Statistics Netherlands

Discussion Paper

Time-series small area estimation for unemployment based on a rotating panel survey

The views expressed in this paper are those of the author and do not necessarily reflect the policies of Statistics Netherlands.

2014 | 17

Harm Jan Boonstra
6 June 2014

A time-series multi-level model is used to estimate municipal unemployment based on the Dutch Labour Force Survey (LFS) at a quarterly frequency. The model includes random municipality effects as well as random municipality-by-quarter effects. The latter are modeled as independent effects or random walks over time, or as a sum of both independent and random walk components. The model is fit using quarterly LFS data from 2003 to 2008. The LFS uses a rotating panel design so that in each quarter the total sample consists of five waves. Because the waves differ according to their time-in-panel and observation strategy, systematic biases arise. These rotation group biases are handled in the model by including measurement effects for the second to fifth waves. The time-series multi-level model is fit using a Gibbs sampler. Based on its output quarterly unemployment estimates and standard errors for over 400 municipalities are computed. Our application to LFS data therefore serves as an example of large-scale Bayesian inference from survey data. The time-series estimates are compared to cross-sectional estimates based on basic area and unit-level models that borrow strength over space but not over time.

Key words: Small area estimation, Time-series models, Multi-level models, Measurement error, Labour Force Survey, Unemployment, Gibbs sampler

Contents

1 Introduction 4

2 Data 5

3 Cross-sectional small area estimation 6
3.1 Combining data from all waves 7

4 Time-series small area estimation 8
4.1 Initial estimates 8
4.2 Time-series multi-level model 13

5 Selecting predictors and random effects for the time-series model 14

6 Comparison of estimates 19

7 Discussion 25

I Gibbs sampler for the time-series model 31
I.1 Gibbs sampling 32
I.2 Redundant multiplicative reparameterization 32
I.3 Derivation of the full conditionals 33
I.4 The Deviance Information Criterion 36
I.5 Implementation 36

II Results for time-series Model `7 vuw' 37

1 Introduction

Data from the Dutch Labour Force Survey (LFS) are used to estimate labour status at the municipal level. Until now these estimates are produced annually by means of direct generalized regression estimation (GREG, see e.g. Särndal et al. (1992)) for municipalities with at least 30 thousand inhabitants. For municipalities with 10 to 30 thousand inhabitants, three-year averages of GREG estimates are used in order to reduce the variance. No estimates are published for municipalities with fewer than 10 thousand inhabitants, because even three-year averages are deemed too unstable in this case. Direct estimation is not even possible for a few small municipalities for which no observations are available.

To improve the municipal estimates, a model-based small area estimation (SAE) strategy will be adopted in the near future (Boonstra et al., 2011; Boonstra and Michiels, 2013). For this purpose, a linear models with random municipality effects are considered, in particular, the Battese-Harter-Fuller unit-level model. Several registrations provide auxiliary information at the unit level from which covariates are selected to use in this model.

The continuous nature of the LFS allows to borrow strength not only from other areas, but also over time. A structural time-series model is already being used to estimate national monthly labour status for 6 sex by age classes (van den Brakel and Krieg, 2009). We would like to combine the borrowing of strength over time of the time-series model with borrowing of strength over space of cross-sectional small area models to further improve estimates of municipal unemployment. In Boonstra et al. (2011) a time-series extension of an area-level SAE model was used to estimate municipal unemployment fractions for a sequence of 9 years. The time-series model was fit to direct annual municipal unemployment estimates based on data from the first wave of the LFS rotating panel only. Here we extend that study to a quarterly frequency, and at the same time use data from all waves of the rotating panel design. This requires two generalizations of the previously used time-series model: first it is necessary to account for autocorrelations among the direct estimates implied by the rotating panel design, and second we need to address the systematic biases of direct estimates based on follow-up waves relative to first wave estimates.

