# Hierarchical Bayesian time series multilevel models for consistent small area estimates at different frequencies and regional levels

Harm Jan Boonstra
Jan van den Brakel

**December 4, 2018**

**Abstract**

**A small area estimation method is developed to produce monthly unemployment figures at a provincial level and quarterly figures at a municipal level. To this end a time series multilevel model is proposed where monthly direct estimates for municipalities and accompanying variance estimates form the input. Consistent estimates for monthly provincial figures and quarterly municipal figures are derived by aggregating the monthly municipal predictions obtained with the time series multilevel model. The model is formulated in an hierarchical Bayesian framework and fitted using MCMC simulations. The model borrows strength over time and space in several ways. Municipalities belonging to the same province share a common provincial smooth trend model. Deviations from this overall trend for the separate municipalities are modelled with local level models. This is a parsimonious alternative for modelling a full covariance matrix between the trend innovations. The model also borrows information from auxiliary series derived from a claimant counts register. Regression coefficients for claimant counts vary between municipalities by modelling random slopes. The regression coefficients can vary over time by modelling the innovations of the regression coefficients with random walk or a smooth trend model. Another way of including cross-sectional correlations is obtained by modelling the spatial effects among random domain intercepts and among random slopes for the regression coefficients with a spatial autoregressive model. To allow for the diversity of municipalities and possibly volatile time-dependence, non-normally distributed municipal random effects and trend innovations are investigated by using global-local shrinkage priors. It is found that the estimates based on the time series multilevel models compare favourably to estimates based on cross-sectional multilevel models that are currently used to produce official annual provincial and municipal unemployment figures.**

*Key words: time series multilevel models, Fay-Herriot models, hierarchical Bayesian models, non-normal priors, global local priors, Gibbs sampler*

This work was funded by the European Union under grant no. 07131.2017.003-2017.596.

# 1 Introduction

Data from the Dutch Labour Force Survey (LFS) are used to estimate labour status at various aggregation levels. National estimates are produced monthly, provincial estimates quarterly, and municipal estimates annually. Many more figures are produced for several demographic subgroups.

Until 2015 municipal estimates were produced annually by means of direct generalized regression estimation (GREG, see e.g. Särndal et al. (1992)), but only for municipalities with at least 30 thousand inhabitants. For municipalities with 10 to 30 thousand inhabitants, three-year moving averages of GREG estimates were used in order to reduce the variance. No estimates were published for municipalities with fewer than 10 thousand inhabitants. To improve the municipal estimates, a model-based small area estimation (SAE) strategy has been adopted starting 2015 (Boonstra et al., 2011; Boonstra and Michiels, 2013). The model used for this purpose is the Battese-Harter-Fuller basic unit-level model (Battese et al., 1988; Rao and Molina, 2015), a linear multilevel model with random municipality effects. Based on this model, estimates of labour status are produced annually for all provinces and municipalities.

The continuous nature of the LFS allows to borrow strength not only from other areas, but also over time. A structural time-series model (STM) is already being used since 2010 to estimate monthly figures about labour status at the national level and a breakdown for six gender by age classes (van den Brakel and Krieg, 2009, 2015). These official monthly publications are based on a state-space model that is used as a form of small area estimation by borrowing strength over time, and that also accounts for rotation group bias (RGB) and serial correlation in the survey errors due to rotating panel design of the Dutch Labour Force Survey. This model uses five series of GREG estimates observed in the separate waves of the rotating panel as input series. The model accounts for RGB by benchmarking the labour force estimates to the level of the first wave of the panel.

The aim of this paper is to obtain improved small area estimates by borrowing strength over both space and time. The proposed time series multilevel models account for temporal and cross-sectional correlations, and in addition use related auxiliary series obtained from a register of claimant counts. Both monthly estimates of provincial unemployment and quarterly estimates of municipal unemployment are computed based on the same time series model. This implies that the estimates will be numerically consistent. This is an advantage compared to using different models for estimates at different aggregation levels.

Previous accounts of regional small area estimation of unemployment, where strength is borrowed over both time and space, include Rao and Yu (1994); Datta et al. (1999); You et al. (2003); You (2008a); Pfeffermann and Burck (1990); Pfeffermann and Tiller (2006), see also Rao and Molina (2015) for an overview. In Boonstra (2014) several multilevel time-series models have been applied to the estimation of annual unemployment levels for Dutch municipalities. In Boonstra and van den Brakel (2016) time series small area estimators are developed for monthly provincial unemployment figures. In that paper a comparison is made between state space models fitted in a frequentist framework and multilevel time series models fitted with the Gibbs sampler in an hierarchical Bayesian framework. Based on the results, it was decided to further develop time series small area estimators for this project using time series multilevel models fitted with a Gibbs sampler in an hierarchical Bayesian framework.

The models considered are applied to direct municipal estimates at a monthly frequency and are extensions of the well-known Fay-Herriot model (Fay and Herriot, 1979). First, direct estimates are computed for each municipality in each month using the GREG estimator that uses auxiliary information to reduce non-response bias. In addition the GREG estimates are calibrated to the official monthly publications at the national level and their breakdown in six domains. In this way the input series account for RGB in the same way as the monthly national figures. The LFS rotating panel design induces an autocorrelation structure among the monthly GREG estimates. These correlations are estimated along with variances for the GREG estimates. The GREG estimates and their estimated variances and covariances are subsequently modelled in an hierarchical Bayesian time series multilevel model. The model is augmented with month-specific covariates to ensure that the small area predictions are consistent with the official monthly publication at the national level, following an internal benchmarking approach described in Bell et al. (2013).

To borrow strength over time, smooth trend models or local level trend models are defined at different aggregation levels. To borrow strength over space, the correlation between innovations of these trend models can be modelled. An alternative and more parsimonious approach is obtained by defining an overall smooth trend at the provincial level for all municipalities belonging to the same province (on a monthly frequency). Deviations from this overall trend for the separate municipalities are modelled with a local level model at municipal level (on a quarterly frequency). Another way of borrowing strength over space is obtained by modelling spatial effects between random intercepts of the municipalities.

Regression components can be included to borrow strength from auxiliary series derived from a claimant count register. Regression coefficients are made domain-specific by modelling random slopes at the level of municipalities. Since relations between LFS trends and claimant counts are not necessarily time-invariant, a model with dynamic regression coefficients is considered. This is achieved by modelling the regression coefficients with a random walk, which is standard in state space models. As an alternative we also consider a smooth trend model for the time-varying regression coefficients.

To allow for the diversity of municipalities and for possibly volatile time-dependence, non-normally distributed random domain effects, spatial effects, trend innovations, and dynamic regression coefficient innovations are considered. In particular, we consider t-distributions, the horseshoe prior and Laplace distributions, which can all be understood in terms of scale mixtures of normal distributions (Andrews and Mallows, 1974).

Models with different combinations of fixed and random effects are compared based on the Deviance Information Criterion (DIC) (Spiegelhalter et al., 2002) and Widely Applicable Information Criterion or Watanabe-Akaike Information Criteria (WAIC) (Watanabe, 2010, 2013). In addition we use posterior predictive checks (Gelman et al., 1996; You, 2008b), and compare with both the direct estimates and the cross-sectional small area estimates.

This paper contributes to the existing literature on small area estimation by proposing an hierarchical Bayesian extension of the Fay-Herriot model that borrows strength over time and space as well as from auxiliary series. The approach is very general and flexible since it allows modelling correlations between trend innovations, the correlation between random domain effects and the random slopes of regression coefficients, and allows using spatial correlations and non-normally distributed random effects as well. The approach is computationally efficient, since it can handle large data sets. In this application 388 time series with a length of 66 observations are combined in a single model. The structure of the finally chosen model is partly based on the output requirements. The model is defined at the most detailed regional and temporal levels of interest and estimates at different levels of interest are obtained by aggregation, thus guaranteeing that all estimates are mutually numerically consistent.

This report is structured as follows. In Section 2 the LFS data used in this study are described. Section 3 describes how the monthly municipal GREG estimates are computed. Section 4 discusses the multilevel models used to model the GREG estimates. In Section 5 the results for several multilevel time series models are compared discussed and compared with the direct estimates and the cross-sectional small area estimates. Section 6 concludes with a discussion. Details about the Gibbs sampler used to estimate the models are given in the appendix.

# 2 Data from the Dutch Labour Force Survey

The Dutch LFS is a household survey conducted according to a rotating panel design in which the respondents are interviewed five times at quarterly intervals. In the years considered in this study, the first wave of the panel consists of data collected by means of web interviewing where the non-response is followed up by either computer assisted personal interviewing (CAPI) or telephone interviewing (CATI), whereas for the four follow-up waves data is collected only by means of CATI. For a more detailed description of the sampling design, we refer to Boonstra et al. (2008). There have been several changes of design mostly concerning the modes of observation. These changes have led to discontinuities that can be accounted for by adding intervention effects to the time-series model (van den Brakel and Krieg, 2015). Here we do not account for any discontinuities because we only use LFS data from after the last redesign.

The rotating panel design results in partial sample overlap between the subsequent sample periods, which gives rise to autocorrelation in the sampling errors. Another consequence of the rotating panel design is that there are systematic differences between the outcomes of the subsequent waves. This is a well known phenomenon in rotating panel designs and is generally termed rotation group bias (RGB) (Bailar, 1975). In the Dutch LFS the estimates for the unemployed labour force in the first wave are systematically higher compared to the follow-up waves, which is the result of many possible causes, including selection, mode and panel effects, see van den Brakel and Krieg (2009).

Since 2010 a multivariate structural time series model is used to publish monthly figures about the Dutch LFS at the national level and a breakdown by gender and three age classes. This model is used as a form of small area estimation to produce sufficiently precise estimates and accounts for the two aforementioned aspects of the rotating panel design. The model benchmarks the time series estimates for the target variables to the level of the first wave, which is considered to be the most reliable. Additionally the model accounts for the autocorrelation in the sampling error due to the sample overlap of the panel design.

In the present study we use 66 months of LFS data that became available after the last redesign, from 2012-04 until 2017-09. Data from all five waves of the rotating panel are used. The

Netherlands is divided into twelve provinces and as of 2017 into 388 municipalities. The aim is to estimate unemployment figures for provinces monthly and for municipalities quarterly. For the municipal estimates the 2017 classification is used for the whole study period.

The target variable considered is the fraction of unemployed in a domain, and is defined as $Y_{it} = \sum_{j\in i} y_{ijt}/N_{it}$, with $y_{ijt}$ equal to one if person $j$ from domain $i$ in period $t$ is unemployed and zero otherwise and $N_{it}$ the population size in domain $i$ and period $t$.

The next section discusses the computation of direct estimates and then Section 4 continues with the description of the multilevel time-series models. These estimates will be compared to each other and also to the cross-sectional small area estimates (SAE). As mentioned in the introduction, a cross-sectional SAE method based on the Battese-Harter-Fuller unit-level model is currently used to produce the official figures on annual municipal unemployment. These estimates along with estimated standard errors are actually computed quarterly, and so we compare quarterly municipal estimates based on the time-series models to them. Current annual official figures on provincial unemployment are also based on a Battese-Harter-Fuller unit-level model, and these will also be compared to the (quarterly) provincial estimates based on the time-series models. The provincial cross-sectional SAE estimates are benchmarked to agree with the national-level estimate, and the municipal SAE estimates are in turn benchmarked to agree with the provincial estimates. This is necessary because the estimates for the different aggregation levels are based on separate models. Due to the different nature of the cross-sectional unit-level SAE model compared to the aggregate-level time-series models discussed in Section 4, there are necessarily some systematic differences between them regarding the way they deal with non-sampling error such as RGB and non-response bias. In the cross-sectional models the RGB is handled by including wave number as a categorical predictor and using the coefficient of the first wave for prediction. Some further differences exist between the covariates used in the direct monthly estimates discussed below and the unit-level SAE model, where the latter uses a more extensive set of covariates and thus might be slightly better in reducing non-response bias. Further details on the cross-sectional SAE estimation method can be found in Boonstra and Michiels (2013).

# 3 Direct estimates

Let $\hat{\hat{Y}}_{it}$ denote the GREG estimate of the unemplopyment fraction for area $i$ in period $t$ based on data observed in all 5 waves of period $t$. The GREG estimator can be expressed as (Särndal et al., 1992)

$$\hat{\hat{Y}}_{it} = \frac{\sum_{j\in s} w_{ijt} y_{ijt} \delta_{ijt}}{\sum_{j\in s} w_{ijt} \delta_{ijt}}, \qquad (1)$$

with $\delta_{ijt}$ a domain indicator which is equal to one if person $j$ belongs to domain $i$ in period $t$, and $w_{ijt}$ the regression weight for person $j$ from domain $i$ in period $t$. These weights are derived from a linear regression model at the household level using several auxiliary variables available from population registers as explanatory variables, see (Särndal et al., 1992, Chapter 6).

The available auxiliary variables are listed in Table 3.1. The model selected to compute the GREG estimates is

$$ru + inct + inc6 + hhtype + region + ethn + gender * age21 + mlf * age3 * gender \qquad (2)$$

| variable | categories |
|---|---|
| gender | male, female |
| inct | type of income: salary, payment, unknown |
| inc6 | yearly income in euros in six categories: |
| | <3.000, 3.000-<10.000, 10.000-<15.000, 15.000-<20.000, 20.000-<30.000, 30.000+ |
| age3 | 15-24, 25-44, 45-64 |
| age21 | 0-14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25-29, 30-34, 35-39, 40-44, 45-49, |
| | 50-54, 55-59, 60-64, 65-74, 75+ |
| region | subdivision of the Netherlands in 43 regions |
| hhtype | single, household with children, other |
| ethn | native, Western immigrant, non-Western immigrant |
| ru | not registered as unemployed, employed and registered, |
| | registered unemployed < 1 year, registered unemployed 1-4 years, |
| | registered unemployed>4 years |
| mlf | monthly labour force estimates: employed, unemployed, not in the labour force |

**Table 3.1 Auxiliary variables used for GREG weighting**

The interactions denoted by '$*$' imply that the main effects are also included.

Note the presence of registered unemployment. Despite being based on a very different concept of unemployment, it is a strong predictor for the unemployment variable of interest. The last term contains the official monthly estimates obtained with the state space model. This is a breakdown of employment status (employed, unemployed, not in the labour force) by gender times three age classes. This term is included to enforce consistency between the domain estimates on the municipal and provincial level with the official monthly figures at the national level. As a result, the GREG estimates, which are based on monthly samples obtained in all five waves, are adjusted for RGB, consistently with the official monthly publications. More details about RGB and the way it is handled in the monthly estimates can be found in van den Brakel and Krieg (2009, 2015); Boonstra and van den Brakel (2016)

The time series models also require variance estimates corresponding to the GREG estimates that are used as input series for these models. We use the following cross-sectionally smoothed estimates of the design variances of the GREG estimates,

$$v(\hat{\hat{Y}}_{it}) = \frac{1}{m_{it}} \frac{1}{m_t} \sum_{i=1}^{M_A} m_{it} \hat{\sigma}_{it}^2 \equiv \hat{\sigma}_t^2 / m_{it}\,, \tag{3}$$

with $m_t$ the number of households in the sample of period $t$ and $m_{it}$ the number of households in the sample of period $t$ belonging to domain $i$, $M_A$ the number of municipalities, and

$$\hat{\sigma}_{it}^2 = \frac{1}{\hat{N}_{it}^2} \frac{m_t}{(m_t - 1)} \left( \sum_{h=1}^{m_t} \hat{E}_{iht}^2 - \frac{1}{m_t} \left( \sum_{h=1}^{m_t} \hat{E}_{iht} \right)^2 \right). \tag{4}$$

Here, $\hat{E}_{iht} = \sum_{j \in h} w_{ijt} \hat{e}_{ijt}$ is the sum over the weighted residuals of persons belonging to the same household, and

$$\hat{N}_{it} = \sum_{j \in s} w_{ijt} \delta_{ijt} \tag{5}$$

denotes the estimated population size of domain $i$. The within-area variances $\hat{\sigma}_{it}^2$ are pooled over the domains to obtain more stable variance estimates.

The panel design induces several non-zero correlations among initial estimates for the same province and different time periods and waves. These positive correlations are due to partial overlap of the sets of sample units on which the estimates are based. Such correlations exist between estimates for the same municipalities in months $t_1, t_2$ whenever $t_2 - t_1 \leq 12$. The covariances between $\hat{\hat{Y}}_{it_1}$ and $\hat{\hat{Y}}_{it_2}$ are estimated as (see e.g. Kish (1965))

$$v(\hat{\hat{Y}}_{it_1}, \hat{\hat{Y}}_{it_2}) = \frac{m_{it_1t_2}}{\sqrt{m_{it_1}m_{it_2}}} \hat{\rho}_{t_1t_2} \sqrt{v(\hat{\hat{Y}}_{it_1}) v(\hat{\hat{Y}}_{it_2})}, \tag{6}$$

where $m_{it_1t_2}$ is the number of units in the overlap, i.e. the number of observations on the same units in area $i$ for periods $(t_1)$ and $(t_2)$. The estimated (auto)correlation coefficient $\hat{\rho}_{t_1t_2}$ is computed as the correlation between the residuals of the linear regression models underlying the GREG estimator at $t_1$ and $t_2$, based on the overlap of both samples over all areas. This way they are pooled over areas, as are the variances $\hat{\sigma}^2_{it}$.

The Dutch municipalities have very diverse population sizes. The population sizes for the target population aged 15-75 ranges from less than 700 to almost 700000. This also means that the data is highly imbalanced over the municipalities. For 158 of the 25608 month-municipality combinations there is actually no response at all, i.e. $m_{it} = 0$, and in these cases the direct estimates are not defined. The data to fit the multi-level models therefore consists of $25608 - 158 = 25450$ month-municipality combinations. Based on the fitted models, predictions are computed for all 25608 combinations.

# 4 Time-series small area estimation

## 4.1 Time series multilevel models

The multilevel time series models for small area prediction are an extension of the area level model proposed by Fay and Herriot (1979). The data for the time-series model consist of time series of GREG estimates at the municpal level as well as variance estimates and covariance estimates induced by the rotating panel design as described in Section 3. A major advantage of the area level model is that the GREG estimates account for the sample design and reduce non-repsonse bias. A unit-level model for all sample periods that accounts for all temporal and cross-sectional effects would become very complex and challenging to fit.

