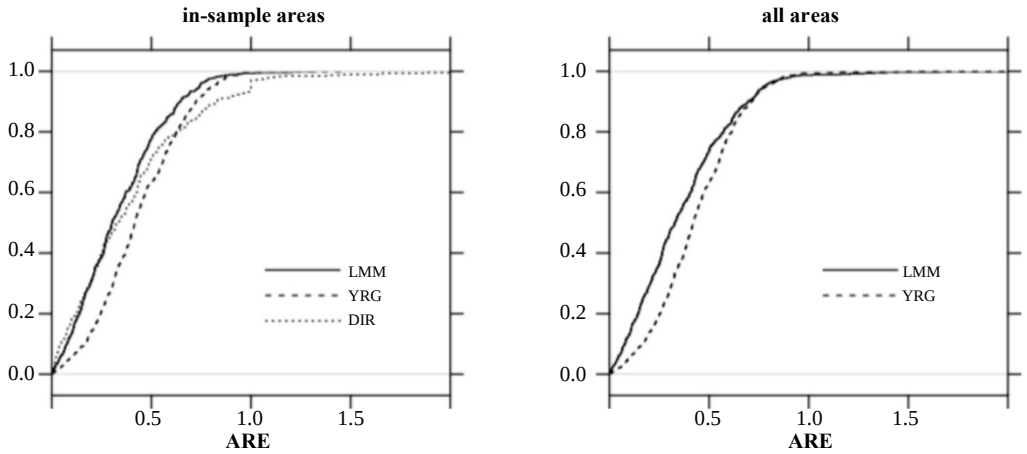


Figure 5.6 Empirical distribution of ARE_i, equation (5.1), for in-sample areas (left panel) and for all areas (right panel).

6 Final remarks

In this paper we develop a new area level SAE method that uses a Latent Markov Model (LMM) as the linking model. In LMMs (Bartolucci et al., 2013), the characteristic of interest, and its evolution in time, is represented by a latent process that follows a Markov chain, usually of first order. Under the assumption of normality for the conditional distribution of the response variables given the latent variables, the model is estimated using an augmented data Gibbs sampler. The proposed model has been applied to quarterly data from the Italian LFS from 2004 to 2014. The model-based method has been found to be effective for developing LMAs level estimates of unemployment incidence and the reduction in the coefficient of variation compared to the direct estimator is quite evident. The proposed approach is also more accurate than the direct and the time-series model-based estimator proposed by You et al. (2003) in reproducing census data. An advantage of this methodology is that it also provides a clustering of the small areas in homogeneous groups.

LMMs can be seen as an extension of latent class models to longitudinal data. In this regard, our approach represents an extension of the latent class SAE model proposed by Fabrizi et al. (2016). Moreover, LMMs may be seen as an extension of Markov chain models to control for measurement errors and can easily handle multivariate data, providing a very flexible modeling framework. The approach could be extended using spatial correlation information, and it could consider different distributions for the manifest variables, such as Poisson, Binomial, and Multinomial responses. In this scenario, we could fit unmatched sampling and linking models and handle departures from the normality assumption, but a Gibbs sampler cannot be used any longer, and Metropolis-Hastings sampling is an option. The proposed univariate model can account for measurement errors, but the extension to multivariate framework could be also possible, taking into account the conditional independence assumption.

In this application we have not accounted explicitly for the serial correlation induced by the rotating panel design. A natural way to take the different features of this design into account, such as the rotating group bias and the autocorrelation of the survey errors, is to use state space-model specifications, as in Pfeffermann (1991), Pfeffermann and Rubin-Bleuer (1993) and, more recently, Van den Brakel and Krieg (2015) and Boonstra and Van den Brakel (2016). In this context, it would also be interesting to extend to SAE the LMM with serial correlation in the measurement model proposed by Bartolucci and Farcomeni (2009). State space-model specifications can also be a useful tool to capture and model the strong trend and seasonality of this type of data.

Acknowledgements

The work of Bertarelli, Ranalli, D'Aló and Solari has been developed under the support of the project PRIN-2012F42NS8 "Household wealth and youth unemployment: new survey methods to meet current challenges". The Authors are grateful to the Associated Editor and two anonymous Referees who provided very useful comments on earlier versions of this paper.