The time-series multi-level model used to borrow strength over space and time includes a combination of several effects: fixed (or unmodeled) effects for a selection of demographic predictors, quarter and possibly period, and random (or modeled) area effects for municipalities and area-by-period interactions. Period main effects can be modeled as fixed effects because the total sample size per period is large. 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 (2008), see also Rao (2003) for an overview.

The time-series multi-level model is fit using a Gibbs sampler. Models with different combinations of fixed and random effects are compared to each other based on the Deviance Information Criterion (DIC). The estimates resulting from the time-series model are also compared to direct estimates and to cross-sectional estimates that only borrow strength over space, i.e. across municipalities.

This report is structured as follows. In Section 2 the LFS data as used in this study are described. Section 3 describes small area estimation based on cross-sectional models, i.e. estimation for each period separately. Section 4 discusses the multi-level time-series models considered. The models used are assessed in Section 5. In Section 6 the results from time-series and cross-sectional models are compared. Section 7 contains a discussion of the results as well as some ideas on further work. Details of the Gibbs sampling procedure used to fit the time-series model can be found in I.

2 Data

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 first wave of the panel, data are collected by means of computer assisted personal interviewing (CAPI), whereas in the four follow-up waves, data are collected by means of computer assisted telephone interviewing (CATI). For a more detailed description of the sampling design, we refer to Boonstra et al. (2008).

Fig. 1 illustrates the rotating panel design for five subsequent quarters. The figure shows that the waves that constitute a quarterly dataset are independent in the sense that they are composed of different households. Between subsequent quarters there is an overlap of four waves, inducing positive sampling autocorrelation. The sampling autocorrelation decreases with the separation between quarters and vanishes for periods more than four quarters apart.

Unfortunately, the different waves give rise to systematic differences in unemployment estimates. These differences, generally termed rotation group bias (Bailar, 1975) when viewed relative to the first wave, have many possible causes, including selection, mode and panel effects, see van den Brakel and Krieg (2009).

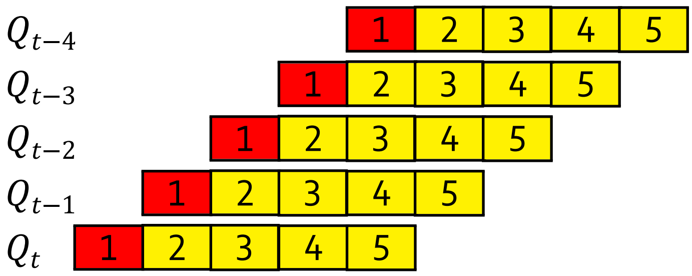

Figure 1 LFS rotating panel design for quarters $t-4,\ldots,t$. The first wave is conducted through CAPI (red) whereas the second to fifth waves are conducted through CATI (yellow). Vertically aligned squares represent observations from the same households (barring panel attrition) at different quarters. Horizontally aligned squares combine into a quarterly dataset.

In the present study we use 24 quarters of LFS data from 2003 to 2008. Data from all five waves of the rotating panel are used. For now, we have simplified the analysis by considering only those municipalities that are not affected by mergers or other administrative changes between 2003 and 2008. Consequently, of the 441 municipalities that existed in 2008, only a subset of 414 are retained.

Quarterly national sample sizes vary between 16 and 22 thousand in the first wave and between 10 and 16 thousand in the fifth wave. Municipal sample sizes are very diverse, ranging from 0 to over 800 for single wave quarterly samples.

LFS data are available at the level of units, i.e. persons. A wealth of auxiliary data from several registrations is also available at the unit level. Among these auxiliary variables is registered unemployment, a strong predictor for the unemployment variable of interest.

3 Cross-sectional small area estimation

In this section we discuss the estimation of municipal unemployment based on data of a single quarter. The cross-sectional models considered are the basic area and unit-level models, also known as Fay-Herriot (Fay and Herriot, 1979) and Battese-Harter-Fuller (Battese et al., 1988) models, respectively, see Rao (2003) for an overview.