The multilevel time-series models are applied to smooth the initial estimates, thereby reducing standard errors for monthly municipal figures by borrowing strenth over time and space and by using time series of claimant counts at the municipal level as auxiliary series in the model. The estimated models are used to make predictions for monthly provincial unemployment fractions by aggregating over the municipalities within provinces and quarterly municipal unemployment fractions by aggregating over the months within each calendar quarter.

The numbers of areas and time periods at the most detailed level considered, i.e. municipalities and months, are denoted by $M_A$ and $M_T$, respectively. For the description of the multilevel time-series model the initial estimates $\hat{\hat{Y}}_{it}$ are combined into a vector $\hat{\hat{Y}} = (\hat{\hat{Y}}_{11}, \hat{\hat{Y}}_{12}, \ldots, \hat{\hat{Y}}_{1M_T}, \hat{\hat{Y}}_{21}, \ldots \hat{\hat{Y}}_{2M_T}, \ldots \hat{\hat{Y}}_{M_A1}, \ldots \hat{\hat{Y}}_{M_AM_T})'$, where it is understood that empty domains are treated as missing.

The total length of $\hat{\hat{Y}}$ is therefore $M = M_A * M_T = 388(areas) * 66(months) = 25608$. Similarly, the variance estimates $v(\hat{\hat{Y}}_{it})$ are put in the same order along the diagonal of a $M \times M$ covariance matrix $\Phi$. The covariance matrix $\Phi$ is not diagonal because of the correlations induced by the panel design. It is a block diagonal matrix, where the blocks correspond to the municipalities and variances and covariances are estimated using (3) and (6), respectively.

The multilevel models considered for modeling the vector of direct estimates $\hat{\hat{Y}}$, take the general linear additive form

$$\hat{\hat{Y}} = X\beta + \sum_{\alpha} Z^{(\alpha)} v^{(\alpha)} + e\,, \tag{7}$$

where $X$ is a $M \times p$ design matrix for a $p$-vector of fixed effects $\beta$, and the $Z^{(\alpha)}$ are $M \times q^{(\alpha)}$ design matrices for $q^{(\alpha)}$-dimensional random effect vectors $v^{(\alpha)}$. Here the sum over $\alpha$ runs over several possible random effect terms at different levels, such as provincial-level smooth trends, municipal random intercepts, white noise at the municipality-by-month level, etc. This is explained in more detail in §4.2 below. The sampling errors $e = (e_{11}, e_{12}, \dots, e_{1M_T}, e_{21}, \dots M_A M_T)'$ are taken to be normally distributed as

$$e \sim N(0, \Sigma) \tag{8}$$

where either $\Sigma = \Phi = \bigoplus_{i=1}^{M_A} \Phi_i$ with $\Phi_i$ the covariance matrix for the GREG estimates for municipality $i$, or $\Sigma = \bigoplus_{i=1}^{M_A} \lambda_i \Phi_i$ in which $\lambda_i$ are municipality-specific variance scale parameters to be estimated. As described in Section 3 the design variances in $\Phi$ are pooled over municipalities and because of the discrete nature of the unemployment data they thereby lose some of their dependence on the unemployment level. It was found that incorporating the variance scale factors $\lambda_i$ allows the model to rescale the estimated design variances to a level that better fits the data. The variance scale parameters $\lambda_i$ are assigned independent inverse chi-squared priors with degree of freedom parameter and scale parameter equal to 1:

$$\lambda_i \sim \text{Inv}{-}\chi^2(1,1)\,, \quad \text{for} \quad i = 1, \dots M_A. \tag{9}$$

Equations (7) and (8) define the likelihood function

$$p\left(\hat{\hat{Y}} | \theta, \Sigma\right) = N\left(\hat{\hat{Y}} | \theta, \Sigma\right), \tag{10}$$

where $\theta = X\beta + \sum_{\alpha} Z^{(\alpha)} v^{(\alpha)}$ is the month-by-municipality vector of unemployment fractions from which all quantities of interest can be obtained by aggregation.

For the fixed effects part in (7) we use as a base model

$$period + prov * t + urb * cc, \tag{11}$$

where $period$ is the categorical variable for month (from 2012-04 to 2017-09), $t$ denotes quantitative time (in months), $prov$ is the categorical variable for the 12 provinces, $urb$ is the degree of urbanisation in 5 classes, and $cc$ is the fraction of claimant counts. The unemployment claimant counts are available at the month-by-municipality level, and are not the same as the registered unemployed variable used for GREG weighting.

The prior used for the $p$-vector of fixed effects is very weakly informative,

$$\beta \sim N(0, 100 I_p)\,, \tag{12}$$

with $I_p$ the identity matrix of order $p$. The standard error value of 10 is very large with regard to the scale of the unemployment fraction relative to the covariate scales.

## 4.2 Random effect model terms

The second term on the right hand side of (7) consists of a sum of contributions to the linear predictor by random effects or varying coefficient terms. The random effect vectors $v^{(\alpha)}$ for different $\alpha$ are assumed to be independent, but the components within a vector $v^{(\alpha)}$ are possibly correlated to accommodate temporal or spatial correlation. We refer to Fahrmeir et al. (2004), Zhao et al. (2006), Rue et al. (2009), as well as Hastie and Tibshirani (1990) and Wood (2017) for earlier and more extensive discussions of structured additive regression and related models. To describe the general model for each vector $v^{(\alpha)}$ of random effects, we suppress superscript $\alpha$ in what follows for notational convenience.

Each random effects vector $v$ is assumed to be distributed as

$$v \sim N(0, A \otimes V)\,, \tag{13}$$

where $V$ and $A$ are $d \times d$ and $l \times l$ covariance matrices, respectively, and $A \otimes V$ denotes the Kronecker product of $A$ with $V$. The total length of $v$ is therefore $q = dl$, and these coefficients may be thought of as corresponding to $d$ effects allowed to vary over $l$ levels of a factor variable, e.g. provincial intercepts ($d = 12$ provinces) varying over time ($l = 66$ months). The covariance matrix $V$ is allowed to be parameterized in three different ways. Most generally, it is an unstructured, i.e. fully parameterized covariance matrix. More parsimonious forms are $V = \mathrm{diag}(\sigma^2_{v;1}, \dots, \sigma^2_{v;d})$ or $V = \sigma_v^2 I_d$. If $d = 1$ the three parameterizations are equivalent. The covariance matrix $A$ describes the covariance structure between the levels of the factor variable, and is assumed to be known. It is typically more convenient to use the precision matrix $Q_A = A^{-1}$ as it is sparse for many common temporal and spatial correlation structures (Rue and Held, 2005). The sparsity of both $Q_A$ and the design matrix $Z$ is exploited in computations.

The following priors for the (hyper)parameters in $V$ are used:

- For a fully parameterized covariance matrix $V$ in (13) we use the scaled-inverse Wishart prior as proposed in O'Malley and Zaslavsky (2008) and recommended by Gelman and Hill (2007). Conditionally on a $d$-dimensional vector parameter $\xi$,

  $$V|\xi \sim \mathrm{Inv-Wishart}\left(V|\nu, \mathrm{diag}(\xi)\Psi\mathrm{diag}(\xi)\right) \tag{14}$$

  with $\nu = d + 1$, and $\Psi = I_d$. The vector $\xi$ is assigned a normal distribution $N(0, I_d)$.
- All other variance parameters appearing in a diagonal matrix $V$ in (13) are assigned, conditionally on an auxiliary parameter $\xi$, inverse chi-squared priors with 1 degree of freedom and scale parameter $\xi^2$. Each parameter $\xi$ is assigned a $N(0, 1)$ prior. Marginally, the standard deviation parameters have half-Cauchy priors. Gelman (2006) demonstrates that these priors are better default priors than the more common inverse gamma priors.

More specifically, the following random effect terms are considered for inclusion in (7):

1. Random intercepts for municipality. In this case $A = I_{M_A}$ and $V = \sigma_1^2$, and the corresponding design matrix is the $M \times M_A$ indicator matrix for municipalities. This can be readily extended to a vector of both municipal random intercepts and slopes for the fraction of claimant counts. In that case $V$ is a $2 \times 2$ covariance matrix, parameterized by variance parameters for the intercepts ($\sigma^2_{1,\mathrm{int}}$) and slopes ($\sigma^2_{1,\mathrm{sl}}$) and a correlation parameter ($\rho_1$). The $M \times (2M_A)$ design matrix then combines the municipality indicator columns for the intercepts and the same indicator columns multiplied by the fraction of claimant counts for the slopes.

2. Spatial municipality effects. Here $V = \sigma_2^2$, and the $M_A \times M_A$ precision matrix $Q_A$ has diagonal values equal to the number of neighbours $a_i$ of each municipality, and off-diagonal elements $-1$ for neighbouring municipalities and 0 elsewhere. The design matrix is the same as for the independent random municipality effects. The rank of $Q_A$ is $M_A - c$ where $c$ is the number of connected components, i.e. the number of clusters of municipalities that are not neighbouring any muncipality in another cluster. In the case of the adjacency structure of Dutch municipalities that we used there are four such disconnected clusters since some of the island municipalities share no border with the mainland municipalities. The singularity of $Q_A$ is dealt with by constraining the sum of each cluster's coefficients to be zero. This way the spatial coefficient of some islands is constrained to zero but this is not a big issue since independent municipal random effects are always included in the model. This spatial component corresponds to an intrinsic conditional autoregressive (ICAR) model for the coefficients (Besag and Kooperberg (1995), Rue and Held (2005)),
$$v_i | v_{j \neq i} \sim N\left(\frac{\sum_{j \in \mathrm{nb}(i)} v_j}{a_i}, \frac{\sigma_2^2}{a_i}\right) \qquad (15)$$
for each spatial municipality coefficient conditional on the others, $\mathrm{nb}(i)$ denoting the municipalities that neighbour municipality $i$. The combination (sum) of independent effects and spatial ICAR effects has been popularized by the Besag-York-Mollié model (Besag, York, and Mollie, Besag et al.). As in the case of the independent municipality effects, the spatial component is extended with random slopes for claimant counts. The matrix $V$ is then a $2 \times 2$ covariance matrix parameterized by $\sigma_{2,\mathrm{int}}^2$, $\sigma_{2,\mathrm{sl}}^2$ and $\rho_2$.
3. Provincial smooth trends over the months. The precision matrix in this case is the $M_T \times M_T$ band matrix (see e.g. Rue and Held (2005))
$$Q_A = \begin{pmatrix} 1 & -2 & 1 & & & & & \\ -2 & 5 & -4 & 1 & & & & \\ 1 & -4 & 6 & -4 & 1 & & & \\ & 1 & -4 & 6 & -4 & 1 & & \\ & & \ddots & \ddots & \ddots & & & \\ & & & 1 & -4 & 6 & -4 & 1 \\ & & & & 1 & -4 & 5 & -2 \\ & & & & & 1 & -2 & 1 \end{pmatrix}. \qquad (16)$$
A full covariance matrix for the trend innovations can be considered to allow for cross-sectional correlations, or a diagonal matrix to allow for different variance hyperparameters for the trend innovations. Both possibilities have been applied but tend to overfit the data. Therefore a common variance parameter for the 12 provincial trends is used in this application, i.e. $V = \sigma_3^2 I_{12}$. The design matrix is a $M \times (12M_T)$ indicator matrix for the province-month combinations. The precision matrix (16) has two singular vectors, $\iota_{M_T} = (1, 1, \dots, 1)$ and $(1, 2, \dots, M_T)'$. This means that the corresponding specification (13) is completely uninformative about the overall levels and linear trends by province. In order to prevent unidentifiability among various terms in the model, the provincial overall levels and linear trends are removed from $v$ by imposing the constraints $Rv = 0$, where $R = I_{12} \otimes R_{\mathrm{RW2}}(M_T)$ where $R_{\mathrm{RW2}}(M_T)$ is the $2 \times M_T$ matrix with the two singular vectors as its rows (Rue and Held, 2005). Note that the provincial overall levels and linear trends are included in the vector $\beta$ of fixed effects in (11).
4. Municipal random walks over quarters. Here $l = M_Q = 22$, the number of quarters. The

precision matrix is now the $M_Q \times M_Q$ band matrix

$$Q_A = \begin{pmatrix} 1 & -1 & & & & \\ -1 & 2 & -1 & & & \\ & -1 & 2 & -1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & -1 & 2 & -1 \\ & & & & -1 & 1 \end{pmatrix}, \qquad (17)$$

corresponding to a first-order random walk over quarters. For each municipality a random walk or local level trend is included, so $d = M_A$. Here also we use a common variance parameter for all municipal random walks, so $V = \sigma_4^2 I_{M_A}$. The design matrix is a $M \times (M_Q M_A)$ indicator matrix for the municipality-quarter combinations. Random walks varying by month would also have been possible, but have not been used since the focus is eventually on municipal estimates at a quarterly frequency. The random walk precision matrix has a single singular vector $R_{\text{RW1}}(M_Q) = \iota_{M_Q} = (1, 1, \dots, 1)$. Therefore constraints $Rv = 0$ are imposed where $R = I_{M_A} \otimes R_{\text{RW1}}(M_Q)$ to remove the municipal main effects from $v$. These main effects are represented instead by the municipal random effects.

5. A dynamic regression coefficient for the fraction of claimant counts. We consider both smooth and local level dynamic regression coefficients, so $Q_A$ is either (17) or (16), of dimension $M_T \times M_T$. In addition, $V = \sigma_5^2$. The design matrix equals the $M \times M_T$ matrix with each column equal to the fraction of claimant counts multiplied by the indicator variable for the corresponding month. Constraints $Rv = 0$ are imposed where $R$ is either $R_{\text{RW1}}(M_T)$ or $R_{\text{RW2}}(M_T)$ In the case of a smooth dynamic coefficient a linear time trend $cc * t$ is added to the fixed effects part (11) of the model.
6. White noise. In order to allow for unexplained variation white noise at the municipal-month level can be included. So $A = I_M$ and $V = \sigma_6^2$, and the design matrix is $I_M$.

In a small area estimation context, model (7) can be regarded as a generalization of the Fay-Herriot area-level model. The Fay-Herriot model only includes a single vector of uncorrelated random effects over the levels of a single factor variable (typically areas). The models used in this paper contain various combinations of uncorrelated and correlated random effects over two aggregation levels of time (month and quarter) and region (munipality and province). Earlier accounts of multilevel time-series models extending the Fay-Herriot model are Rao and Yu (1994); Datta et al. (1999); You (2008a). Datta et al. (1999) and You (2008a) use time-series models with independent area effects and first-order random walks over time for each area. In Rao and Yu (1994) a model is used with independent random area effects and a stationary autoregressive AR(1) instead of a random walk model over time. In You et al. (2003) the random walk model was found to fit the Canadian unemployment data slightly better than AR(1) models with autocorrelation parameter fixed at 0.5 or 0.75. Compared to the aforementioned references the models considered here contain more structured random effect terms corresponding to multiple temporal and spatial aggregation levels. Moreover, smooth trends are considered instead of or in addition to first-order random walks or autoregressive components, and more flexible ways of shrinkage are considered, as discussed next.

## 4.3 Non-normally distributed random effects

We also investigate a generalisation of (13) to non-normal distributions of random effects. There exists a vast literature on this topic, see for example Carter and Kohn (1996) in the state space modeling context, Datta and Lahiri (1995) and Fabrizi and Trivisano (2010) in the small area estimation context, and Lang et al. (2002) and Brezger et al. (2007) in the context of more general structured additive regression models. Here we connect to the more recent global-local shrinkage framework (Carvalho et al., 2010; Polson and Scott, 2010; Tang et al., 2018) and introduce additional local scale parameters associated with the precision matrix $Q_A$. In order to do so, first rewrite $Q_A = D'_A D_A$ where $D_A$ is an $l_D \times l$ matrix. For independent random effects $l_D = l$ and $D_A = I_l$. For the time-series components, $l_D$ corresponds to the number of innovations, which is $l - 1$ for a first-order random walk and $l - 2$ for a smooth trend. For a first-order random walk $D_A$ is the differencing matrix, which when acting on an $l$-vector of coefficients yields the $l - 1$ differences between subsequent elements. It has values $-1$ along the main diagonal and $1$ along the first superdiagonal, and otherwise $0$. Similarly, for a smooth trend $D_A$ is a second-order differencing matrix with values $1$ along the main diagonal, $-2$ along the first superdiagonal and $1$ along the second superdiagonal. For a spatial ICAR component, $D_A$ is the oriented incidence matrix corresponding to the graph defining the spatial neighbourhood structure. Here $l_D$ is the number of edges of the graph, i.e. the number of different neighbour pairs and is in this case much larger than $l$, the number of vertices, i.e. regions. Each row (edge) has a value $-1$ for the region (column) defining the start and a value $1$ for the region defining the end of the edge.

Now $Q_A = D'_A D_A$ is replaced by (see Lang et al. (2002))

$$\tilde{Q}_A = D'_A \mathrm{diag}(\omega_1, \dots \omega_{l_D})^{-1} D_A \tag{18}$$

where the $l_D$ parameters $\omega_k$ are local variance parameters. Several distributions can be considered for the components of $\omega$. We have tried

1. inverse-chi-squared distributions yielding t-distributed innovations
2. independent half-Cauchy distributions on the standard deviation parameters $\sqrt{\omega_k}$. This leads to the so-called horseshoe prior for the innovations.
3. exponential distributions. This gives rise to a double exponential or Laplace distribution for the innovations, and can be viewed as a Bayesian version of lasso shrinkage (Tibshirani, 1996; Park and Casella, 2008).

In the end we settled for the exponential prior as it seemed to work best in the current application.

## 4.4 Estimating the time-series multilevel model

The model is fit using Markov Chain Monte Carlo (MCMC) sampling, in particular the Gibbs sampler (Geman and Geman, 1984; Gelfand and Smith, 1990). The multilevel models considered belong to the class of additive latent Gaussian models with random effect terms being Gaussian Markov Random Fields (GMRFs), and we make use of the sparse matrix and block sampling techniques described in Rue and Held (2005) for efficiently fitting such models to the data. Besides that, the parameterization in terms of the above-mentioned auxiliary parameters $\xi$ (Gelman et al., 2008), greatly improves the convergence of the Gibbs sampler used. See Appendix A for more details on the Gibbs sampler used, including specifications of the full conditional distributions.