Appendix A

Model estimation

In the following we first illustrate Bayesian estimation and model selection based on a MCMC algorithm which is implemented in a data augmentation framework (Tanner and Wong, 1987).

A.1 Data augmentation method

In order to estimate the small area parameters $\boldsymbol{\Theta}$, the measurement parameters $\boldsymbol{\phi}_{\text{obs}}$, and the latent parameters $\boldsymbol{\phi}_{\text{lat}}$, we follow a data augmentation approach. We recall that the observed data consist of the direct estimates $\hat{\theta}_{it}$, the corresponding smoothed $\widehat{\text{MSE}}_{it}$, and the covariate vectors $\mathbf{x}_{it}$, with $i = 1, \ldots, m$ and $t = 1, \ldots, T$. Moreover, the data augmentation approach explicitly introduces the latent variables U_{it} treated as missing data, the values of which are updated during the MCMC algorithm that is, therefore, based on a complete data likelihood. In this context, the use of conjugate priors to the complete data likelihood allows us to sample from the conditional posterior of the latent states in a straightforward way. Since the state space is finite, sampling the latent states conditionally given the model parameters is also simple.

To generate samples from the joint posterior distribution of the model parameters and latent states, the proposed MCMC algorithm proceeds as follows. Let $\hat{\boldsymbol{\Theta}}$ be the matrix of realizations of the available direct estimates that is defined as in (4.1), with each θ_{it} replaced by $\hat{\theta}_{it}$, and let $\mathbf{U}$ be the matrix of the latent variable U_{it}, with elements organized as in $\hat{\boldsymbol{\Theta}}$. Then the posterior distribution of all model parameters and latent variables, given the observed data, has the following expression:

$$p\left(\mathbf{U}, \boldsymbol{\phi}_{\text{lat}}, \boldsymbol{\phi}_{\text{obs}}, \boldsymbol{\Theta} \middle| \hat{\boldsymbol{\Theta}}\right) \propto p\left(\mathbf{U} \middle| \boldsymbol{\phi}_{\text{lat}}\right) \pi\left(\boldsymbol{\phi}_{\text{lat}}\right) \pi\left(\boldsymbol{\phi}_{\text{obs}}\right) p\left(\boldsymbol{\Theta} \middle| \mathbf{U}, \boldsymbol{\phi}_{\text{obs}}\right) p\left(\hat{\boldsymbol{\Theta}} \middle| \boldsymbol{\Theta}\right).$$

The MCMC algorithm alternates between sampling the latent variables and the parameters from the corresponding full conditional distribution. This scheme is repeated for R iterations. At the end of each iteration r, $r = 1, \ldots, R$, the sampled model parameters and latent variables are obtained and are denoted by $\mathbf{U}^{(r)}$, $\boldsymbol{\phi}_{\text{lat}}^{(r)}$, $\boldsymbol{\phi}_{\text{obs}}^{(r)}$, and $\boldsymbol{\Theta}^{(r)}$. More precisely, each iteration consists in:

1. drawing $\mathbf{U}^{(r)}$ from $p\left(\mathbf{U} \middle| \boldsymbol{\phi}_{\text{lat}}^{(r-1)}, \boldsymbol{\phi}_{\text{obs}}^{(r-1)}, \boldsymbol{\Theta}^{(r-1)}\right)$;
2. drawing $\boldsymbol{\phi}_{\text{lat}}^{(r)}$ from $p\left(\boldsymbol{\phi}_{\text{lat}} \middle| \mathbf{U}^{(r)}\right)$;
3. drawing $\boldsymbol{\phi}_{\text{obs}}^{(r)}$ from $p\left(\boldsymbol{\phi}_{\text{obs}} \middle| \mathbf{U}^{(r)}, \boldsymbol{\Theta}^{(r-1)}\right)$;
4. drawing $\boldsymbol{\Theta}^{(r)}$ from $p\left(\boldsymbol{\Theta} \middle| \mathbf{U}^{(r)}, \boldsymbol{\phi}_{\text{obs}}^{(r)}, \hat{\boldsymbol{\Theta}}\right)$.