Let y_{ij} denote the variable of interest for unit j in area (municipality) i. The total number of areas is m_A, and area i has population and sample sizes N_i and n_i, respectively. Note that in this section we suppress time indices, since the models considered here are applied to the data of each period separately. The quantities of interest θ_i are the area population fractions of unemployed, $\theta_i = \frac{1}{N_i}\sum_{j=1}^{N_i} y_{ij}$.

The basic unit-level model expresses y_{ij} as

$$y_{ij} = \beta' x_{ij} + v_i + \epsilon_{ij}\,, \qquad \epsilon_{ij} \overset{iid}{\sim} N(0,\sigma^2)\,, \qquad v_i \overset{iid}{\sim} N(0,\sigma_v^2)\,, \tag{1}$$

where x_{ij} is a vector of covariates known for the whole population, β are the corresponding coefficients, v_i are area effects, σ^2 is the within-area variance parameter, and σ_v^2 is the between-area variance parameter. Although the variable of interest, unemployment, is a binary variable, we model it using a gaussian linear model, mainly for practical reasons. See Boonstra et al. (2011) for some arguments towards justifying the use of a linear model in this situation.

As an alternative, the basic area-level model is specified at the level of municipalities. First, initial estimates $\hat{\theta}_i$ are computed using a direct estimation method that does not smooth over areas. Corresponding variance estimates ψ_i are also computed. The $(\hat{\theta}_i, \psi_i)$ constitute the data at the area level and they are modeled by the Fay-Herriot model,

$$\hat{\theta}_i = \theta_i + e_i\,, \qquad e_i \overset{ind}{\sim} N(0,\psi_i)\,, \tag{2}$$

$$\theta_i = \beta' \bar{X}_i + v_i\,, \qquad v_i \overset{iid}{\sim} N(0,\sigma_v^2)\,, \tag{3}$$

where β is a vector of regression coefficients corresponding to the vector of auxiliary area-level population means $\bar{X}_i$, and σ_v^2 is the between-area variance parameter of the area effects v_i. The area-level model is more restricted in the number of predictors it can accommodate because of the fewer degrees of freedom at the aggregate level. The initial estimates $\hat{\theta}_i$ are taken to be survey regression estimates (Battese et al., 1988), and ψ_i are corresponding smoothed variance estimates. The survey regression estimates do make use of unit-level auxiliary information to reduce non-response bias. The sampling variances ψ_i are treated as given, so uncertainty about them is not taken into account.

We employ a Bayesian approach to deal with the uncertainty in the between-area variance parameters by integrating over the posterior density for $\frac{\sigma_v^2}{\sigma^2}$ in case of the unit-level model (Datta and Ghosh, 1991), or σ_v^2 in case of the area-level model, see e.g. Rao (2003). Hierarchical Bayes estimates thus obtained based on both unit and area-level models are computed with R (R Development Core Team, 2009) using package hbsae (Boonstra, 2012).

3.1 Combining data from all waves

Within a single quarter, data from the five waves can be considered independent. However, estimates based on follow-up waves are biased relative to first wave estimates. To account for these relative biases, wave effects are added to the model. This approach is followed in Boonstra and Michiels (2013). The relative biases are treated as measurement errors, although they may be partly due to selection effects not explained by the unit-level auxiliary variables used.