For each model considered, the Gibbs sampler is run in three independent chains with randomly generated starting values. Each chain is run for 2500 iterations. The first 500 draws are discarded as a "burn-in sample". From the remaining 2000 draws from each chain, we keep every second draw to save some memory while reducing the effect of autocorrelation between successive draws. This leaves $3 * 1000 = 3000$ draws to compute estimates and standard errors. The convergence of the MCMC simulation is assessed using trace and autocorrelation plots as well as the Gelman-Rubin potential scale reduction factor (Gelman and Rubin, 1992), which diagnoses the mixing of the chains. The potential scale reduction factors for the model parameters in all models considered are usually very close to 1, but always below 1.1. Also, the estimated Monte Carlo simulation errors (accounting for the remaining autocorrelation in the chains) are small compared to the posterior standard errors for all parameters, so that the number of retained draws is sufficient for our purposes.

The estimands of interest can be expressed as functions of the parameters, and applying these functions to the MCMC output for the parameters results in draws from the posteriors for these estimands. In this paper we summarize those draws in terms of their mean and standard deviation, serving as estimates and standard errors, respectively. All estimands considered can be expressed as linear predictors, i.e. as linear combinations of the model parameters. In particular, the MCMC simulation vectors of the municipality-by-month linear predictor $X\beta + \sum_\alpha Z^{(\alpha)} v^{(\alpha)}$ are computed. Afterwards, these are aggregated to the main estimation levels of interest: province-by-month and municipality-by-quarter.

# 5 Results

## 5.1 Measures for model assessment

Several multilevel models of the form (7) have been fitted to the data, and in order to assess the models we use different criteria. Frequently applied model selection criteria in hierarchical Bayesian settings are the Widely Applicable Information Criterion or Watanabe-Akaike Information Criterion (WAIC) (Watanabe, 2010, 2013) and the Deviance Information Criterion (DIC) (Spiegelhalter et al., 2002). They are popular because they are easy to compute from MCMC simulation output and because of their ability to make a reasonable trade-off between model fit and model complexity. The DIC is defined as

$$\begin{aligned} \mathrm{DIC} &= -2\left(\log p\left(\hat{\hat{Y}} \mid E_{\mathrm{post}}\theta\right) - p_{\mathrm{DIC}}\right), \\ p_{\mathrm{DIC}} &= 2\left(\log p\left(\hat{\hat{Y}}|E_{\mathrm{post}}\theta\right) - E_{\mathrm{post}} \log p\left(\hat{\hat{Y}}|\theta\right)\right), \end{aligned} \tag{19}$$

where $p\left(\hat{\hat{Y}}|E_{\mathrm{post}}\theta\right)$ is the log-likelihood (10) evaluated at the posterior mean $E_{\mathrm{post}}\theta$ of the model parameters, and $p_{\mathrm{DIC}}$ is an estimate of the effective number of model parameters. The effective number of model parameters is used as a penalty for model complexity and is closely related to the effective number of parameters proposed by Hodges and Sargent (2001) for linear multilevel models where each fixed effect contributes one degree of freedom and random effects contribute a value in the range between zero and the number of random effects, depending on the size of the variance component(s). Models with lower DIC values are preferred.

We use two WAIC measures differing by their estimate of the effective number of model parameters (Gelman et al., 2014),

$$\mathrm{WAIC}_k = -2\sum_{i=1}^{M} \log E_{\mathrm{post}} p\left(\hat{\hat{Y}}_i|\theta\right) + 2p_{\mathrm{WAIC},k}\,, \quad k = 1,2$$

$$p_{\mathrm{WAIC},1} = 2\left(\sum_{i=1}^{M}\left(\log E_{\mathrm{post}} p\left(\hat{\hat{Y}}_i|\theta\right) - E_{\mathrm{post}} \log p\left(\hat{\hat{Y}}_i|\theta\right)\right)\right), \tag{20}$$

$$p_{\mathrm{WAIC},2} = \sum_{i=1}^{M} \mathrm{var}_{\mathrm{post}}\left(p\left(\hat{\hat{Y}}_i|\theta\right)\right).$$

Here $p\left(\hat{\hat{Y}}_i|\theta\right)$ is the pointwise-likelihood for month-municipality combination $i$, and $\mathrm{var}_{\mathrm{post}}$ denotes the variance with respect to the posterior distribution for the model parameters. The WAIC is often seen as an improvement on the DIC, because $p_{\mathrm{WAIC},1}$ and $p_{\mathrm{WAIC},2}$ cannot become negative, unlike $p_{\mathrm{DIC}}$, and WAIC is also defined for singular models (Vehtari et al., 2017). Models with lower WAIC values are preferred.

Model adequacy is also evaluated using posterior predictive checks. This implies that draws from the posterior predictive distribution are simulated and compared with the originally observed data to study systematic discrepancies and to evaluate how well a model is able to reproduce certain aspects of the observed data (Gelman et al., 1996). Posterior predictive p-values are calculated for four different test statistics. They are defined as $p_k = \Pr(T_k(\tilde{\tilde{Y}}) \geq T_k(\hat{\hat{Y}})|\hat{\hat{Y}})$, where $T_k$ is a test statistic and $\tilde{\tilde{Y}}$ denotes replicate data based on the posterior predicitive distribution. The posterior predictive p-values are estimated from the MCMC output as the average over the $S$ Monte Carlo samples

$$\hat{p}_k = \frac{1}{S}\sum_{s=1}^{S} I\left(T_k(\tilde{\tilde{Y}}^S) \geq T_k(\hat{\hat{Y}})\right),$$

with $I(A)$ the indicator function with value one if the condition $A$ is fulfilled and zero otherwise. Ideally, a model should be able to reproduce important aspects of the observed data, which is the case if the observed $T_k(\hat{\hat{Y}})$ is in the bulk of the histogram of replicates $T_k(\tilde{\tilde{Y}}^S)$. Therefore p-values close to zero or one are indications of a poor fit regarding the test statistic. Posterior predictive p-values are computed for the following test statistics:

1. $T_1(y) = \frac{1}{M-1}\sum_{i=1}^{M}(y_i - \bar{y})^2$, the unweighted variance of the (replicate) data-vector. Here $\bar{y}$ is the mean of $y_i$.
2. $T_2(y) = \frac{1}{\sum_i N_i}\sum_{i=1}^{M} N_i(y_i - \bar{y})^2$, i.e. a weighted variance where the weights are the municipal population sizes $N_i$.
3. $T_3(y) = \bar{y}$, the unweighted mean of the (replicate) data-vector.
4. $T_4(y) = \frac{1}{\sum_i N_i}\sum_{i=1}^{M} N_i y_i$, i.e. a weighted mean where the weights are the municipal population sizes $N_i$.

Finally, we define the following two discrepancy measures to evaluate and compare the time-series multilevel models, as well as compare them with the quarterly cross-sectional estimates based on a unit-level small area estimation model. The first measure is the Mean Relative Bias (MRB) and summarizes the differences between model estimates and direct estimates averaged over time, as percentage of the latter. For a given model $M$, the $MRB_i$ is defined as

$$MRB_i = \frac{\sum_t(\hat{\theta}_{it}^{\mathrm{M}} - \hat{\theta}_{it}^{\mathrm{direct}})}{\sum_t \hat{\theta}_{it}^{\mathrm{direct}}} \times 100\%\,. \tag{21}$$

This benchmark measure shows for each region $i$ how much the model-based estimates deviate from the direct estimates. The discrepancies should not be too large as one may expect that the direct estimates averaged over time are close to the true average level of unemployment. The second discrepancy measure is the Relative Reduction of the Standard Errors (RRSE) and measures the percentages of reduction in standard error of the model-based compared to the direct estimates, i.e.,

$$RRSE_i = 100\% \times \frac{1}{M_T} \sum_t (se(\hat{\theta}_{it}^{\text{direct}}) - se(\hat{\theta}_{it}^{\text{M}}))/se(\hat{\theta}_{it}^{\text{direct}}), \quad (22)$$

for a given model $M$. Both measures are evaluated at the two regional levels of provinces and municipalities. Note however that for the RRSE measure it can make a difference whether the time-averaging in 22 is over months or quarters. For the provincial estimates both will be considered.

## 5.2 Model results

In this section results for eleven different models are presented. An overview of the models is given in Table 5.1. See Section 4 for the description of the general model form, and in particular Subsection 4.2 for a detailed description of the random effect model components used. Table 5.2 shows the posterior means of the variance components of these random effect terms under the models considered.

The first three, baseline, models are abbreviated as BLM1, BLM2 and BLM3. All three models have the same base fixed effects model part defined by (11) and (12). The variable $period$ is included because the monthly national-level aggregates of the direct estimates correspond to the officially published labour force figures obtained themselves using a time-series model, and so any further smoothing over time at the national level is not desired. The component $prov * t$ is a linear trend for seperate provinces and is part of the smooth trend model for provinces discussed below. The component $urb * cc$ corresponds to the fraction of claimant counts for each of five degrees of urbanisation. The models BLM1, BLM2 and BLM3 share the same trend model. This consists of separate smooth trend models for the provinces with a scalar variance structure, denoted RW2P. This implies that the innovations for the trend share the same variance hyperparameter and no correlations between provinces are assumed. All municipalities belonging to the same province share the same provincial smooth trend. Deviations from this overall trend are modeled with municipality-specific local level trends, denoted RW1M. A scalar variance structure is assumed also for the innovations of the municipality trends. This is a parsimoneous alternative for a trend model with a full covariance matrix for the trend innovations of municipalities as the provincial trends induce a positive correlation between municipal trends within the same province (Boonstra and van den Brakel, 2016). The model is used to compute monthly provincial unemployment figures and quarterly municipal unemployment figures. Therefore the provincial trends are modelled at a monthly frequency and the municipal trends are modelled at a quarterly frequency. This implies that the provincial smooth trends generate a new innovation each month, while the municipal local level trends produce only one new innovation for the three months within each quarter. Finally the overall municipal levels are represented in each model as random intercepts.

The first model BLM1 uses $\Sigma = \Phi = \bigoplus_i \Phi_i$ in (8), i.e. the matrix with estimated design-variances and covariances pooled over municipalities, as discussed in Section 3. Model BLM2 is the same but uses instead $\Sigma = \bigoplus_i \lambda_i \Phi_i$ for the survey errors. This covariance matrix is parameterized in terms of municipal scale factors $\lambda_i$, and turns out to fit the data better, as can be seen from the information

| Model name | fixed effects | random effects | sampling error variance |
|---|---|---|---|
| BLM1 | $period + prov * t + urb * cc$ | MI + RW2P + RW1M | no scaling |
| BLM2 | $period + prov * t + urb * cc$ | MI + RW2P + RW1M | scaling |
| BLM3 | $period : vn + prov * t + urb * cc$ | MI + RW2P + RW1M | scaling |
| **Model name** | **Description** | | |
| DR1 | BLM3 + local level dynamic regression coefficient for $cc$ | | |
| DR2 | BLM3 + smooth dynamic regression coefficient for $cc$ | | |
| DR2RS | DR2 + random slopes for $cc$ at municipal level | | |
| GL1 | BLM3 with Laplace prior for municipal random effects | | |
| DR2RSL | DR2RS with Laplace priors for municipal random effects and dynamic regression coeff. | | |
| SPDR2RS | DR2RS + spatial random municipal intercepts and slopes for $cc$ | | |
| WN | BLM3 + white noise | | |
| WNSPDR2RSL | SPDR2RS with the Laplace priors of DR2RSL + white noise | | |

**Table 5.1 Summary of the time-series models considered. The abbreviations for the random effects terms of the baseline models BLM1, BLM2, BLM3 stand for municipal random intercepts (MI), smooth provincial trends (RW2P) and local level municipal trends with quarterly innovations (RW1M).**

| | $\sigma_{1,\mathrm{int}}$ | $\sigma_{1,\mathrm{sl}}$ | $\rho_1$ | $\sigma_{2,\mathrm{int}}$ | $\sigma_{2,\mathrm{sl}}$ | $\rho_2$ | $\sigma_3$ | $\sigma_4$ | $\sigma_5$ | $\sigma_6$ |
|---|---|---|---|---|---|---|---|---|---|---|
| BLM1 | 0.41 | | | | | | 0.005 | 0.056 | | |
| BLM2 | 0.46 | | | | | | 0.003 | 0.122 | | |
| BLM3 | 0.25 | | | | | | 0.004 | 0.132 | | |
| DR1 | 0.25 | | | | | | 0.003 | 0.124 | 3.0 | |
| DR2 | 0.25 | | | | | | 0.003 | 0.125 | 0.45 | |
| DR2RS | 0.16 | 8.0 | -16 | | | | 0.003 | 0.126 | 0.45 | |
| GL1 | 0.27 | | | | | | 0.004 | 0.128 | | |
| DR2RSL | 0.17 | 8.4 | -10 | | | | 0.003 | 0.123 | 0.43 | |
| SPDR2RS | 0.12 | 4.9 | -19 | 0.22 | 9.7 | -12.0 | 0.003 | 0.127 | 0.50 | |
| WN | 0.26 | | | | | | 0.004 | 0.147 | | 0.61 |
| WNSPDR2RSL | 0.14 | 6.7 | -18 | 0.49 | 20.0 | -7.8 | 0.003 | 0.139 | 0.46 | 0.61 |

**Table 5.2 Posterior means ($\times 100$) of variance components under the models considered. The parameter names correspond to the random effect model components introduced in Section 4.2.**

criteria in Table 5.3. Figure 5.1 shows that the estimated scale factors are almost linearly related to the unemployment fraction, and therefore they restore the dependence between estimate and variance that was removed by pooling the design variances over municipalities, which was done because of the large sampling variances of the estimated design variances themselves.

The posterior predictive p-values in Table 5.4 also show a defect of model BLM1: it takes the value 1 for the variance (weighted or unweighted), meaning that predicted data based on model BLM1 has too much variablility compared to the original data. Model BLM2 is somewhat better in this respect (p-values around 0.8), but it has very low p-values for the means of the estimates, meaning that predicted data based on model BLM2 tend to be negatively biased. This is clear also from Figure 5.2, which shows the monthly national aggregates for direct estimates and model estimates based on BLM2 and BLM3.

The bias of BLM2 at the national level is clearly undesirable. Model BLM3 removes this bias, at least approximately, by exploiting the internal benchmarking approach described in Bell et al. (2013). Loosely speaking, by including fixed effects for the columns of $\Sigma W$ for a given $M \times q$ matrix $W$ in the

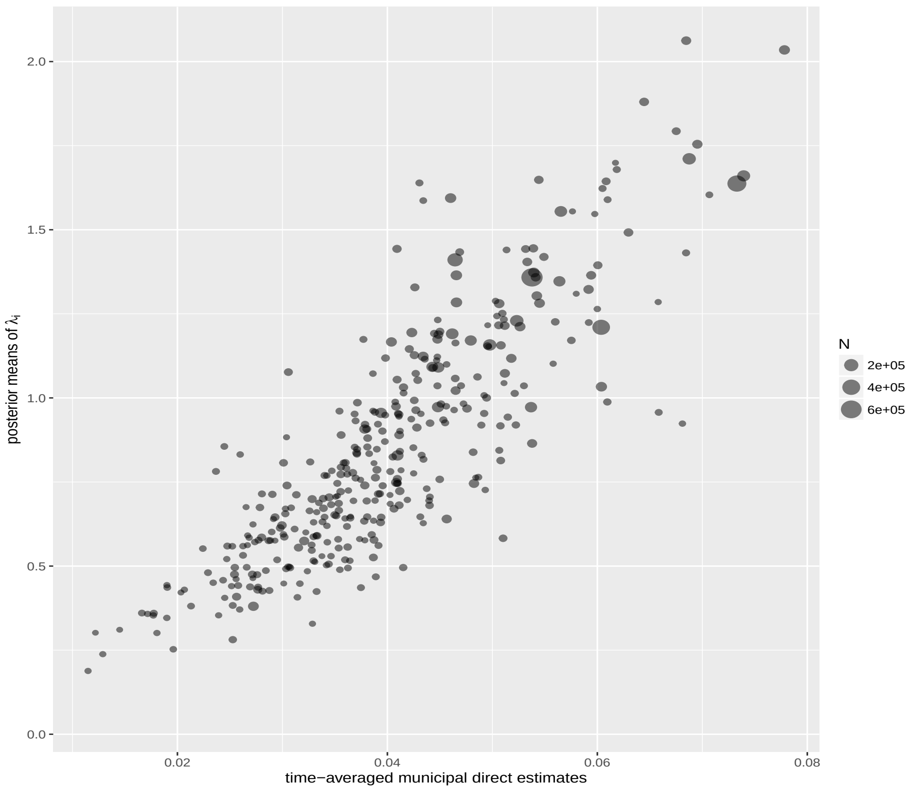


**Figure 5.1 Posterior means of the variance factors $\lambda_i$ of model BLM2, versus the time-averages of the direct estimates for the municipalities. The dot size relates to the population size of a municipality.**

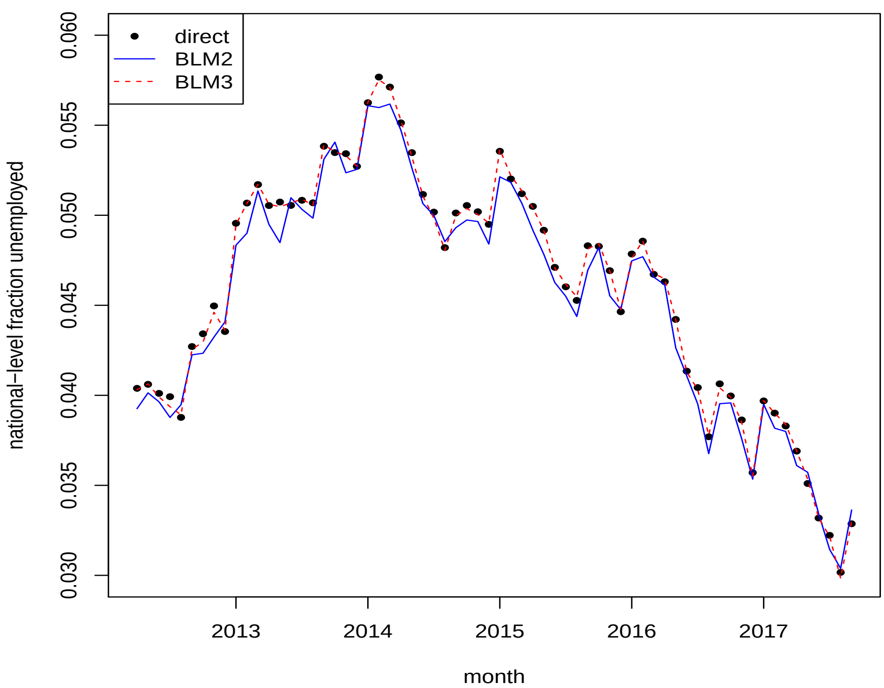


**Figure 5.2 Monthly time-series of national aggregates of direct estimates (black dots) and model estimates based on models BLM2 (blue line) and BLM3 (red dashed line).**

| | DIC | pDIC | WAIC1 | pWAIC1 | WAIC2 | pWAIC2 | mean llh |
|---|---|---|---|---|---|---|---|
| BLM1 | -102522 | 256 | -99967 | 508 | -99935 | 524 | 51389 |
| BLM2 | -104676 | 755 | -100903 | 1519 | -100537 | 1703 | 52716 |
| BLM3 | -104719 | 686 | -101091 | 1328 | -100774 | 1486 | 52702 |
| DR1 | -104716 | 682 | -101094 | 1317 | -100777 | 1476 | 52699 |
| DR2 | -104720 | 678 | -101097 | 1314 | -100781 | 1472 | 52699 |
| DR2RS | -104720 | 683 | -101103 | 1326 | -100788 | 1483 | 52702 |
| GL1 | -104730 | 690 | -101102 | 1342 | -100781 | 1502 | 52710 |
| DR2RSL | -104734 | 683 | -101113 | 1335 | -100795 | 1494 | 52709 |
| SPDR2RS | -104729 | 684 | -101115 | 1329 | -100798 | 1487 | 52707 |
| WN | -104895 | 2126 | -101226 | 2638 | -100352 | 3075 | 53511 |
| WNSPDR2RSL | -104923 | 2168 | -101270 | 2689 | -100383 | 3133 | 53546 |