In the following we illustrate in details each of the above steps. In this regard, note that our illustration is referred to the case where all elements of $\hat{\boldsymbol{\Theta}}$ are available. However, in our application, some elements of this matrix are missing. This requires minor adjustments to the MCMC algorithm, consisting in imputing the missing values by a Gibbs sampler and sampling directly from its full conditional distribution.

A.1.1 Simulation of $\mathbf{U}^{(r)}$

Each latent variable U_{it} is drawn separately from the corresponding full conditional distribution, which is of multinomial type with specific parameters. In particular, we have that

$$U_{it} \left| U_{i,t-1}^{(r)}, U_{i,t+1}^{(r-1)}, \boldsymbol{\phi}_{\text{lat}}^{(r-1)}, \boldsymbol{\phi}_{\text{obs}}^{(r-1)}, \boldsymbol{\Theta}^{(r-1)} \right. \sim \text{Multi}_k\left(\mathbf{q}_{it}\right), \quad t = 1, \ldots, T, \quad i = 1, \ldots, m, \qquad \text{(A.1)}$$

where $U_{i,t-1}^{(r)}$ disappears for $t = 1$ and $U_{i,t+1}^{(r)}$ disappears for $t = T$. Moreover, the probability vector $\mathbf{q}_{it}$ is defined as follows:

- for $t = 1$, $\mathbf{q}_{it}$ has elements proportional to

$$\pi_u^{(r-1)} \pi_{u \mid U_{i2}^{(r-1)}}^{(r-1)}, \quad u = 1, \ldots, k;$$

- for $t = 2, \ldots, T-1$, $\mathbf{q}_{it}$ has elements proportional to

$$\pi_{u \mid U_{i,t-1}^{(r)}}^{(r-1)} \pi_{U_{i,t+1}^{(r-1)} \mid u}^{(r-1)}, \quad u = 1, \ldots, k;$$

- for $t = T$, $\mathbf{q}_{it}$ has elements proportional to

$$\pi_{u \mid U_{i,T-1}^{(r)}}^{(r-1)}, \quad u = 1, \ldots, k.$$

A.1.2 Simulation of $\boldsymbol{\phi}_{\text{lat}}^{(r)}$

Recalling that $\boldsymbol{\phi}_{\text{lat}} = \{\boldsymbol{\pi}, \boldsymbol{\Pi}\}$, we first draw $\boldsymbol{\pi}^{(r)}$ from the full conditional distribution:

$$\boldsymbol{\pi} \,|\, \mathbf{U}^{(r)} \sim \text{Dirichlet}\left(\mathbf{1}_k + \mathbf{n}_1\right),$$

where $\mathbf{n}_1 = \left(n_{11}, \ldots, n_{1k}\right)'$ and n_{1u} is the number of areas in state u at time 1, with $u = 1, \ldots, k$. Moreover, we draw each row of matrix $\boldsymbol{\Pi}$ from the distribution

$$\boldsymbol{\pi}_{\bar{u}} \,|\, \mathbf{U}^{(r)} \sim \text{Dirichlet}\left(\mathbf{1}_k + \mathbf{n}_{\bar{u},t}\right), \quad t = 2, \ldots, T,$$

where $\mathbf{n}_{\bar{u},t} = \left(n_{\bar{u},t_1}, \ldots, n_{\bar{u},t_k}\right)'$ and $n_{\bar{u},t_u}$ is the number of areas moving from state $\bar{u}$ to state u at time t, with $t = 2, \ldots, T$ and $u, \bar{u} = 1, \ldots, k$.