For the purpose of extending the unit-level model (1) with a measurement error term we distinguish between the true variable of interest with components Y_{ij} and measured values y_{ij}. The model including a measurement error term can now be written as

$$\begin{aligned} y_{ij} &= Y_{ij} + \alpha' z_{ij} + e_{ij}\,, \qquad e_{ij} \overset{iid}{\sim} N(0, \sigma_m^2)\,, \\ Y_{ij} &= \beta' x_{ij} + v_i + \epsilon_{ij}\,, \qquad \epsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)\,, \qquad v_i \overset{iid}{\sim} N(0, \sigma_v^2)\,, \end{aligned} \tag{4}$$

where the first line represents the measurement part of the model. We simplify this model by assuming $\sigma_m^2 = 0$, so that the measurement error is a bias of size $\alpha' z_{ij}$ depending on known variables z_{ij} with unkown coefficients α. It is assumed that z does not predict the true variable Y, but only the measurement bias in the observed values y. As we take z_{ij} to be the vector of dummy variables for the second to fifth waves, such an assumption is warranted. The corresponding coefficients can be interpreted as the rotation group bias relative to the first wave, controlling for the covariates x.

From (4) with $\sigma_m^2 = 0$ it follows that

$$y_{ij} = \alpha' z_{ij} + \beta' x_{ij} + v_i + \epsilon_{ij}\,, \qquad \epsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)\,, \qquad v_i \overset{iid}{\sim} N(0, \sigma_v^2)\,,$$

so that the likelihood function of the model parameters is exactly the same as for the model without measurement error but with covariate vector $(z', x')'$ and coefficients $(\alpha', \beta')'$. Prediction is based on x only, so the coefficient α is not involved directly in prediction, only through its indirect effects on the partial regression coefficients β.

For the basic area-level model, we deal with rotation group bias through the initial estimates. That is, the survey regression estimates are based on a unit-level regression model including regressors for waves 2 to 5 with zero area population means.

4 Time-series small area estimation

To fit a model to all unit-level data of all periods at once would be very challenging, not only because of the size of the combined dataset, but also because such a model would become quite complex since it must account for many effects at once. In addition to area and time effects the model would also need to include person or household effects to account for repeated measurements of the same persons/households. We therefore specify the time-series model at the area, i.e. municipal, level. This means that the data for the time-series model consist of a set of initial estimates at the area level, computed separately for each time period.

Modeling and estimation are then divided into two stages. In the first stage, described in Section 4.1, initial estimates are computed, using auxiliary information at the unit level to reduce non-response bias. The initial estimates are accompanied by variance estimates as well as estimates of covariances between estimates that are dependent due to the rotating panel design. In the second stage a multi-level time-series model is applied to smooth the initial estimates and to reduce rotation group bias. As in (Rao and Yu, 1994; Datta et al., 1999; You, 2008), we use multi-level time-series models that extend the basic area-level model by including area-by-period random effects in addition to area random effects. The models considered contain a combination of random area effects, independent area-by-period effects and area-by-period effects modeled as random walks over time, as explained further in Section 4.2.

4.1 Initial estimates

Let $\hat{\hat{Y}}_{itp}$ denote the initial estimate for area i and period t based on data from wave p. The initial estimates used as input for the time-series small area model are survey regression estimates (Battese et al., 1988)

$$\hat{\hat{Y}}_{itp} = \bar{y}_{itp} + \hat{\beta}'_{tp}(\bar{X}_{it} - \bar{x}_{itp}), \tag{5}$$

where $\bar{y}_{itp}$, $\bar{x}_{itp}$ denote sample means, $\bar{X}_{it}$ is the vector of population means of the covariates x, and $\hat{\beta}_{tp}$ are estimated regression coefficients. The coefficients are estimated separately for each period and each wave, but they are based on the national samples combining data from all areas.

Even though the regression coefficients used in (5) are not area-specific, the survey regression estimate for a particular area is a direct estimate in the sense that it is primarily based on the data obtained in that area, i.e. on $\bar{y}_{itp}$. The second term in (5) is a correction term that reduces bias due to selection effects, i.e. differences between the observed data and population with regard to the distribution of auxiliary variables x