**Table 5.3 Information criteria and effective numbers of model parameters for all models: DIC and two versions of WAIC. The last column displays the posterior mean of the log-likelihood.**

| | var | weighted var | mean | weighted mean |
|---|---|---|---|---|
| BLM1 | 1.00 | 1.00 | 0.82 | 0.53 |
| BLM2 | 0.83 | 0.80 | 0.12 | 0.05 |
| BLM3 | 0.87 | 0.80 | 0.77 | 0.45 |
| DR1 | 0.87 | 0.79 | 0.76 | 0.46 |
| DR2 | 0.86 | 0.82 | 0.75 | 0.44 |
| DR2RS | 0.87 | 0.83 | 0.74 | 0.46 |
| GL1 | 0.87 | 0.82 | 0.74 | 0.45 |
| DR2RSL | 0.86 | 0.85 | 0.71 | 0.46 |
| SPDR2RS | 0.90 | 0.83 | 0.71 | 0.47 |
| WN | 0.79 | 0.72 | 0.74 | 0.47 |
| WNSPDR2RSL | 0.80 | 0.75 | 0.71 | 0.47 |

**Table 5.4 Posterior predictive p-values for mean and variance of the estimates, both unweighted and weighted by municipal population sizes.**

model, the model predictions $\hat{\theta}$ satisfy the benchmark conditions $W'\hat{\theta} = W'\hat{\hat{Y}}$. Choosing $W$ to be the $M \times M_T$ indicator-matrix for month with each column multiplied by the $M$-vector $N$ of population sizes we get the desired benchmark restrictions. For the $\Sigma$ in the predictor matrix $\Sigma W$ a plug-in estimate $\hat{\Sigma} = \bigoplus_i \hat{\lambda}_i \Phi_i$ is used, where $\hat{\lambda}_i$ are the posterior means obtained from model BLM2. To include this in the fixed effects part of the model, define the quantitative covariate $vn$ to be the vector $\Sigma N$, and replace $period$ in (11) by $period : vn$, where the ':' indicates that no main effects for $period$ are included. (We are using the R language (R Core Team, 2018) conventions for model formulae). The variable $vn$ turned out to contain for some of the smallest municipalities and a few months several outlying values, which have been removed by cutting off all values below 20 and above 200. This prevents a few unrealistic predictions for these month-municipality combinations, but otherwise has little effect. Note that using $vn$ as a covariate also means that for prediction for non-observed month-municipality combinations a value for this variable must be available. In these cases a variance estimate of 1 has been used for computing $vn$, which is relatively large compared to the variance estimates for non-missing month-municipality combinations. Figure 5.2 shows that the national-level bias is indeed almost non-existent for BLM3. There are only slight discrepancies due to using the above-mentioned cut-off values, using the plug-in estimate for $\lambda_i$ from model BLM2, and more generally Monte Carlo error of the MCMC estimates. Model BLM1 has only slightly larger national-level bias than model BLM3. This is not shown in the figure, but is consistent with the posterior predictive p-value close to 0.5 for the weighted mean in Table 5.4. BLM1 also approximates the internal benchmark condition, since $\Sigma N$ for BLM1 is well-represented by the fixed effects term $period$, due to the fact that pooled variances are used and sampling fractions are equal over municipalities.

Based on the above comparison, model BLM3 is the best of the baseline models. We next compare different extensions of BLM3. The first extension is to allow for a time-varying regression coefficient for the fraction of claimant counts at the national level. This is realised by including random effects term 5 described in Section 4.2 with either first or second-order random walk innovations. These models are abbreviated as DR1 and DR2, respectively. Figure 5.3 shows the estimated time trends of both models. They are composed of the posterior means of the main effect for claimant counts plus the varying effects, and in case of the second-order random walk also the linear time trend fixed effect. In terms of the model-fit measures the dynamic regression coefficient does not change much, except that the WAIC measures suggest that especially the smooth dynamic regression coefficient of model DR2 yields a small improvement over model BLM3.

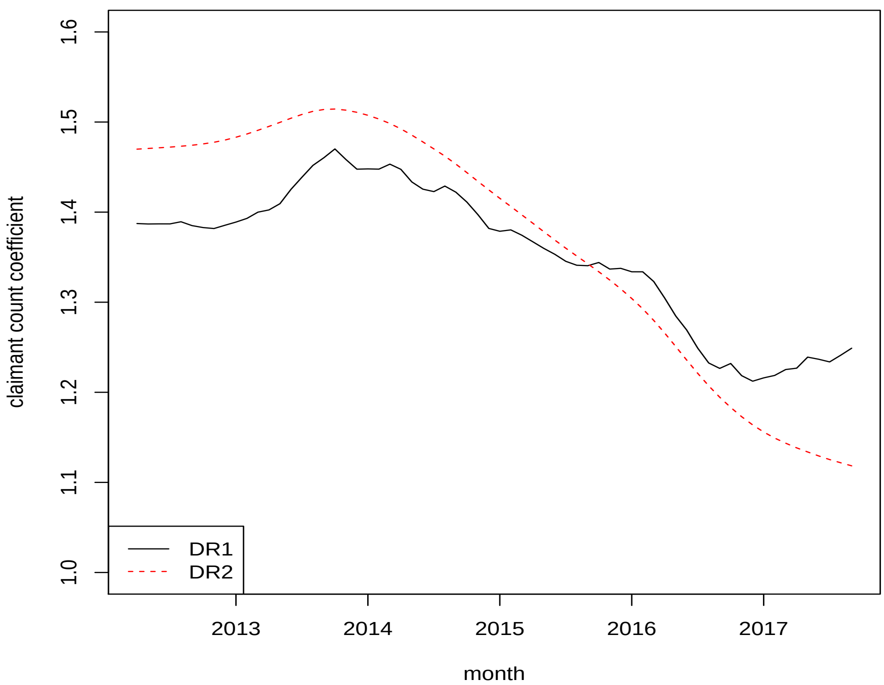


**Figure 5.3 Posterior mean of the regression coefficient for claimant counts varying over month, based on first-order (black) and second-order (red, dashed) random walks.**

The next model considered is a further extension of DR2 with claimant count slopes varying by municipality. A full $2 \times 2$ covariance matrix is used to model the variances of and correlation between municipal slopes and intercepts, as described under random effect term number 1 in Section 4.2. This model is abbreviated as DR2RS. Its WAIC values are slightly smaller (better) than for model DR2, but otherwise the model assessment measures are similar.

Another modification considered is the use of non-normal priors for random effects. Model GL1 is model BLM3 where the normal distribution for the municipality random effects is replaced by a Laplace distribution. The Laplace distribution places more mass around zero and has fatter tails than the normal distribution. This is illustrated by Figure 5.4 where the posterior means of the municipal random intercepts are displayed for models BLM3 and GL1. The Laplace prior hardly shrinks small municipality intercepts further to zero, but does allow a few more extreme values. The smallest value corresponds to a very small municipality with small monthly sample sizes, which happens to have no observations of unemployed persons in the study period. The two largest values correspond to two large cities Groningen and Rotterdam. We have also tried to use a horseshoe prior, which is more extreme in the sense that it places even more mass around zero and has even fatter tails than the Laplace distribution, and is very suitable in a sparse situation where many effects are in fact zero. As mentioned in Section 4.2 the horseshoe prior did not work as well as the Laplace prior, perhaps because most random municipality effects are actually nonzero. According to the information criteria, the Laplace priors of model GL1 yield a small improvement over model BLM3, see Table 5.3.

The next model listed in Table 5.1, DR2RSL, combines model DR2RS with the Laplace global local prior of model GL1. The Laplace prior is used both for the municipal random intercepts and slopes and for the innovations of the smooth dynamic regression coefficient for claimant counts. From Table 5.3 it can be seen that the two small improvements of models DR2RS and GL1 over the baseline model BLM3 result in a larger improvement of DR2RSL regarding the information criteria.

Another potential model improvement is to account for the spatial neighbourhood structure of municipalities, by including spatial correlations over municipalities in addition to the clustering of municipalities within provinces already implicitly accounted for in the models. Model SPDR2RS is an extension of model DR2RS in which besides the unstructured (iid) municipal random intercepts and slopes for claimant counts also the same effects with spatial ICAR correlation are included. The posterior mean estimates of these four types of effects under this model are displayed in Figure 5.5 (plots have been made using R package `tmap` (Tennekes, 2018)). Compared to model DR2RS, this model yields a modest improvement regarding the information criteria of about 10 units.

The models considered thus far do not contain a white noise term, i.e. an unstructured random effects term at the data-level of municipality-by-month. Such a term might capture some remaining unstructured contributions to the signal (the linear predictor), instead of perhaps attributing them to the sampling noise. To study this, we first consider model WN, which is model BLM3 extended by a white noise term. Table 5.3 shows that the white noise term results in a large increase of the effective number of model parameters as estimated by $p_{\text{DIC}}$, $p_{\text{WAIC},1}$ and $p_{\text{WAIC},2}$. At the same time the model fit becomes much better. However, the information criteria considered disagree about the trade-off between model-fit and model complexity. According to DIC and WAIC1, WN is a much better model, whereas according to WAIC2 it is much worse than BLM3. The discrepancy between WAIC1 and WAIC2 is due to the large difference in the estimated effective number of model parameters, but it is not clear to us which estimate is to be preferred in this case. Looking at the other model evaluation measures, we see from Table 5.4 that adding white noise yields posterior predictive p-values for the variance statistics a little closer to 0.5, implying that the spread of data generated by model WN is closer to that of the observed data. A more notable difference is that the RRSE measures at both provincial and municipal levels (Tables 5.6 and 5.8) become smaller. So the estimated standard errors under model WN are larger than under model BLM3, which seems reasonable. The white noise term decreases the risk of underfitting and obtaining over-optimistic standard errors. We have also tried to include an unstructured random effects term at the province-by-month level, but this did not seem to be an improvement and had no noticeable effect on the RRSE values. Figure 5.6 shows the small but noticeable difference between models BLM3 and WN regarding their flexibility; the latter tends to follow the direct estimates more closely.

The last model considered combines all model extensions that have been found to be improvements: the dynamic smooth regression coefficient and random slopes for claimant counts with Laplace priors, spatial municipal intercepts and slopes with Laplace priors, and white noise. This is the model with lowest DIC and WAIC1. Its WAIC2 is lower than that of WN, but still high compared to the other models. This is due to the white noise term, which has an adversarial effect on WAIC2 but otherwise seems to be an improvement. So we choose this model as the model to base inferences on.

Figures 5.7 and 5.8 show the monthly provincial estimates based on the chosen model, and compare them to the direct GREG estimates and the cross-sectional SAE estimates. Note that the latter are only available per quarter. It is clear that the estimates based on the time-series multilevel model capture the mean level over time of the GREG estimates better than do the cross-sectional estimates. The difference is especially large for the provinces Groningen, Friesland and Flevoland, where the cross-sectional estimates are negatively biased relative to the GREG estimates, and the province Zeeland where the cross-sectional estimates are positively biased relative to the GREG estimates. This is also apparent from Table 5.5, which contains the mean relative bias measures defined in (21), computed at the provincial level. It can also be seen that the estimated standard errors based on the time-series multilevel model are smaller than the standard errors based on the cross-sectional model, even though the latter refer to quarters. This is in agreement with Table 5.6, which contains the percentage reduction of standard errors relative to those of the direct estimates at both quarterly and monthly levels. Table 5.6 demonstrates that the gain relative to the direct estimates is larger at the more disaggregated monthly level, as one would expect.

| | GR | FR | DR | OV | FL | GD | UT | NH | ZH | ZL | NB | LB |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CS-SAE | -9.6 | -10.7 | -4.2 | -0.4 | -9.9 | 0.2 | 2.3 | 0.9 | 0.5 | 12.8 | 1.6 | -1.5 |
| BLM1 | -1.0 | 0.5 | -0.8 | 0.3 | 1.3 | 0.3 | -0.5 | 0.1 | 0.1 | -0.2 | -0.4 | 0.9 |
| BLM2 | -3.2 | -3.0 | -1.6 | -0.7 | 0.7 | -1.5 | -0.8 | -1.0 | -1.8 | -0.7 | -1.2 | -0.5 |
| BLM3 | -2.1 | -1.1 | 0.3 | 1.2 | 1.3 | -0.2 | 0.3 | 0.2 | -0.9 | 1.2 | -0.1 | 0.9 |
| DR1 | -1.9 | -1.1 | 0.3 | 1.2 | 1.4 | -0.2 | 0.2 | 0.3 | -0.9 | 1.3 | -0.1 | 1.0 |
| DR2 | -2.0 | -1.1 | 0.3 | 1.1 | 1.3 | -0.2 | 0.2 | 0.3 | -0.9 | 1.4 | -0.1 | 1.0 |
| DR2RS | -2.1 | -1.0 | 0.2 | 1.1 | 1.6 | -0.2 | 0.2 | 0.3 | -0.8 | 1.1 | -0.1 | 0.9 |
| GL1 | -1.8 | -1.1 | 0.2 | 1.1 | 1.3 | -0.1 | 0.3 | 0.2 | -0.8 | 1.1 | -0.1 | 0.9 |
| DR2RSL | -1.8 | -0.8 | 0.1 | 1.0 | 1.5 | -0.1 | 0.2 | 0.3 | -0.8 | 1.3 | -0.1 | 0.8 |
| SPDR2RS | -2.1 | -1.2 | 0.3 | 1.0 | 1.4 | -0.1 | 0.4 | 0.2 | -0.8 | 1.4 | 0.0 | 0.8 |
| WN | -1.3 | -1.1 | 0.1 | 1.1 | 1.5 | -0.1 | -0.0 | 0.0 | -0.5 | 1.2 | -0.3 | 0.9 |
| WNSPDR2RSL | -1.3 | -0.9 | 0.2 | 1.0 | 1.5 | -0.1 | -0.0 | 0.1 | -0.4 | 1.2 | -0.3 | 0.7 |

**Table 5.5 Mean relative biases (in %) as defined in (21) for the cross-sectional SAE estimates (first line) and the different time-series models, at the provincial level.**

Similar conclusions can be drawn from Tables 5.7 and 5.8 concerning the municipal-level estimates: the (mean over municipalities of the) mean relative bias of the time-series model estimates is clearly smaller than that for the cross-sectional SAE model. The estimated standard errors are also smaller under the time-series model compared to the cross-sectional model, though the differences are not as large as in the provincial-level case. Figure 5.9 shows that the cross-sectional SAE method shrinks the estimates more towards the mean than does the time-series model. Note that the figure only displays time-averages. Naturally, the bias-variance trade-off is made differently by the time-series model because it can make use of more information over time.