A.1.3 Simulation of $\boldsymbol{\phi}_{\text{obs}}^{(r)}$

Considering that $\boldsymbol{\phi}_{\text{obs}} = \{\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_k, \sigma_1^2, \ldots, \sigma_k^2\}$, we first draw each $\boldsymbol{\beta}_u$, $u = 1, \ldots, k$, from the full conditional distribution:

$$\boldsymbol{\beta}_u \,|\, \mathbf{U}^{(r)}, \boldsymbol{\Theta}^{(r)} \sim N_p\left(\boldsymbol{\eta}_{1,u}, \boldsymbol{\Sigma}_{1,u}\right),$$

where

$$\begin{aligned}
\boldsymbol{\eta}_{1,u} &= \boldsymbol{\Lambda}_{1,u}^{-1} \sum_{i=1}^{m}\sum_{t=1}^{T} \theta_{it} I\left(U_{it} = u\right) \mathbf{x}_{it}, \\
\boldsymbol{\Sigma}_{1,u} &= \sigma_u^2 \boldsymbol{\Lambda}_{1,u}^{-1}, \\
\boldsymbol{\Lambda}_{1,u} &= \sum_{i=1}^{m}\sum_{t=1}^{T} \mathbf{x}_{it}\mathbf{x}_{it}' I\left(U_{it} = u\right) + \boldsymbol{\Lambda}_0,
\end{aligned}$$

with $I(\cdot)$ denoting the indicator function equal to 1 if its argument is true and to 0 otherwise. Then, we draw each σ_u^2 from

$$\sigma_u^2 \,|\, \mathbf{U}^{(r)}, \boldsymbol{\Theta}^{(r)} \sim \text{IG}\left(a_{1,u}, b_{1,u}\right),$$

with

$$\begin{aligned}
a_{1,u} &= a_0 + \frac{n_{.u}}{2}, \\
b_{1,u} &= b_0 + \frac{1}{2}\left(\sum_{i=1}^{m}\sum_{t=1}^{T} \theta_{it}^2 I\left(U_{it} = u\right) + \boldsymbol{\eta}_0'\boldsymbol{\Lambda}_0\boldsymbol{\eta}_0 - \boldsymbol{\eta}_{1,u}'\boldsymbol{\Lambda}_{1,u}\boldsymbol{\eta}_{1,u}\right),
\end{aligned}$$

where $n_{.u} = \sum_{t=1}^{T} n_{tu}$ is the number of areas in state u regardless of the specific time occasion.

A.1.4 Simulation of $\boldsymbol{\Theta}^{(r)}$

The goal of SAE is to predict each θ_{it}, $i = 1, \ldots, m$, $t = 1, \ldots, T$, based on the model and the observed data. This amounts to draw these elements from

$$\theta_{it} \,|\, \mathbf{U}^{(r)}, \boldsymbol{\phi}_{\text{obs}}^{(r)}, \hat{\theta}_{it} \sim N(\hat{\theta}_{it}^{(r)}, \gamma_{it}^{(r)}\psi_{it}),$$

where

$$\hat{\theta}_{it}^{(r)} = \gamma_{it}^{(r)}\hat{\theta}_{it} + \left(1 - \gamma_{it}^{(r)}\right)\mathbf{x}_{it}'\boldsymbol{\beta}_u^{(r)}, \tag{A.2}$$

with $\gamma_{it}^{(r)} = \sigma_u^{2(r)} \Big/ \left(\sigma_u^{2(r)} + \psi_{it}\right)$.

A.2 Model selection: The Chib estimator

The method proposed in Chib (1995) can be applied to perform model selection starting from the Gibbs sampler output. It is known that the posterior density can be written as the ratio of the product of the likelihood function and the priors divided by the marginal likelihood:

$$p\left(\mathbf{U}, \boldsymbol{\phi}_{\text{lat}}, \boldsymbol{\phi}_{\text{obs}}, \boldsymbol{\Theta} \,\middle|\, \hat{\boldsymbol{\Theta}}\right) = \frac{p\left(\mathbf{U} \,|\, \boldsymbol{\phi}_{\text{lat}}\right)\pi\left(\boldsymbol{\phi}_{\text{lat}}\right)\pi\left(\boldsymbol{\phi}_{\text{obs}}\right)p\left(\boldsymbol{\Theta} \,|\, \mathbf{U}, \boldsymbol{\phi}_{\text{obs}}\right)p\left(\hat{\boldsymbol{\Theta}} \,|\, \boldsymbol{\Theta}\right)}{p\left(\hat{\boldsymbol{\Theta}}\right)}. \tag{A.3}$$

Therefore, it is possible to write the marginal likelihood of the data $\hat{\boldsymbol{\Theta}}$ as

$$p\left(\hat{\boldsymbol{\Theta}}\right) = \frac{p\left(\hat{\boldsymbol{\Theta}} \,|\, \boldsymbol{\Theta}\right)p\left(\mathbf{U} \,|\, \boldsymbol{\phi}_{\text{lat}}\right)\pi\left(\boldsymbol{\phi}_{\text{lat}}\right)\pi\left(\boldsymbol{\phi}_{\text{obs}}\right)p\left(\boldsymbol{\Theta} \,|\, \mathbf{U}, \boldsymbol{\phi}_{\text{obs}}\right)}{p\left(\mathbf{U}, \boldsymbol{\phi}_{\text{lat}}, \boldsymbol{\phi}_{\text{obs}}, \boldsymbol{\Theta} \,\middle|\, \hat{\boldsymbol{\Theta}}\right)}, \tag{A.4}$$

for any $\mathbf{U}$, $\boldsymbol{\phi}_{\text{lat}}$, $\boldsymbol{\phi}_{\text{obs}}$, $\boldsymbol{\Theta}$ and $\hat{\boldsymbol{\Theta}}$. We drop the dependence on k for ease of notation. This is the model selection criterion used in Section 5. Then, choosing specific values of the latent variables and model parameters, denoted by $\bar{\mathbf{U}}$, $\bar{\boldsymbol{\phi}}_{\text{lat}}$, $\bar{\boldsymbol{\phi}}_{\text{obs}}$, and $\bar{\boldsymbol{\Theta}}$, we can estimate $\log p\left(\hat{\boldsymbol{\Theta}}\right)$ through the following decomposition:

$$\begin{aligned}\log p\left(\hat{\boldsymbol{\Theta}}\right) = \; & \log p\left(\hat{\boldsymbol{\Theta}} \,\middle|\, \bar{\boldsymbol{\Theta}}\right) + \log p\left(\bar{\mathbf{U}} \,\middle|\, \bar{\boldsymbol{\phi}}_{\text{lat}}\right) + \log \pi\left(\bar{\boldsymbol{\phi}}_{\text{lat}}\right) + \log \pi\left(\bar{\boldsymbol{\phi}}_{\text{obs}}\right) \\ & + \log p\left(\bar{\boldsymbol{\Theta}} \,\middle|\, \bar{\mathbf{U}}, \bar{\boldsymbol{\phi}}_{\text{obs}}\right) - \log p\left(\bar{\mathbf{U}}, \bar{\boldsymbol{\phi}}_{\text{lat}}, \bar{\boldsymbol{\phi}}_{\text{obs}}, \bar{\boldsymbol{\Theta}} \,\middle|\, \hat{\boldsymbol{\Theta}}\right).\end{aligned} \tag{A.5}$$

The use of the log transformation is motivated by numerical stability (Chib, 1995).

The first five terms at the right hand side of (A.5) can be computed directly from the assumed distributions of the parameters and the data. On the other hand obtaining the last component is more challanging. By the law of total probability, $p\left(\bar{\mathbf{U}}, \bar{\boldsymbol{\phi}}_{\text{lat}}, \bar{\boldsymbol{\phi}}_{\text{obs}}, \bar{\boldsymbol{\Theta}} \,\middle|\, \hat{\boldsymbol{\Theta}}\right)$ may be decomposed as

$$p\left(\bar{\mathbf{U}}, \bar{\boldsymbol{\phi}}_{\text{lat}}, \bar{\boldsymbol{\phi}}_{\text{obs}}, \bar{\boldsymbol{\Theta}} \,\middle|\, \hat{\boldsymbol{\Theta}}\right) = p\left(\bar{\mathbf{U}} \,\middle|\, \bar{\boldsymbol{\phi}}_{\text{lat}}, \bar{\boldsymbol{\phi}}_{\text{obs}}, \bar{\boldsymbol{\Theta}}, \hat{\boldsymbol{\Theta}}\right)p\left(\bar{\boldsymbol{\phi}}_{\text{lat}} \,\middle|\, \bar{\boldsymbol{\phi}}_{\text{obs}}, \bar{\boldsymbol{\Theta}}, \hat{\boldsymbol{\Theta}}\right)p\left(\bar{\boldsymbol{\phi}}_{\text{obs}} \,\middle|\, \bar{\boldsymbol{\Theta}}, \hat{\boldsymbol{\Theta}}\right)p\left(\bar{\boldsymbol{\Theta}} \,\middle|\, \hat{\boldsymbol{\Theta}}\right). \tag{A.6}$$