We are primarily interested in quarterly municipal estimates even though the time-series models considered also provide monthly estimates. Figures 5.10 and 5.11 compare the quarterly estimates

| | GR | FR | DR | OV | FL | GD | UT | NH | ZH | ZL | NB | LB |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| quarterly | | | | | | | | | | | | |
| CS-SAE | 31.7 | 34.0 | 38.4 | 27.8 | 35.9 | 18.2 | 23.7 | 12.0 | 7.9 | 40.0 | 15.2 | 24.9 |
| BLM1 | 50.2 | 49.6 | 52.9 | 47.8 | 51.9 | 41.2 | 45.7 | 38.9 | 35.2 | 53.3 | 39.6 | 45.8 |
| BLM2 | 49.6 | 51.1 | 55.1 | 51.3 | 46.6 | 47.1 | 48.6 | 41.5 | 37.7 | 63.0 | 45.7 | 50.4 |
| BLM3 | 46.5 | 49.5 | 54.4 | 48.8 | 44.3 | 46.7 | 48.3 | 36.5 | 28.8 | 63.4 | 45.3 | 49.8 |
| DR1 | 47.2 | 50.2 | 54.4 | 49.6 | 45.1 | 47.9 | 48.7 | 36.8 | 28.9 | 63.8 | 45.9 | 50.7 |
| DR2 | 47.6 | 50.8 | 55.1 | 50.1 | 44.5 | 48.1 | 49.1 | 37.2 | 29.8 | 64.3 | 46.1 | 50.8 |
| DR2RS | 47.8 | 50.7 | 55.4 | 50.4 | 45.2 | 48.6 | 49.0 | 37.5 | 29.7 | 64.5 | 46.6 | 51.2 |
| GL1 | 46.6 | 50.2 | 54.0 | 48.5 | 43.3 | 46.7 | 47.9 | 36.3 | 29.2 | 63.1 | 45.3 | 49.7 |
| DR2RSL | 48.6 | 50.6 | 55.4 | 50.3 | 44.6 | 48.6 | 49.6 | 37.2 | 30.1 | 64.6 | 46.4 | 50.9 |
| SPDR2RS | 47.6 | 50.7 | 54.5 | 50.2 | 44.5 | 48.5 | 48.7 | 37.4 | 29.5 | 64.2 | 46.1 | 50.7 |
| WN | 42.2 | 46.5 | 50.7 | 46.2 | 38.4 | 44.5 | 44.3 | 37.0 | 30.1 | 60.3 | 43.3 | 46.4 |
| WNSPDR2RSL | 44.0 | 47.8 | 51.4 | 47.0 | 39.0 | 46.1 | 45.5 | 37.8 | 30.6 | 61.6 | 43.7 | 47.3 |
| monthly | | | | | | | | | | | | |
| BLM1 | 67.1 | 66.8 | 69.3 | 63.2 | 69.7 | 55.6 | 61.3 | 52.6 | 48.1 | 70.5 | 53.2 | 61.8 |
| BLM2 | 66.8 | 68.7 | 70.9 | 65.5 | 67.1 | 59.5 | 63.6 | 54.9 | 50.4 | 76.2 | 57.0 | 64.9 |
| BLM3 | 64.1 | 65.7 | 70.2 | 63.7 | 63.0 | 60.4 | 63.3 | 49.0 | 39.9 | 77.3 | 58.1 | 64.8 |
| DR1 | 64.5 | 66.1 | 70.1 | 64.0 | 63.3 | 60.9 | 63.6 | 49.1 | 39.8 | 77.5 | 58.3 | 65.1 |
| DR2 | 64.7 | 66.4 | 70.6 | 64.3 | 63.2 | 61.1 | 63.8 | 49.4 | 40.4 | 77.8 | 58.5 | 65.3 |
| DR2RS | 64.7 | 66.3 | 70.7 | 64.4 | 63.5 | 61.2 | 63.7 | 49.4 | 40.2 | 77.9 | 58.7 | 65.5 |
| GL1 | 64.1 | 66.0 | 70.0 | 63.6 | 62.5 | 60.4 | 63.2 | 48.9 | 40.0 | 77.1 | 58.1 | 64.7 |
| DR2RSL | 65.1 | 66.2 | 70.7 | 64.3 | 63.1 | 61.3 | 64.0 | 49.3 | 40.4 | 78.0 | 58.6 | 65.3 |
| SPDR2RS | 64.6 | 66.3 | 70.2 | 64.3 | 63.1 | 61.2 | 63.6 | 49.4 | 40.2 | 77.8 | 58.5 | 65.2 |
| WN | 56.0 | 60.4 | 63.7 | 57.7 | 53.1 | 55.3 | 54.7 | 44.3 | 37.5 | 71.8 | 53.4 | 59.0 |
| WNSPDR2RSL | 56.4 | 60.7 | 63.9 | 58.0 | 53.4 | 56.3 | 55.2 | 44.6 | 38.1 | 72.6 | 53.6 | 59.6 |

**Table 5.6 Mean RRSE (in %) for quarterly and monthly municipal model-based estimates. The first line contains the measures for the cross-sectional SAE estimates, which are only available quarterly. For the time-series multilevel models the measures are given at both temporal aggregation levels.**

based on the selected time-series model with the direct estimates and estimates based on the cross-sectional SAE model for a selection of municipalities ordered by decreasing population size. For the largest municipalities (Figure 5.10) the model-based estimates are generally close to the direct estimates. There are some systematic differences for example for Groningen where the model-based estimates are mostly below the direct estimates. This is the case also for the time-series model despite the fact that for Groningen the random intercept and slope for claimant counts are large and positive, and the random effects distribution is Laplace. For some municipalities systematic differences between time-series and cross-sectional model-based estimates are visible, for example for Enschede, Roermond (Figure 5.10) and Appingedam (Figure 5.11). In most but not all of such cases the time-series model estimates are closer to the average of the direct estimates. The last municipality displayed in Figure 5.11 is Schiermonnikoog, an island with fewer than 1000 inhabitants. The information coming from the direct estimates is very sparse in this case: the sample sizes are very small and over the period considered no unemployed have been among the respondents, and in some months there has been no response at all. It is also a municipality with a strong seasonal pattern in the claimant count register data. This pattern is also visible in the model-based estimates, especially the time-series ones.

| | Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|---|---|---|---|---|---|---|
| CS-SAE | -47.4 | -8.7 | 1.6 | 7.7 | 16.8 | 213.0 |
| BLM1 | -39.4 | -5.2 | 1.2 | 5.7 | 10.8 | 186.3 |
| BLM2 | -43.3 | -6.2 | 0.2 | 2.1 | 6.6 | 166.7 |
| BLM3 | -31.4 | -6.1 | 1.0 | 4.1 | 9.5 | 166.6 |
| DR1 | -30.9 | -6.1 | 1.1 | 4.2 | 9.7 | 166.8 |
| DR2 | -31.3 | -6.1 | 1.2 | 4.1 | 9.4 | 167.9 |
| DR2RS | -32.3 | -6.3 | 1.0 | 4.1 | 8.8 | 165.9 |
| GL1 | -30.0 | -6.1 | 1.1 | 4.1 | 9.3 | 166.5 |
| DR2RSL | -31.3 | -6.1 | 0.9 | 4.0 | 9.0 | 164.4 |
| SPDR2RS | -33.2 | -5.5 | 1.1 | 3.9 | 8.2 | 169.5 |
| WN | -30.2 | -5.8 | 1.0 | 4.0 | 9.0 | 165.5 |
| WNSPDR2RSL | -32.0 | -5.2 | 0.9 | 3.8 | 8.2 | 169.3 |

**Table 5.7 Summary statistics of mean relative biases (in %) for municipalities, for the cross-sectional SAE estimates (first line) and the different time-series models.**

| | Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|---|---|---|---|---|---|---|
| CS-SAE | 20.4 | 68.3 | 74.5 | 72.4 | 78.8 | 91.3 |
| BLM1 | 53.8 | 77.5 | 81.4 | 80.4 | 84.5 | 93.1 |
| BLM2 | 43.8 | 75.2 | 79.8 | 78.2 | 83.0 | 92.7 |
| BLM3 | 37.6 | 78.8 | 83.3 | 81.3 | 86.3 | 94.1 |
| DR1 | 36.9 | 79.2 | 83.7 | 81.6 | 86.5 | 94.2 |
| DR2 | 38.3 | 79.2 | 83.7 | 81.6 | 86.4 | 94.2 |
| DR2RS | 37.2 | 78.7 | 83.4 | 81.3 | 86.4 | 94.2 |
| GL1 | 35.1 | 78.7 | 83.2 | 81.1 | 86.1 | 94.0 |
| DR2RSL | 35.2 | 78.5 | 83.4 | 81.2 | 86.3 | 93.7 |
| SPDR2RS | 37.6 | 78.9 | 83.5 | 81.3 | 86.3 | 94.3 |
| WN | 25.6 | 70.4 | 76.8 | 74.2 | 80.8 | 92.4 |
| WNSPDR2RSL | 24.4 | 70.5 | 76.9 | 74.3 | 81.1 | 92.1 |

**Table 5.8 Summary statistics of mean RRSE (in %) for quarterly municipal model-based estimates. The first line contains the measures for the cross-sectional SAE estimates.**

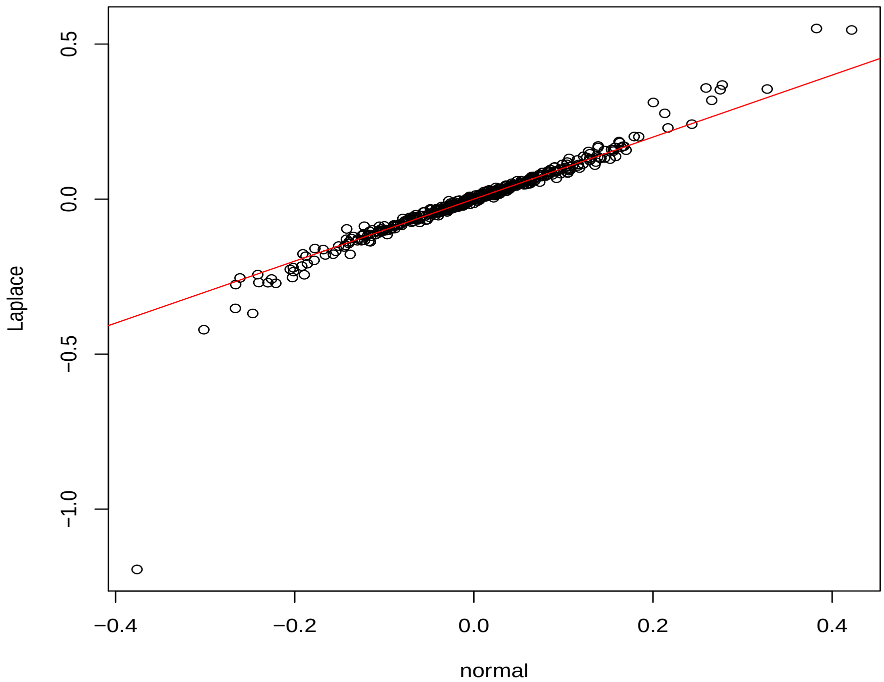


**Figure 5.4 Posterior means of the municipal intercepts based on model BLM3 (normal) and model GL1 (Laplace).**

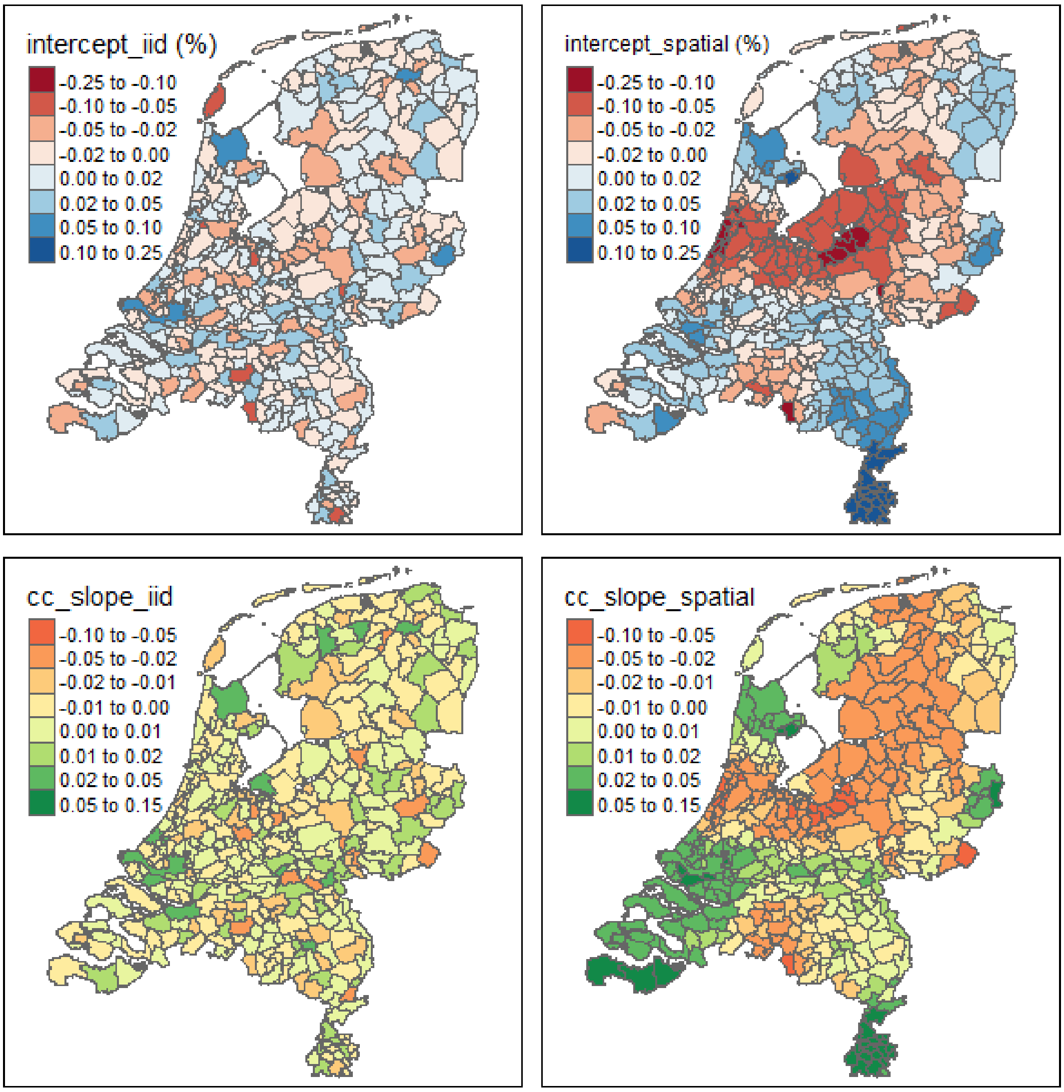


**Figure 5.5 Municipal random intercepts (upper) as well as slopes for claimant counts (lower). Both independent (unstructured) effects (left) and ICAR spatial effects (right) are shown. Note that the intercept sizes have been multiplied by 100 for better readability.**

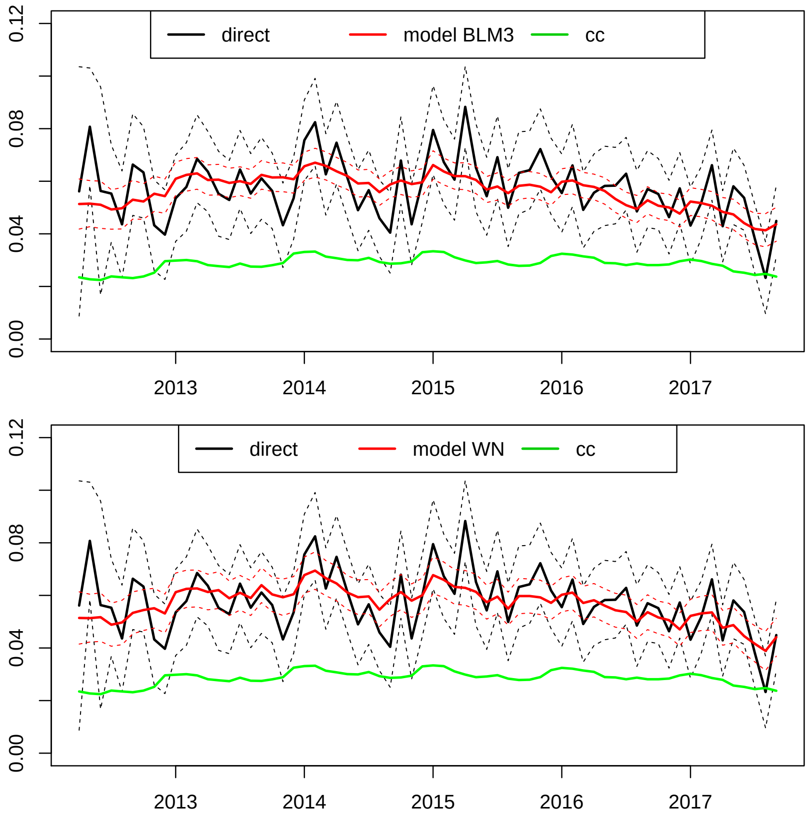


**Figure 5.6 Monthly estimates for one of the provinces, Groningen, based on models BLM3 (above) and WN (below). The thin dashed lines represent the lower and upper limits of approximate 95% intervals. The green line is the fraction of claimant counts.**

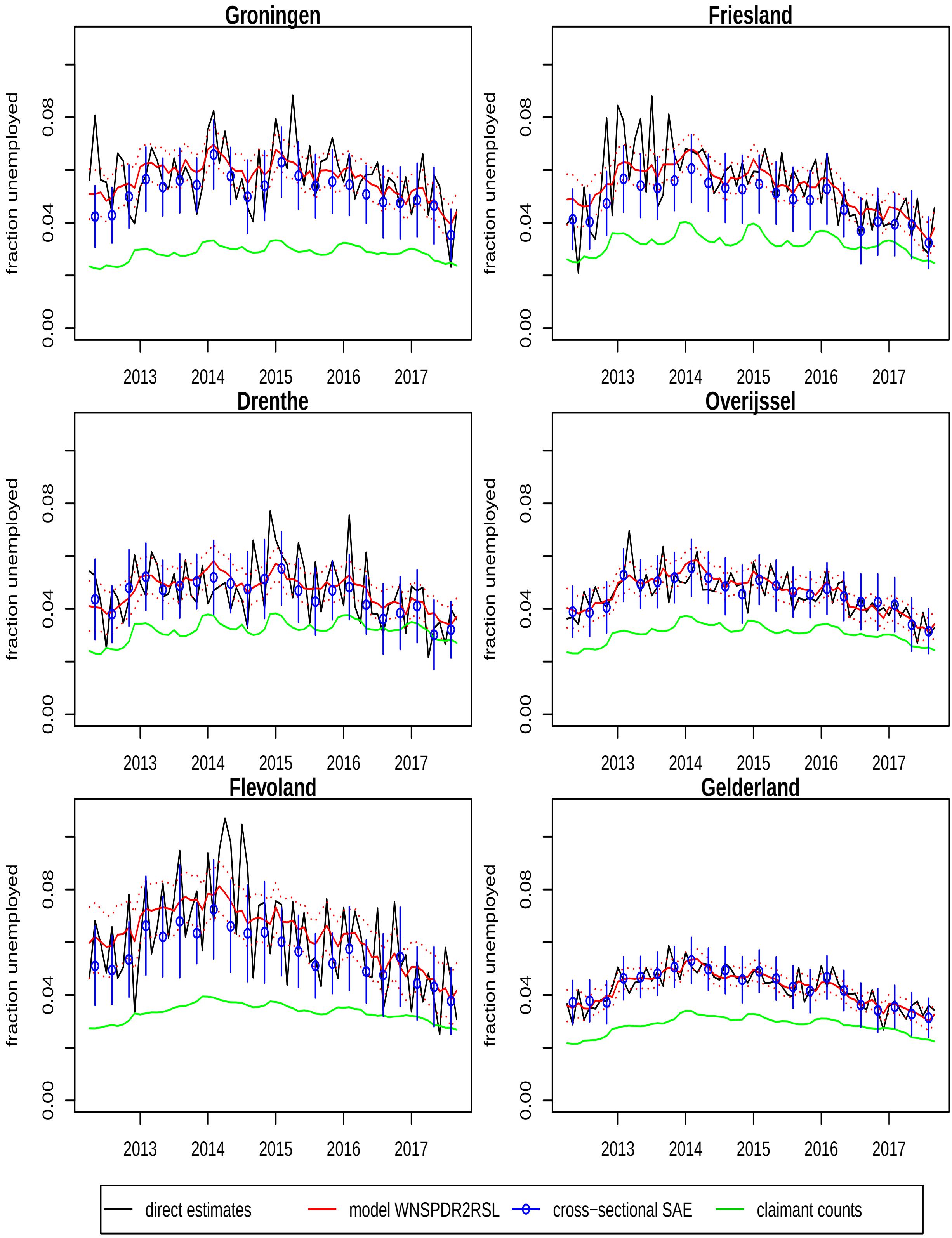


**Figure 5.7 Estimates for the northern and eastern provinces. GREG (black) and time-series model estimates (red) are monthly, whereas the cross-sectional small area estimates are available only on a quarterly basis. The thin red dashed lines represent the lower and upper limits of approximate 95% intervals for the time-series model estimates. The blue bars represent approximate 95% intervals for the cross-sectional SAE estimates.**