Following Chib (1995), we compute the first term of (A.6) following the Gibbs scheme outlined in Section A.1, whereas, the other three terms are estimated from the Gibbs output. In particular, we estimate

$$p\left(\bar{\boldsymbol{\phi}}_{\text{lat}} \,\middle|\, \bar{\boldsymbol{\phi}}_{\text{obs}}, \bar{\boldsymbol{\Theta}}, \hat{\boldsymbol{\Theta}}\right) = \int p\left(\bar{\boldsymbol{\phi}}_{\text{lat}} \,\middle|\, \bar{\mathbf{U}}, \bar{\boldsymbol{\phi}}_{\text{obs}}, \bar{\boldsymbol{\Theta}}, \hat{\boldsymbol{\Theta}}\right)p\left(\bar{\mathbf{U}} \,\middle|\, \bar{\boldsymbol{\phi}}_{\text{obs}}, \bar{\boldsymbol{\Theta}}, \hat{\boldsymbol{\Theta}}\right)d\bar{\mathbf{U}}$$

as $R^{-1}\sum_{r=1}^{R} p\left(\bar{\boldsymbol{\phi}}_{\text{lat}} \,\middle|\, \mathbf{U}^{(r)}, \bar{\boldsymbol{\phi}}_{\text{obs}}, \bar{\boldsymbol{\Theta}}\right)$, based on R draws from a reduced Gibbs sampling where $\mathbf{U}$ is not updated. In order to estimate

$$p\left(\bar{\boldsymbol{\phi}}_{\text{obs}} \,\middle|\, \bar{\boldsymbol{\Theta}}, \hat{\boldsymbol{\Theta}}\right) = \int p\left(\bar{\boldsymbol{\phi}}_{\text{obs}} \,\middle|\, \bar{\mathbf{U}}, \bar{\boldsymbol{\phi}}_{\text{lat}}, \bar{\boldsymbol{\Theta}}, \hat{\boldsymbol{\Theta}}\right)p\left(\bar{\mathbf{U}}, \bar{\boldsymbol{\phi}}_{\text{lat}} \,\middle|\, \bar{\boldsymbol{\Theta}}, \hat{\boldsymbol{\Theta}}\right)d\bar{\mathbf{U}}\, d\bar{\boldsymbol{\phi}}_{\text{lat}},$$

we use $R^{-1}\sum_{r=1}^{R} p\left(\overline{\boldsymbol{\phi}}_{\text{obs}} \mid \mathbf{U}^{(r,1)}, \overline{\boldsymbol{\phi}}_{\text{lat}}^{(r,1)}, \overline{\boldsymbol{\Theta}}\right)$. Finally, to estimate

$$p\left(\overline{\boldsymbol{\Theta}} \mid \hat{\boldsymbol{\Theta}}\right) = \int p\left(\overline{\boldsymbol{\Theta}} \mid \overline{\mathbf{U}}, \overline{\boldsymbol{\phi}}_{\text{lat}}, \overline{\boldsymbol{\phi}}_{\text{obs}}, \hat{\boldsymbol{\Theta}}\right) p\left(\overline{\mathbf{U}}, \overline{\boldsymbol{\phi}}_{\text{lat}}, \overline{\boldsymbol{\phi}}_{\text{obs}} \mid \hat{\boldsymbol{\Theta}}\right) d\overline{\mathbf{U}}\, d\overline{\boldsymbol{\phi}}_{\text{lat}}\, d\overline{\boldsymbol{\phi}}_{\text{obs}},$$

we use $R^{-1}\sum_{r=1}^{R} p\left(\overline{\boldsymbol{\Theta}} \mid \mathbf{U}^{(r,2)}, \boldsymbol{\phi}_{\text{lat}}^{(r,2)}, \boldsymbol{\phi}_{\text{obs}}^{(r,2)}\right)$, with R draws from a third reduced Gibbs sampling.