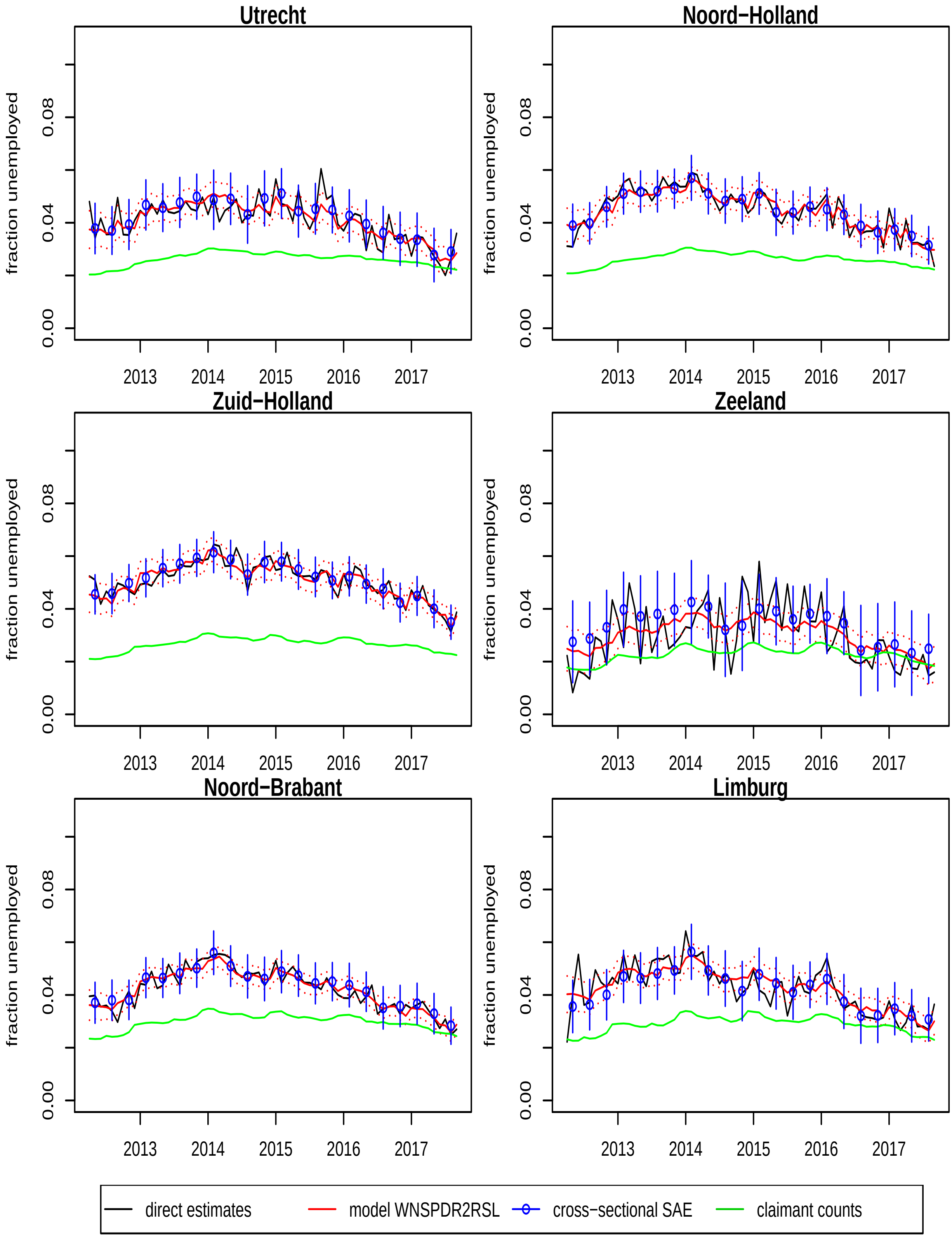


**Figure 5.8 Estimates for the southern and western provinces. GREG (black) and time-series model estimates (red) are monthly, whereas the cross-sectional small area estimates are available only on a quarterly basis. The thin red dashed lines represent the lower and upper limits of approximate 95% intervals for the time-series model estimates. The blue bars represent approximate 95% intervals for the cross-sectional SAE estimates.**

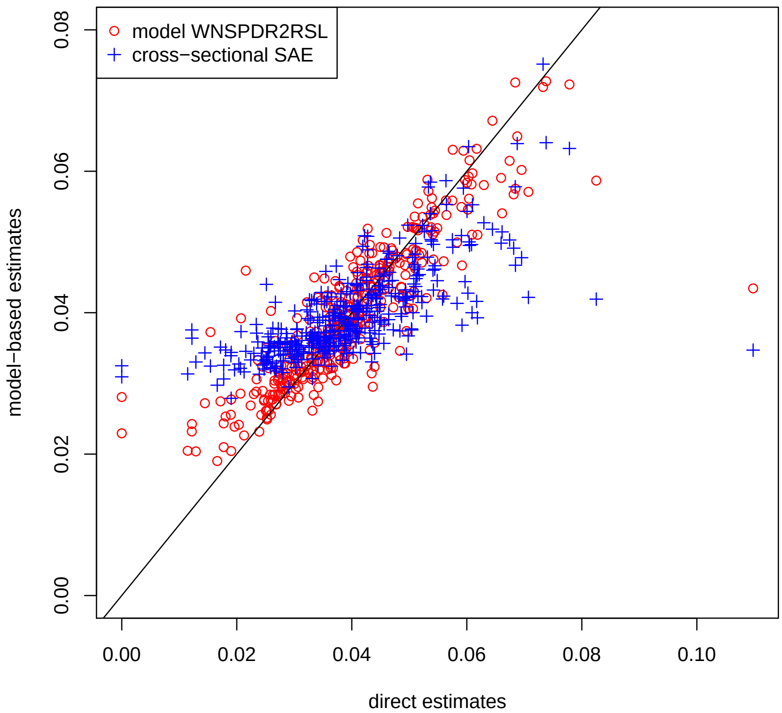


**Figure 5.9 Scatterplot of model-based vs direct municipal estimates averaged over time.**

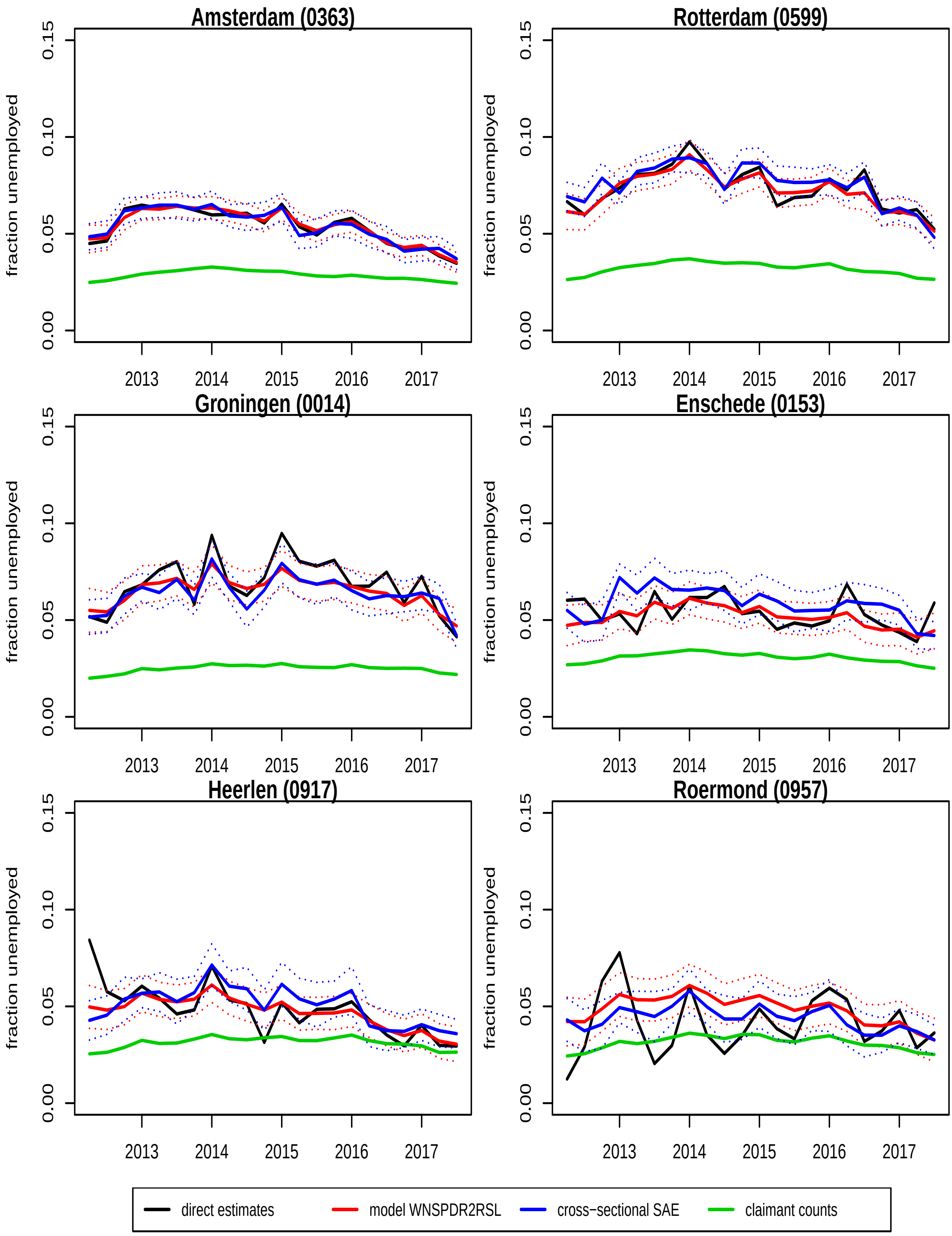


**Figure 5.10 Time-series of quarterly estimates for a selection of large to medium-sized municipalities. The thin dashed lines represent the lower and upper limits of approximate 95% intervals for the time-series and cross-sectional model-based estimates. For reference the claimant count series is also displayed.**

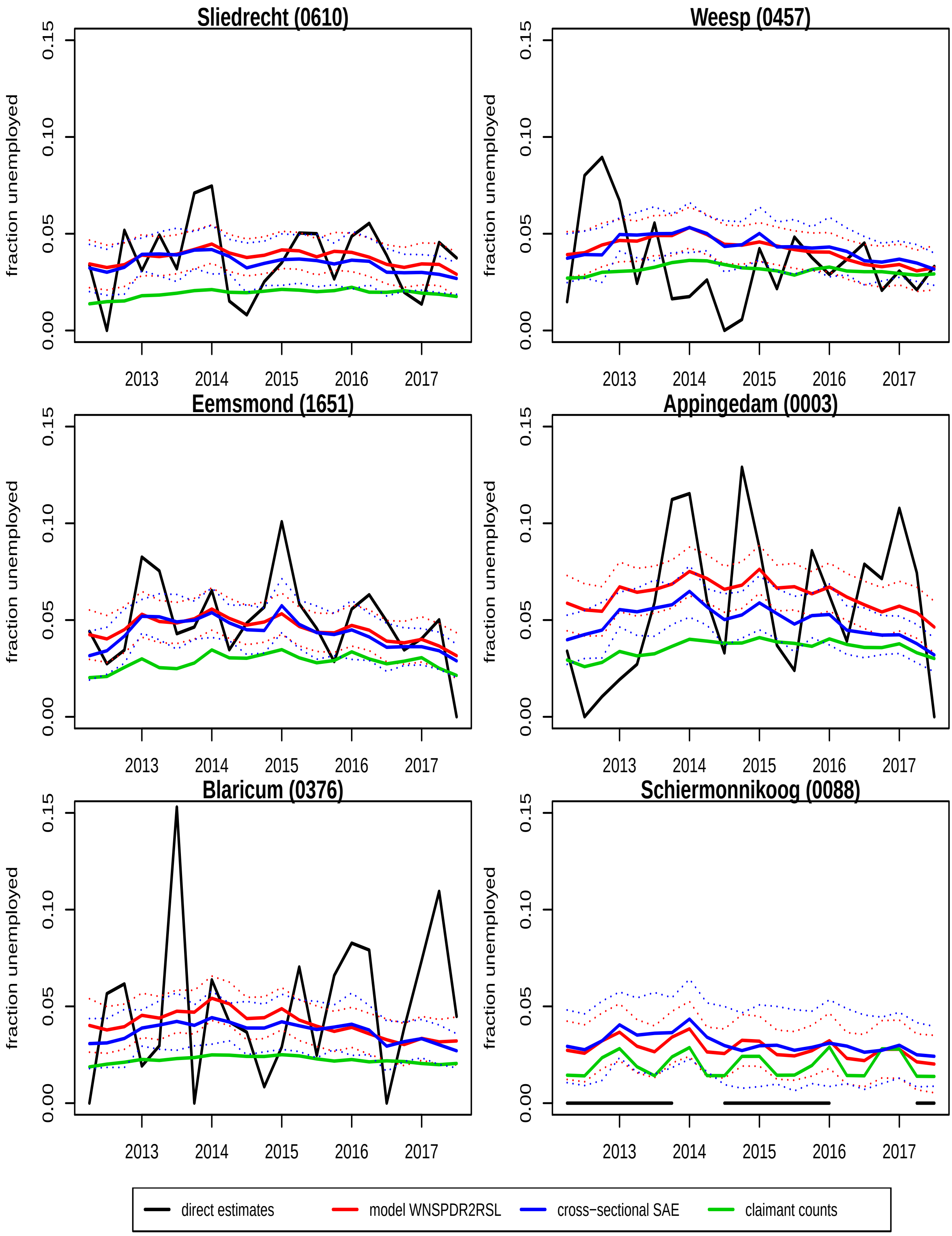


**Figure 5.11 Time-series of quarterly estimates for a selection of medium-sized to small municipalities. The thin dashed lines represent the lower and upper limits of approximate 95% intervals for the time-series and cross-sectional model-based estimates. For reference the claimant count series is also displayed.**

# 6 Discussion

In most countries labour force figures are based on surveys, which are used to produce multiple output tables. Monthly figures are typically produced at a high regional level due to the lack of sufficient sample data. But also for quarterly and annual figures there is an increasing demand for detailed publications at a level for which the sample mass becomes too small to apply standard design-based estimation techniques. Over the last decade the use of model-based small area estimation methods has become more and more accepted in the production of official statistics. Since 2010 Statistics Netherlands indeed uses a multivariate structural time series model (STM) for the publication of official monthly labour force figures at a high regional level. Since 2015 a Battese-Harter-Fuller unit-level model is in place for the production of annual provincial and municipal unemployment figures. A consequence of using different models for different output tables is that numerical consistency between marginals of the various output tables is disturbed.

In this paper a small area estimation strategy is developed that avoids consistency problems between different output tables. The models developed take into account the aggregation levels of interest by means of various random effects terms. The main interest focuses on monthly provincial unemployment and quarterly municipal unemployment. Therefore a time series multilevel model is specified at the most detailed level of month-by-municipality. From the estimated model, monthly provincial estimates are obtained by aggregating monthly municipal predictions over municipalities and quarterly municipal figures by aggregating the same predictions over months.

Input for this model consists of municipal GREG estimates for unemployment. These estimates are calibrated to the monthly figures published with the multivariate STM, which entails a correction for rotation group bias. The time series multilevel model also uses pooled design variances of the GREG estimates as well as design covariances to account for the serial correlation in the sampling errors due to the sample overlap of the rotating panel design.

The input series are direct estimates at the most detailed level. To obtain more accurate small area predictions, a time series multilevel model is developed in an hierarchical Bayesian framework and fitted with the Gibbs sampler. A first finding is that pooling the design variances disturbs the dependency between the standard errors and the level of the estimates. This relation is restored with a parameterized design covariance matrix, by including scale factors for the municipalities in the covariance matrix of the sampling errors. To achieve numerical consistency with the monthly figures at the national level we apply an internal benchmarking procedure, which uses a weighted version of the parameterized design covariance matrix as a predictor matrix that augments the fixed effects design matrix of the model. A plug-in estimate for this parameterized design covariance matrix is obtained by fitting an initial time series model to derive the posterior means of the municipal scale parameters.

The finally selected time series model borrows strength over time and space by defining independent smooth trend models for the provinces. All municipalities within a province share the same smooth trend model. Each municipality has a separate local level trend to account for deviations from the overall provincial trend. In this way (positive) correlations between trend innovations of municipalities within provinces are modelled indirectly. This is a parsimonious alternative for a model with a full covariance matrix among the trend innovations. The time series multilevel model also borrows strength from claimant count series that are included as a dynamic regression component, allowed to vary over time according to a first or second-order random walk.

The claimant count coefficients can also vary between municipalities by means of a random slope component, and the model allows for correlations between the municipal intercepts and slopes. A further extension of the model is realised by including spatial correlations among the random intercepts and among the random slopes for claimant counts using a spatial ICAR model.

In addition, it is investigated to what extent non-normal priors for random effects further improve the model fit over standard normal priors. It turns out that Laplace distributions work well for a number of random effect terms, and they are preferred over t-distributed and horseshoe priors in this application. The Laplace prior is more flexible than the normal prior as it allows for a few outlying random effects and does not seem to overshrink random effects to zero.

Estimates based on the selected time series model improve on the cross-sectional SAE estimates that are currently computed quarterly based on a unit-level Battese-Harter-Fuller model, and published annually. Predictions based on the time series multilevel model are more in line with the level of the GREG series. In other words, on average over time, the bias in the time series multilevel model predictions is smaller than that of the cross-sectional SAE estimates. This is true for both municipal and provincial levels. In addition the time series multilevel model results in a decrease in the standard errors of the model predictions. The decrease is modest at the municipal level, but substantial at the provincial level. More precisely, the time series multilevel model standard errors for quarterly provincial figures are on average 45% smaller than the standard errors of the direct estimates. The average reduction based on the cross-sectional SAE model is about 25%. For the municipal figures the average reduction of the standard errors is 74% based on the time series multilevel model and 72% based on the cross-sectional SAE model. The time series model also produces monthly provincial estimates. For these estimates the average reduction of standard errors compared to the standard errors of the direct estimates is about 56%.