References

Bartolucci, F., and Farcomeni, A. (2009). A multivariate extension of the dynamic logit model for longitudinal data based on a latent Markov heterogeneity structure. *Journal of the American Statistical Association*, 104, 816-831.

Bartolucci, F., Farcomeni, A. and Pennoni, F. (2013). *Latent Markov Models for Longitudinal Data*. Boca Roton, FL: CRC Press.

Bartolucci, F., Lupparelli, M. and Montanari, G.E. (2009). Latent Markov model for longitudinal binary data: An application to the performance evaluation of nursing homes. *The Annals of Applied Statistics*, 3, 611-636.

Bartolucci, F., Pennoni, F. and Francis, B. (2007). A latent Markov model for detecting patterns of criminal activity. *Journal of the Royal Statistical Society, Series A*, 170, 115-132.

Boonstra, H.J. (2012). hbsae: Hierarchical Bayesian small area estimation. *R Package Version 1*.

Boonstra, H.J. (2014). Time-series small area estimation for unemployment based on a rotating panel survey. Technical report, CBS. Available at https://www.cbs.nl/nl-nl/achtergrond/2014/25/time-series-small-area-estimation-for-unemployment-based-on-a-rotating-panel-survey.

Boonstra, H.J., and van den Brakel, J.A. (2016). *Estimation of Level and Change for Unemployment Using Multilevel and Structural Time Series Models*. Discussion paper 2016-10. Statistics Netherlands, Heerlen.

Boys, R., and Henderson, D. (2003). Data augmentation and marginal updating schemes for inference in hidden Markov models. Technical report, Univ. Newcastle.

Carlin, B.P., and Chib, S. (1995). Bayesian model choice via Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society, Series B*, 56, 473-484.

Chib, S. (1995). Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association*, 90, 1313-1321.

D'Alò, M., Di Consiglio, L., Falorsi, S., Ranalli, M.G. and Solari, F. (2012). Use of spatial information in small area models for unemployment rate estimation at sub-provincial areas in Italy. *Journal of the Indian Society of Agricultural Statistics*, 66, 43-53.

Datta, G.S., Lahiri, P., Maiti, T. and Lu, K.L. (1999). Hierarchical Bayes estimation of unemployment rates for the states of the US. *Journal of the American Statistical Association*, 94, 1074-1082.

Fabrizi, E., Montanari, G.E. and Ranalli, M.G. (2016). A hierarchical latent class model for predicting disability small area counts from survey data. *Journal of the Royal Statistical Society, Series A*, 179, 103-132.

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, 269-277.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis*, 1(3), 515-534.

Gelman, A., Jakulin, A., Pittau, M.G. and Su, Y.-S. (2008). A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, 2, 1360-1383.

Germain, S.E. (2010). *Bayesian Spatio-Temporal Modelling of Rainfall Through Non-Homogenous Hidden Markov Models*. Ph.D. thesis, University of Newcastle Upon Tyne.

Ghosh, M., Nangia, N. and Kim, D.H. (1996). Estimation of median income of four-person families: A Bayesian time series approach. *Journal of the American Statistical Association*, 91, 1423-1431.

Harvey, A., and Chung, C.-H. (2000). Estimating the underlying change in unemployment in the UK. *Journal of the Royal Statistical Society, Series A*, 163, 303-309.

Istat (2014). I sistemi locali del lavoro 2011. *Rapporto Annuale 2014*.

Jasra, A., Holmes, C. and Stephens, D. (2005). Markov chain Monte Carlo methods and the label switching problem in Bayesian mixture modeling. *Statistical Science*, 20, 50-67.

Krieg, S., and van der Brakel, J.A. (2012). Estimation of the monthly unemployment rate for six domains through structural time series modelling with cointegrated trends. *Computational Statistics & Data Analysis*, 56, 2918-2933.