Another advantage of the time series multilevel models is that they produce numerically consistent estimates at both municipal and provincial levels and at monthly and quarterly temporal aggregation levels. Currently, the cross-sectional SAE method is applied separately for municipalities and provinces, and an extra benchmark procedure is necessary to get a consistent set of estimates. Note, however, that the predictions obtained with the time series multilevel model are only approximately consistent with the published monthly national figures. This is in part because the internal benchmarking procedure uses a plug-in estimate of the parameterized covariance matrix of the sampling errors. But exact numerical consistency would not be possible anyway using a simulation-based approach because of Monte Carlo error. In the end a final 'cosmetic' benchmark update remains necessary for exact numerical consistency with published national figures.

The current model accounts indirectly for seasonal effects, since the domain predictions are benchmarked to the GREG estimates at the national level. These GREG estimates are themselves calibrated to the monthly STM estimates at the national level, where seasonal effects are explicitly included in the model. A point for further research is to extend the model with seasonal effects. Some first attempts, however, did not result in an improved overall model fit. If seasonally adjusted regional figures are of interest, then it is important to include regional seasonal effects nonetheless. However, a difficulty in obtaining seasonally adjusted figures is that a seasonal effect is present in the claimant count series too. An alternative would be to treat the claimant count series as a dependent series with its own trend and seasonal effects, and allow for correlations between the trend and possibly seasonal innovations of the LFS unemployment and claimant counts.

Finally, a further obvious model extension is to account for the effects of the survey redesigns of the LFS in 2010 and 2012. These redesigns have led to changes in non-sampling error and consequently to level changes in the time series of direct unemployment estimates. To account for such changes, intervention variables may be added as additional model components. It must then be investigated whether discontinuities should be modelled as fixed or random effects and whether non-normal priors result in better fits. An alternative would be to account for the redesign effects by using GREG input series that have already been adjusted for these effects.

# References


Andrews, D. F. and C. L. Mallows (1974). Scale mixtures of normal distributions. *Journal of the Royal Statistical Society. Series B (Methodological)*, 99--102.

Bailar, B. (1975). The effects of rotation group bias on estimates from panel surveys. *Journal of the American Statistical Association 70*(349), 23--30.

Bates, D. and D. Eddelbuettel (2013). Fast and elegant numerical linear algebra using the RcppEigen package. *Journal of Statistical Software 52*(5), 1--24.

Bates, D. and M. Maechler (2010). *Matrix: Sparse and Dense Matrix Classes and Methods*. R package version 0.999375-44.

Battese, G., R. Harter, and W. Fuller (1988). An error-components model for prediction of county crop areas using survey and satellite data. *Journal of the American Statistical Association 83*(401), 28--36.

Bell, W., G. Datta, and M. Ghosh (2013). Benchmarking small area estimators. *Biometrika 100*(1), 189--202.

Besag, J. and C. Kooperberg (1995). On conditional and intrinsic autoregressions. *Biometrika 82*(4), 733--746.

Besag, J., J. York, and A. Mollie. Bayesian image restoration with two applications in spatial statistics (with discussion) ann inst stat math. 1991; 43: 1--59. doi: 10.1007. *BF00116466.[Cross Ref]*.

Boonstra, H. J. (2014). Time-series small area estimation for unemployment based on a rotating panel survey. Technical Report 201417, https://www.cbs.nl/nl-nl/achtergrond/2014/25/time-series-small-area-estimation-for-unemployment-based-on-a-rotating-panel-survey, Statistics Netherlands.

Boonstra, H. J. (2018). *mcmcsae: MCMC Small Area Estimation*. R package version 0.9.

Boonstra, H. J., B. Buelens, K. Leufkens, and M. Smeets (2011). Small area estimates of labour status in dutch municipalities. Technical Report 201102, https://www.cbs.nl/nl-nl/achtergrond/2011/02/small-area-estimates-of-labour-status-in-dutch-municipalities, Statistics Netherlands.

Boonstra, H. J. and J. Michiels (2013). Evaluatie van kleine-domeinschatters voor arbeidspositie op regionaal niveau. Technical Report in Dutch, PPM-2013-03-11-HBTA, Statistics Netherlands.

Boonstra, H. J. and J. van den Brakel (2016). Estimation of level and change for unemployment using multilevel and structural time series models. Technical Report 201610, https://www.cbs.nl/nl-nl/achtergrond/2016/37/estimation-of-level-and-change-for-unemployment, Statistics Netherlands.

Boonstra, H. J., J. van den Brakel, B. Buelens, S. Krieg, and M. Smeets (2008). Towards small area estimation at statistics netherlands. *Metron LXVI*(1), 21--49.

Brezger, A., L. Fahrmeir, and A. Hennerfeind (2007). Adaptive gaussian markov random fields with applications in human brain mapping. *Journal of the Royal Statistical Society: Series C (Applied Statistics) 56*(3), 327--345.

Carter, C. K. and R. Kohn (1996). Markov chain monte carlo in conditionally gaussian state space models. *Biometrika 83*(3), 589--601.

Carvalho, C. M., N. G. Polson, and J. G. Scott (2010). The horseshoe estimator for sparse signals. *Biometrika 97*(2), 465--480.

Datta, G., P. Lahiri, T. Maiti, and K. Lu (1999). Hierarchical bayes estimation of unemployment rates for the states of the u.s. *Journal of the American Statistical Association 94*(448), 1074--1082.

Datta, G. S. and P. Lahiri (1995). Robust hierarchical bayes estimation of small area characteristics in the presence of covariates and outliers. *Journal of Multivariate Analysis 54*(2), 310--328.

Eddelbuettel, D. and R. Francois (2011). Rcpp: Seamless r and c++ integration. *Journal of Statistical Software 40*(8), 1--18.

Fabrizi, E. and C. Trivisano (2010). Robust linear mixed models for small area estimation. *Journal of Statistical Planning and Inference 140*(2), 433--443.

Fahrmeir, L., T. Kneib, and S. Lang (2004). Penalized structured additive regression for space-time data: a bayesian perspective. *Statistica Sinica*, 731--761.

Fay, R. and R. Herriot (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.

Gelfand, A. and A. Smith (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association 85*, 398--409.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis 1*(3), 515--533.

Gelman, A., D. V. Dyk, Z. Huang, and W. Boscardin (2008). Using redundant parameterizations to fit hierarchical models. *Journal of Computational and Graphical Statistics 17*(1), 95--122.

Gelman, A. and J. Hill (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.

Gelman, A., J. Hwang, and A. Vehtari (2014). Understanding predictive information criteria for bayesian models. *Statistics and computing 24*(6), 997--1016.

Gelman, A., X.-L. Meng, and H. Stern (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica*, 733--760.

Gelman, A. and D. Rubin (1992). Inference from iterative simulation using multiple sequences. *Statistical Science 7*(4), 457--472.

Geman, S. and D. Geman (1984). Stochastic relaxation, gibbs distributions and the bayesian restoration of images. *IEEE Trans. Pattn Anal. Mach. Intell. 6*, 721--741.

Guennebaud, G., B. Jacob, et al. (2010). Eigen v3. http://eigen.tuxfamily.org.

Hastie, T. and R. Tibshirani (1990). *Generalized Additive Models*. Chapman & Hall/CRC.

Hodges, J. and D. Sargent (2001). Counting degrees of freedom in hierarchical and other richly-parameterised models. *Biometrika 88*(2), 367--379.

Kish, L. (1965). *Survey Sampling*. Wiley.

Lang, S., E.-M. Fronk, and L. Fahrmeir (2002). Function estimation with locally adaptive dynamic models. *Computational Statistics 17*(4), 479--500.

Leydold, J. and W. Hormann (2017). *GIGrvg: Random Variate Generator for the GIG Distribution*. R package version 0.5.

Makalic, E. and D. F. Schmidt (2016). A simple sampler for the horseshoe estimator. *IEEE Signal Processing Letters 23*(1), 179--182.

O'Malley, A. and A. Zaslavsky (2008). Domain-level covariance analysis for multilevel survey data with structured nonresponse. *Journal of the American Statistical Association 103*(484), 1405--1418.

Park, T. and G. Casella (2008). The bayesian lasso. *Journal of the American Statistical Association 103*(482), 681--686.

Pfeffermann, D. and L. Burck (1990). Robust small area estimation combining time series and cross-sectional data. *Survey Methodology 16*, 217--237.

Pfeffermann, D. and R. Tiller (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 J. G. Scott (2010). Shrink globally, act locally: Sparse bayesian regularization and prediction. *Bayesian statistics 9*, 501--538.

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rao, J. and I. Molina (2015). *Small Area Estimation*. Wiley-Interscience.

Rao, J. and M. Yu (1994). Small area estimation by combining time series and cross-sectional data. *The Canadian Journal of Statistics 22*, 511--528.

Rue, H. and L. Held (2005). *Gaussian Markov Random Fields: Theory and Applications*. Chapman and Hall/CRC.

Rue, H., S. Martino, and N. Chopin (2009). Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. *Journal of the royal statistical society: Series b (statistical methodology) 71*(2), 319--392.

Särndal, C.-E., B. Swensson, and J. Wretman (1992). *Model Assisted Survey Sampling*. Springer.

Spiegelhalter, D., N. Best, B. Carlin, and A. van der Linde (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society B 64*(4), 583--639.

Tang, X., M. Ghosh, N. S. Ha, and J. Sedransk (2018). Modeling random effects using global--local shrinkage priors in small area estimation. *Journal of the American Statistical Association*, 1--14.

Tennekes, M. (2018). tmap: Thematic maps in R. *Journal of Statistical Software 84*(6), 1--39.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 267--288.

van den Brakel, J. and S. Krieg (2009). Structural time series modelling of the monthly unemployment rate in a rotating panel. *Survey Methodology 35*(2), 177--190.

van den Brakel, J. and S. Krieg (2015). Dealing with small sample sizes, rotation group bias and discontinuities in a rotating panel design. *Survey Methodology 41*(2), 267--296.

Vehtari, A., A. Gelman, and J. Gabry (2017). Practical bayesian model evaluation using leave-one-out cross-validation and waic. *Statistics and Computing 27*(5), 1413--1432.

Watanabe, S. (2010). Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research 11*, 3571--3594.

Watanabe, S. (2013). A widely applicable bayesian information criterion. *Journal of Machine Learning Research 14*, 867--897.

Wood, S. N. (2017). *Generalized additive models: an introduction with R*. Chapman and Hall/CRC.

You, Y. (2008a). An integrated modeling approach to unemployment rate estimation for sub-provincial areas of canada. *Survey Methodology 34*(1), 19--27.

You, Y. (2008b). Small area estimation using area level models with model checking and applications. *SSC annual meeting, Proceedings of the Survey Methods Section*.

You, Y., J. Rao, and J. Gambino (2003). Model-based unemployment rate estimation for the canadian labour force survey: A hierarchical bayes approach. *Survey Methodology 29*(1), 25--32.

Zhao, Y., J. Staudenmayer, B. Coull, and M. Wand (2006). General design bayesian generalized linear mixed models. *Statistical Science 21*(1), 35--51.

# Appendix

# A Gibbs sampler for the multilevel time-series model

For notational convenience we rewrite model (7) as

$$y = \eta + e \tag{A.1}$$

where $y = \hat{\hat{Y}}$ is the data vector consisting of $M$ GREG estimates, $\eta$ is a linear predictor built from various fixed and random effect terms, and $e$ is a vector of survey errors, modeled as

$$e \sim N(0, \oplus_{i=1}^{M_A} \lambda_i \Phi_i)\,, \tag{A.2}$$

in terms of a covariance matrices $\Phi_i$ depending on initial variance estimates treated as known, and scale factors $\lambda_i$, one for each area $i = 1, \dots, M_A$.

All quantities of interest considered can be expressed as linear combinations of the fixed and random effects. Let $\theta$ be such a vector of quantities of interest. Inference about $\theta$ is based on its posterior distribution $p(\theta|y)$. This distribution cannot be obtained in closed form and cannot be directly sampled from. Therefore we use a Markov chain Monte Carlo (MCMC) method, and in particular the Gibbs sampler (Geman and Geman, 1984; Gelfand and Smith, 1990). Using the Gibbs sampler we sample from the joint posterior $p(\psi|\hat{\hat{Y}})$ where $\psi$ is the vector of all model parameters, including $\lambda, \beta$ and the parameters associated with each random effects term. The joint posterior is determined by the model and prior specifications,

$$p(\psi|y) \propto p(\psi)p(y|\psi)\,, \tag{A.3}$$

up to a normalization constant. The Gibbs sampler generates samples from this posterior distribution, and from these samples and the definition of the parameters of interest $\theta$ we obtain posterior samples for the latter.

The Gibbs sampler iteratively samples from the full conditional distributions $p(\psi_g|y, \psi_1, \dots, \psi_{g-1}, \psi_{g+1}, \dots, \psi_G)$, for a suitable decomposition of $\psi$ in blocks $\psi_g$, $g = 1, \dots, G$. The full conditionals for the class of linear multilevel models considered in this paper are easy to sample from: they are normal for all (fixed or random) coefficients and inverse chi-squared or inverse-Wishart for the variance parameters. With $G$ the number of parameter blocks and $K$ the number of simulations, the Gibbs sampling algorithm is as follows:

choose starting values $\psi_g^{(0)}$ for $g = 1, \dots, G$
**for** $k$ in 1 **to** $K$
  **for** $g$ in 1 **to** $G$
    draw $\psi_g^{(k)}$ from $p(\psi_g|y, \psi_1^{(k)}, \dots, \psi_{g-1}^{(k)}, \psi_{g+1}^{(k-1)}, \dots, \psi_G^{(k-1)})$

After convergence, samples can be considered draws from $p(\psi|y)$.

Below we give the full conditional distributions for all model parameters. The vector $\beta$ of fixed effects is sampled in a single block, and each random effect term $v^{(\alpha)}$ is considered a block as well.

## A.1 Full conditional distributions

For the vector of data-level variance parameters we use as prior

$$\lambda_i \overset{\text{ind}}{\sim} \text{Inv}{-}\chi^2(1,1)\,, \tag{A.4}$$

for $i = 1, \dots, M_A$. We have that $\Phi = \bigoplus_{i=1}^{M_A} \Phi_i$, since initial estimates for different areas are uncorrelated. The full conditional for $\lambda_i$ is then

$$p(\lambda_i|y,.) \propto p(\lambda_i) N\,(e|0, \lambda_i \Phi_i)\ , \tag{A.5}$$

independently for $i = 1, \dots, M_A$, where $e = y - \eta$ is the vector of residuals. We use the notation '.' in $p(\lambda_i|y,.)$ to denote conditioning on all other parameters. This yields inverse chi-squared distributions

$$\begin{aligned} p(\lambda_i|y,.) &= \text{Inv}{-}\chi^2(\lambda_i|d_i, s_i)\,,\\ d_i &= n_i + 1\,,\\ s_i &= \frac{1}{d_i}\left(1 + e'_{\{i\}}\Phi_i^{-1} e_{\{i\}}\right)\,, \end{aligned} \tag{A.6}$$

where subscript $\{i\}$ denotes the data units associated with municipality $i$, and $n_i$ is the number of those units. In our case $n_i$ equals the number of months for which the response for municipality $i$ is non-empty, which means that $n_i = M_T$ for most municipalities.

Given a prior distribution $p(\beta) = N(\beta|b_0, \Omega_\beta)$, the full conditional distribution for the vector $\beta$ of fixed effects is

$$\begin{aligned} p(\beta|y,.) &= N(\beta|E_\beta, V_\beta)\,,\\ V_\beta &= \left(X'\Sigma^{-1}X + \Omega_\beta^{-1}\right)^{-1}\,,\\ E_\beta &= V_\beta\left(X'\Sigma^{-1}e_\beta + \Omega_\beta^{-1}b_0\right)\,, \end{aligned}$$

where $\Sigma = \bigoplus_{i=1}^{M_A} \lambda_i \Phi_i$ and $e_\beta = y - \eta + X\beta$ is the vector of 'partial' residuals.

Next, we turn to the full conditional distributions associated with a generic random effect component $Z^{(\alpha)}v^{(\alpha)}$. In the description below, the superscript $\alpha$ is omitted.

Let $Z$ be a $M \times q$ design matrix corresponding to $d$ effects that can vary over the $l$ levels of a factor variable. Let $v$ be the corresponding $q$-vector of random effects, $v = (v_{ik})'_{i=1\dots l\,;\,k=1\dots d} = (v_{11}, v_{12}, \dots)'$, where by convention the last index runs fastest. The random effect contribution to the linear predictor is $Zv$.

We use redundant multiplicative parameterization, which improves convergence of the Gibbs sampler (Gelman et al., 2008), and yields more robust prior distributions for the variance parameters (Gelman, 2006). For that purpose, a $d$-dimensional parameter vector $\xi$ and a $q$-vector $\tilde{v}$ of raw random effects are introduced, which combine to form the original coefficients as

$$v = \Delta_\xi \tilde{v} \qquad \Delta_\xi = I_l \otimes \text{diag}(\xi) = \text{diag}(W\xi)\,, \tag{A.7}$$

where $W = \iota_l \otimes I_d$ is a $q \times d$ indicator-matrix, and $\iota_l$ is an $l$-vector of ones.

Priors on $\xi$ and $\tilde{v}$ are

$$\begin{aligned} \xi &\sim N(0, I_d)\,,\\ \tilde{v} &\sim N(0, A \otimes \tilde{V})\,, \end{aligned} \tag{A.8}$$

where $A$ is a given, possibly degenerate $l \times l$ covariance matrix, and $\tilde{V}$ a parameterized $d \times d$ covariance matrix.

Three different parameterizations of $\tilde{V}$ are considered:

a.) $\tilde{V}$ is an unstructured covariance matrix with prior

$$\tilde{V} \sim \mathrm{Inv-Wish}(\nu_v, \Psi_v)\,, \tag{A.9}$$

with degrees of freedom $\nu_v$, by default taken to be $d+1$, and $d \times d$ scale matrix $\Psi_v$, by default equal to $I_d$.

b.) $\tilde{V} = \mathrm{diag}(\tilde{\sigma}^2_{v;1} \dots \tilde{\sigma}^2_{v;d})$, a diagonal variance matrix with independent inverse chi-squared priors on the variances,

$$\tilde{\sigma}^2_{v;k} \overset{\mathrm{ind}}{\sim} \mathrm{Inv-}\chi^2(\nu_{v;k}, s^2_{v;k}) \tag{A.10}$$

c.) $\tilde{V} = \tilde{\sigma}^2_v I_d$. The prior for the single common variance parameter $\tilde{\sigma}^2_v$ is

$$\tilde{\sigma}^2_v \sim \mathrm{Inv-}\chi^2(\nu_v, s^2_v)\,. \tag{A.11}$$

Note that if $d = 1$, parameterizations a.) and b.) reduce to c.), provided that $\Psi_v$ is identified with $\nu_v s^2_v$.