Lazarsfeld, P.F., Henry, N.W. and Anderson, T.W. (1968). *Latent Structure Analysis*. Houghton Mifflin Boston.

Liu, J.S., Wong, W.H. and Kong, A. (1994). Covariance structure of the Gibbs sampler with applications to the comparisons of estimators and augmentation schemes. *Biometrika*, 81, 27-40.

MacDonald, I.L., and Zucchini, W. (1997). *Hidden Markov and Other Models for Discrete-Valued Time Series*. London: Series Chapman & Hall.

Marhuenda, Y., Molina, I. and Morales, D. (2013). Small area estimation with spatio-temporal Fay-Herriot models. *Computational Statistics & Data Analysis*, 58, 308-325.

Marin, J.-M., Mengersen, K. and Robert, C.P. (2005). Bayesian modelling and inference on mixtures of distributions. *Handbook of Statistics*, 25, 459-507.

Meng, X.-L. (1994). Posterior predictive p-values. *The Annals of Statistics*, 1142-1160.

Pfeffermann, D. (1991). Estimation and seasonal adjustment of population means using data from repeated surveys. *Journal of Business & Economic Statistics*, 9, 163-175.

Pfeffermann, D., and Burck, L. (1990). Robust small area estimation combining time series and cross-sectional data. *Survey Methodology*, 16, 2, 217-237. Paper available at https://www150.statcan.gc.ca/n1/pub/12-001-x/1990002/article/14534-eng.pdf.

Pfeffermann, D., and Rubin-Bleuer, S. (1993). Robust joint modelling of labour force series of small areas. *Survey Methodology*, 19, 2, 149-163. Paper available at https://www150.statcan.gc.ca/n1/pub/12-001-x/1993002/article/14458-eng.pdf.

Pfeffermann, D., and Tiller, R. (2006). Small-area estimation with state-space models subject to benchmark constraints. *Journal of the American Statistical Association*, 101, 1387-1397.

Polson, N.G., and Scott, J.G. (2012). On the half-cauchy prior for a global scale parameter. *Bayesian Analysis*, 7, 4, 887-902.

Rao, J.N.K. (2003). *Small Area Estimation*. Wiley Online Library.

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.

Spezia, L. (2010). Bayesian analysis of multivariate gaussian hidden Markov models with an unknown number of regimes. *Journal of Time Series Analysis*, 31, 1-11.

Statistics Canada (2016). Guide to the Labour Force Survey. Technical report, Statistics Canada, Catalogue 71-543-G, available at http://www.statcan.gc.ca/pub/71-543-g/71-543-g2016001-eng.pdf.

Tanner, M.A., and Wong, W.H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American statistical Association*, 82, 528-540.

Van den Brakel, J.A., and Krieg, S. (2015). Dealing with small sample sizes, rotation group bias and discontinuities in a rotating panel design. *Survey Methodology*, 41, 2, 267-296. Paper available at https://www150.statcan.gc.ca/n1/pub/12-001-x/2015002/article/14231-eng.pdf.

Van der Brakel, J.A., and Krieg, S. (2016). Small area estimation with state space common factor models for rotating panels. *Journal of the Royal Statistical Society, Series A*, 179, 763-791.

Van Dyk, D.A., and Meng, X.-L. (2001). The art of data augmentation. *Journal of Computational and Graphical Statistics*, 10(1), 1-50.

Vermunt, J.K., and Magidson, J. (2002). Latent class cluster analysis. *Applied Latent Class Analysis*, 11, 89-106.

Wiggins, L.M. (1973). *Panel Analysis: Latent Probability Models for Attitude and Behavior Processes*. Jossey-Bass.

Wolter, K. (2007). *Introduction to Variance Estimation*. New York: Springer Science & Business Media.

You, Y., Rao, J.N.K. and Gambino, J. (2003). Model-based unemployment rate estimation for the Canadian Labour Force Survey: A hierarchical Bayes approach. *Survey Methodology*, 29, 1, 25-32. Paper available at https://www150.statcan.gc.ca/n1/pub/12-001-x/2003001/article/6602-eng.pdf.