The prior for the original coefficients $v$ is, given $\xi$ and $\tilde{V}$,

$$v \sim N(0, A \otimes V) \qquad V = \mathrm{diag}(\xi)\,\tilde{V}\,\mathrm{diag}(\xi)\,. \tag{A.12}$$

The $l \times l$ matrix $A$ describes the covariance structure between the levels of the factor variable. It is specified in terms of its inverse $Q_A$, which directly reflects the conditional dependence structure between the levels and is usually sparse.

The precision matrix $Q_A$ may be singular. The singular vectors of $Q_A$ correspond to directions along which the prior is constant, i.e. non-informative. Let $R_A$ be the $l \times r$ matrix of singular vectors such that $Q_A R_A = 0$. The matrix $R = R_A \otimes I_d$ may then be used as a constraint matrix to impose $R\tilde{v} = 0$, or equivalently $Rv = 0$, so that other terms in the model remain identifiable.

First we derive the full conditional for $\xi$, followed by that of $\tilde{V}$ and $\tilde{v}$. For $\xi$,

$$p(\xi|y,.) \propto N(\xi|0, I_d) N(y|Z\Delta_\xi \tilde{v} + \dots, \Sigma) \tag{A.13}$$

Now

$$\Delta_\xi \tilde{v} = \mathrm{diag}(W\xi)\tilde{v} = \mathrm{diag}(\tilde{v})W\xi = \Delta_{\tilde{v}}\xi\,, \tag{A.14}$$

where $\Delta_{\tilde{v}} = \mathrm{diag}(\tilde{v})W$. Therefore,

$$\begin{aligned} p(\xi|y,.) &= N(\xi|E_\xi, V_\xi) \\ V_\xi &= \left(\Delta'_{\tilde{v}} Z' \Sigma^{-1} Z \Delta_{\tilde{v}} + I_d\right)^{-1} \\ E_\xi &= \left(\Delta'_{\tilde{v}} Z' \Sigma^{-1} Z \Delta_{\tilde{v}} + I_d\right)^{-1} \Delta'_{\tilde{v}} Z' \Sigma^{-1} e_v\,, \end{aligned} \tag{A.15}$$

where $e_v = y - \eta + Zv$. Note that everything can be expressed in terms of $v$ instead of $\tilde{v}$ by using $\tilde{v} = \Delta_\xi^{-1} v$. For $d = 1$ or in the case that $\tilde{V} = \sigma^2_v I_d$ is defined in terms of a single variance parameter, (A.15) reduces to

$$\begin{aligned} p(\xi|y,.) &= N(\xi|E_\xi, V_\xi) \\ V_\xi &= \left(\tilde{v}' Z' \Sigma^{-1} Z \tilde{v} + 1\right)^{-1} \\ E_\xi &= \left(\tilde{v}' Z' \Sigma^{-1} Z \tilde{v} + 1\right)^{-1} \tilde{v}' Z' \Sigma^{-1} e_v\,. \end{aligned} \tag{A.16}$$

For the full conditional distribution for $\tilde{V}$ we distinguish between three situations:

a.) In the case that $\tilde{V}$ is a fully parameterised covariance matrix with an inverse Wishart prior,

$$\begin{aligned} p(\tilde{V}|y,.) &\propto \mathrm{Inv-Wish}(\tilde{V}|\nu_v, \Psi_v) N(\tilde{v}|0, A \otimes \tilde{V}) \\ &\propto \left|A \otimes \tilde{V}\right|^{-1/2} e^{-\frac{1}{2}\tilde{v}'(Q_A \otimes \tilde{V}^{-1})\tilde{v}} \\ &\times \left|\tilde{V}\right|^{-(\nu_v+d+1)/2} e^{-\frac{1}{2}\mathrm{tr}(\Psi_v \tilde{V}^{-1})} \\ &\propto \left|\tilde{V}\right|^{-(\nu_v+l^*+d+1)/2} e^{-\frac{1}{2}\mathrm{tr}[(\Psi_v + \tilde{v}_M' Q_A \tilde{v}_M)\tilde{V}^{-1}]}, \end{aligned} \tag{A.17}$$

where $\tilde{v}_M$ is the $l \times d$ matrix such that $\tilde{v} = \mathrm{vec}(\tilde{v}_M')$, i.e. the matrix composed of stacking the $l$ row vectors $\tilde{v}_i'$. If the precision matrix $Q_A = A^{-1}$ is singular and constraints associated with all singular vectors are imposed on $\tilde{v}$ or $v$, then $l^*$ should be taken equal to the rank of $Q_A$.[1) ] Otherwise $l^* = l$. We used the relation

$$|C \otimes D| = |C|^{\mathrm{rank}(D)} |D|^{\mathrm{rank}(C)}, \tag{A.18}$$

as well as the relations $\mathrm{tr}(C'D) = \mathrm{vec}(C)'\mathrm{vec}(D)$ and $\mathrm{vec}(CDE) = (E' \otimes C)\mathrm{vec}(D)$ from which follows

$$\begin{aligned} &\tilde{v}'\left(Q_A \otimes \tilde{V}^{-1}\right)\tilde{v} = \mathrm{vec}(\tilde{v}_M')'\left(Q_A \otimes \tilde{V}^{-1}\right)\mathrm{vec}(\tilde{v}_M') \\ &= \mathrm{vec}(\tilde{v}_M')'\mathrm{vec}(\tilde{V}^{-1}\tilde{v}_M' Q_A) = \mathrm{tr}(\tilde{v}_M \tilde{V}^{-1} \tilde{v}_M' Q_A) \\ &= \mathrm{tr}(\tilde{v}_M' Q_A \tilde{v}_M \tilde{V}^{-1}). \end{aligned} \tag{A.19}$$

So in the case of an unstructured covariance matrix $\tilde{V}$,

$$\begin{aligned} p(\tilde{V}|y,.) &= \mathrm{Inv-Wish}(\tilde{V}|\nu_{v1}, \Psi_{v1}) \\ \nu_{v1} &= \nu_v + l^* \\ \Psi_{v1} &= \Psi_v + \tilde{v}_M' Q_A \tilde{v}_M. \end{aligned} \tag{A.20}$$

Note that $V$ is obtained from (A.12), or more immediately by drawing from an inverse Wishart distribution with the same degrees of freedom, but with scale matrix

$$\mathrm{diag}(\xi)\Psi_v \mathrm{diag}(\xi) + v_M' Q_A v_M, \tag{A.21}$$

where $v_M = \tilde{v}_M \mathrm{diag}(\xi)$.

b.) In the case that $\tilde{V}$ is diagonal and independent inverse chi-squared priors are assigned to the variance parameters $\tilde{\sigma}_{v;k}^2$ for $k = 1, \dots, d$,

$$p(\tilde{\sigma}_{v;1}^2 \dots \tilde{\sigma}_{v;q_0}^2|y,.) \propto \prod_{k=1}^{q_0} \left(\tilde{\sigma}_{v;k}^2\right)^{-\frac{\nu_{v;k}}{2}} e^{-\frac{\nu_{v;k}s_{v;k}^2}{2\tilde{\sigma}_{v;k}^2}} (\tilde{\sigma}_{v;k}^2)^{-l^*/2} e^{-\frac{1}{2\tilde{\sigma}_{v;k}^2}\left(\tilde{v}_M' Q_A \tilde{v}_M\right)_{kk}},$$

since in this case

$$\tilde{v}'(\tilde{V}^{-1} \otimes Q_A)\tilde{v} = \tilde{v}'\left(\oplus_{k=1}^d \frac{1}{\sigma_{v;k}^2} Q_A\right)\tilde{v} = \sum_{k=1}^d \frac{1}{\tilde{\sigma}_{v;k}^2}\left(\tilde{v}_M' Q_A \tilde{v}_M\right)_{kk}.$$

1) All full conditionals are expressed in terms of the precision matrix $Q_A$, and not in terms of the covariance matrix $A$. Strictly speaking, the latter as inverse of $Q_A$ is not well-defined, although it can still be understood as a pseudo-inverse.

Therefore,

$$\begin{aligned}
p(\tilde{\sigma}^2_{v;1} \dots \tilde{\sigma}^2_{v;d}|y,.) &= \prod_{k=1}^{d} \text{Inv}{-}\chi^2(\tilde{\sigma}^2_{v;k}|\nu_{v1;k}, s^2_{v1;k}) \\
\nu_{v1;k} &= \nu_{v;k} + l^* \\
s^2_{v1;k} &= \frac{1}{\nu_{v1;k}} \left(\nu_{v;k} s^2_{v;k} + (\tilde{v}'_M Q_A \tilde{v}_M)_{kk}\right) .
\end{aligned} \tag{A.22}$$

The original variance parameters are obtained by drawing independently from inverse chi-squared distributions with the same degrees of freedom, but with scale parameters

$$\frac{1}{\nu_{v1;k}} \left(\nu_{v;k} s^2_{v;k} \xi^2_k + (v'_M Q_A v_M)_{kk}\right) \tag{A.23}$$

c.) In the case that $\tilde{V} = \tilde{\sigma}^2_v I_d$,

$$p(\tilde{\sigma}^2_v|y,.) \propto \left(\tilde{\sigma}^2_v\right)^{-\frac{\nu_v}{2}} e^{-\frac{\nu_v s^2_v}{2\tilde{\sigma}^2_v}} (\tilde{\sigma}^2_v)^{-dl^*/2} e^{-\frac{1}{2\tilde{\sigma}^2_v}\text{tr}(\tilde{v}'_M Q_A \tilde{v}_M)} \tag{A.24}$$

and so

$$\begin{aligned}
p(\tilde{\sigma}^2_v|y,.) &= \text{Inv}{-}\chi^2(\tilde{\sigma}^2_v|\nu_{v1}, s^2_{v1}) \\
\nu_{v1} &= \nu_v + dl^* \\
s^2_{v1} &= \frac{1}{\nu_{v1}} \left(\nu_v s^2_v + \text{tr}(\tilde{v}'_M Q_A \tilde{v}_M)\right) .
\end{aligned} \tag{A.25}$$

The original variance parameter is $\sigma^2_v = \xi^2 \tilde{\sigma}^2_v$.

Finally, the full conditional distribution for the vector $\tilde{v}$ of random effects is

$$p(\tilde{v}|y,.) \propto N(\tilde{v}|0, A \otimes \tilde{V}) N(y|Z\Delta_\xi \tilde{v} + \dots, \Sigma) , \tag{A.26}$$

implying

$$\begin{aligned}
p(\tilde{v}|y,.) &= N(\tilde{v}|E_{\tilde{v}}, V_{\tilde{v}}) \\
V_{\tilde{v}} &= \left(\Delta'_\xi Z' \Sigma^{-1} Z \Delta_\xi + Q_A \otimes \tilde{V}^{-1}\right)^{-1} \\
E_{\tilde{v}} &= \left(\Delta'_\xi Z' \Sigma^{-1} Z \Delta_\xi + Q_A \otimes \tilde{V}^{-1}\right)^{-1} \Delta'_\xi Z' \Sigma^{-1} e_v .
\end{aligned} \tag{A.27}$$

Note that $\Delta'_\xi = \Delta_\xi$ as it is a diagonal matrix. Since $v = \Delta_\xi \tilde{v}$ and $V = \text{diag}(\xi)\tilde{V}\text{diag}(\xi)$, we can immediately obtain $v$ (conditional on $\xi$ and $\tilde{V}$) by drawing from

$$\begin{aligned}
p(v|y,.) &= N(v|E_v, V_v) \\
V_v &= \left(Z' \Sigma^{-1} Z + Q_A \otimes V^{-1}\right)^{-1} \\
E_v &= \left(Z' \Sigma^{-1} Z + Q_A \otimes V^{-1}\right)^{-1} Z' \Sigma^{-1} e_v .
\end{aligned} \tag{A.28}$$

Constraints are imposed by modifying draws of $\tilde{v}$ or $v$ as follows (Rue and Held, 2005),

$$v \to v^* = v - V_v R \left(R' V_v R\right)^{-1} R' v , \tag{A.29}$$

so that $R' v^* = 0$.

## A.2 Global-local shrinkage

In the case that global-local shrinkage priors are used, the precision matrix $Q_A$ in the full conditionals listed above is a function of $l_D$ local variance scale parameters $\omega_k$,

$$Q_A = D_A' \mathrm{diag}(\omega_1, \dots \omega_{l_D})^{-1} D_A \,, \tag{A.30}$$

where $D_A$ is a given $l_D \times l$ matrix, see Section 4.3 in the main text. The local variance factors are assigned independent priors $p(\omega) = \prod_{i=1}^{l_D} p(\omega_i)$. The full conditional for $\omega$ is

$$\begin{aligned} p(\omega|y,.) &\propto p(\omega) N(\tilde{v}|0, A \otimes \tilde{V}) \\ &\propto \left( \prod_{i=1}^{l_D} p(\omega_i) \omega_i^{-d/2} \right) e^{-\frac{1}{2}\tilde{v}'\left(\left(D_A' \,\mathrm{diag}(\omega_1,\dots,\omega_{l_D})^{-1} \, D_A\right) \otimes \tilde{V}^{-1}\right)\tilde{v}} \\ &= \left( \prod_{i=1}^{l_D} p(\omega_i) \omega_i^{-d/2} \right) e^{-\frac{1}{2}\mathrm{tr}\left(\tilde{v}_M' D_A' \,\mathrm{diag}(\omega_1,\dots,\omega_{l_D})^{-1} \, D_A \tilde{v}_M \tilde{V}^{-1}\right)} \\ &= \left( \prod_{i=1}^{l_D} p(\omega_i) \omega_i^{-d/2} \right) e^{-\frac{1}{2}\mathrm{tr}\left(D_A \tilde{v}_M \tilde{V}^{-1} \tilde{v}_M' D_A' \,\mathrm{diag}(\omega_1,\dots,\omega_{l_D})^{-1}\right)} \\ &= \prod_{i=1}^{l_D} p(\omega_i) \omega_i^{-d/2} e^{-\frac{1}{2} M_{ii}/\omega_i} \,, \end{aligned} \tag{A.31}$$

where $M = D_A \tilde{v}_M \tilde{V}^{-1} \tilde{v}_M' D_A'$. The Laplace priors considered result from taking $p(\omega_i)$ to be (independent) exponential distributions. In particular, with

$$p(\omega_i) = e^{-\omega_i} \,, \tag{A.32}$$

the full conditionals for $\omega_i$ are independent and given by

$$p(\omega_i|y,.) \propto \omega_i^{-d/2} e^{-\left(\omega_i + \frac{M_{ii}}{2}\frac{1}{\omega_i}\right)} \,. \tag{A.33}$$

This can be recognized as a generalized inverse Gaussian (GiG) distribution, whose density for general parameter values $a \geq 0$, $b \geq 0$ and $p$ is defined as

$$\mathrm{GiG}(x|a,b,p) = \frac{(a/b)^{p/2}}{2K_p(\sqrt{ab})} x^{p-1} e^{-\frac{1}{2}(ax + b/x)} \,, \tag{A.34}$$

for $x > 0$. Here $K_p$ is a modified Bessel function of the second kind. So the full conditionals for $\omega_i$ are GiG with parameters $a_i = 2$, $b_i = M_{ii}$ and $p_i = 1 - d/2$.

Taking inverse chi-squared distributions as priors for $\omega_i$ results in Student-t distributed innovations. In particular, with

$$p(\omega_i) = \mathrm{Inv}{-}\chi^2(\omega_i|\nu_\omega, 1) \,, \tag{A.35}$$

the full conditional for $\omega_i$ is again inverse chi-squared

$$p(\omega_i|y,.) = \mathrm{Inv}{-}\chi^2\left(\omega_i|\nu_\omega + d, \frac{\nu_\omega + M_{ii}}{\nu_\omega + d}\right) . \tag{A.36}$$

Finally, a horseshoe prior for the innovations is obtained with independent half-Cauchy priors on the standard deviations $\sqrt{\omega_i}$. A convenient representation for these priors is the following scale mixture of inverse chi-squared distributions for the variances $\omega_i$:

$$p(\omega_i|\kappa_i) = \text{Inv}{-}\chi^2\left(\omega_i|1, \frac{1}{\kappa_i}\right), \qquad p(\kappa_i) = \text{Inv}{-}\chi^2(1,1), \tag{A.37}$$

where $\kappa_i$ are additional (independent) auxiliary parameters. This mixture representation leads to a simple Gibbs sampler in terms of inverse chi-squared full conditionals (Makalic and Schmidt, 2016):

$$\begin{aligned} p(\omega_i|y,.) &= \text{Inv}{-}\chi^2\left(\omega_i|d+1, \frac{M_{ii}+1/\kappa_i}{d+1}\right), \\ p(\kappa_i|y,.) &= \text{Inv}{-}\chi^2\left(\kappa_i|2, \frac{1+1/\omega_i}{2}\right). \end{aligned} \tag{A.38}$$

## A.3 Implementation

The above Gibbs sampler for the broad class of multilevel models as described has been implemented in R (R Core Team, 2018), and is being developed into an R package called `mcmcsae` (Boonstra, 2018). The package makes extensive use of the sparse matrix facilities provided by package `Matrix` (Bates and Maechler, 2010) and also of some dense and sparse matrix routines of the C++ library Eigen (Guennebaud et al., 2010), via R packages `Rcpp` (Eddelbuettel and Francois, 2011) and `RcppEigen` (Bates and Eddelbuettel, 2013). R package `GiGrvg` (Leydold and Hormann, 2017) is used to sample from the generalized inverse Gaussian full conditional distributions (A.33).

## Colophon

*Publisher*
Statistics Netherlands
Henri Faasdreef 312, 2492 JP The Hague
www.cbs.nl

*Prepress*
Statistics Netherlands, Grafimedia

*Design*
Edenspiekermann

*Information*
Telephone +31 88 570 70 70, fax +31 70 337 59 94
Via contact form: www.cbs.nl/information