Table 1 Available covariates

variable	categories
sex	male, female
age3	15-24, 25-44, 45-64
age5	15-24, 25-34, 35-44, 45-54, 55-64
hhtype	single, household with children, other
ethn	native, Western immigrant, non-Western immigrant
prov	12 provinces
ru	registered unemployed or not
ru5	not ru, ru with job, ru $<$ 1 yr, ru 1-4 yrs, ru $>$ 4 yrs

used in the model. The available auxiliary variables are listed in Table 1. The model selected to compute the survey regression estimates is

$$ru \times (sex + age3 + ethn + prov) + ru5 + sex \times age5 + hhtype\,. \qquad (6)$$

Note the prominent role played by registered unemployment; it is a strong predictor for the unemployment variable of interest.[1)]

Fig. 2 illustrates the rotation group biases, i.e. the average differences between estimates based on different rotation groups or waves. Displayed are sample means and survey regression estimates averaged over all periods and areas, weighted by area population sizes. The use of auxiliary information in the survey regression estimates according to model (6) reduces selection effects, giving rise to an increase in unemployment estimates. The effect of using auxiliary information grows with wave number, which makes sense since overall response rates decrease with wave number. The survey regression estimates based on the four follow-up waves are on approximately the same level, still well below the first wave level. To a large extent this remaining difference may well be due to measurement effects. This is further supported by the fact that panel attrition is limited, with response fractions close to 90% in waves two to five, given initial response in the first wave. Fig. 3 displays the survey regression estimates averaged only over areas, and shows the variation over time of the rotation group biases.

The time-series model also requires variance estimates corresponding to the initial estimates. We use cross-sectionally smoothed variance estimates

$$v(\hat{\hat{Y}}_{itp}) = \widehat{\sigma^2_{tp}}/n_{itp}\,, \qquad (7)$$

based on the estimated variance $\widehat{\sigma^2_{tp}}$ of the regression residuals and the area-specific sample sizes n_{itp}. The survey regression estimates are computed as BLUP predictors based on the basic unit-level model where the between-area variance parameter is fixed at a large value. Thus the regression model includes area effects, and $\widehat{\sigma^2_{tp}}$ are smoothed estimates of residual within-area variance.

1) We use the term registered unemployed for all persons registered with the public employment service (in The Netherlands, UWV Werkbedrijf), irrespective of whether they have a (part-time) paid job, whether they are actively searching for a job, or whether they are available for work in the short term.

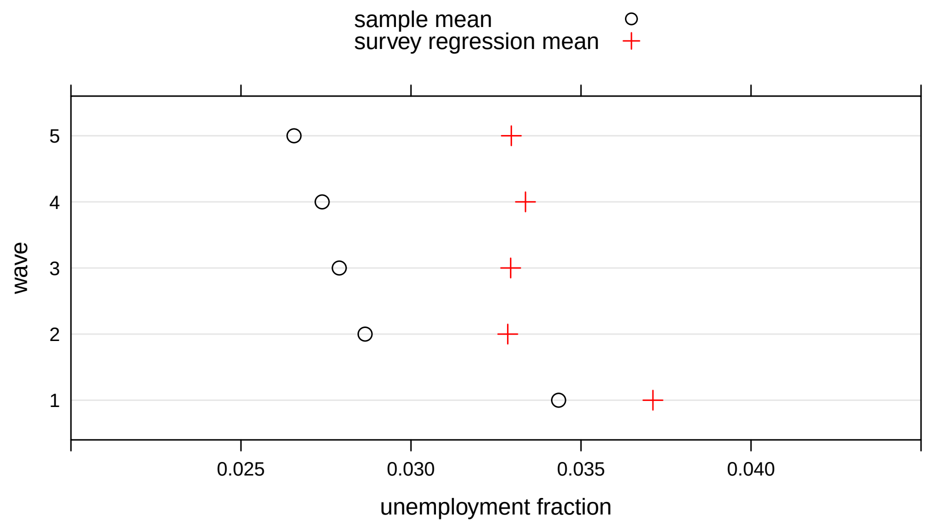

Figure 2 Systematic differences between estimates based on different waves (rotation group biases): sample means vs. survey regression means aggregated over periods and areas.

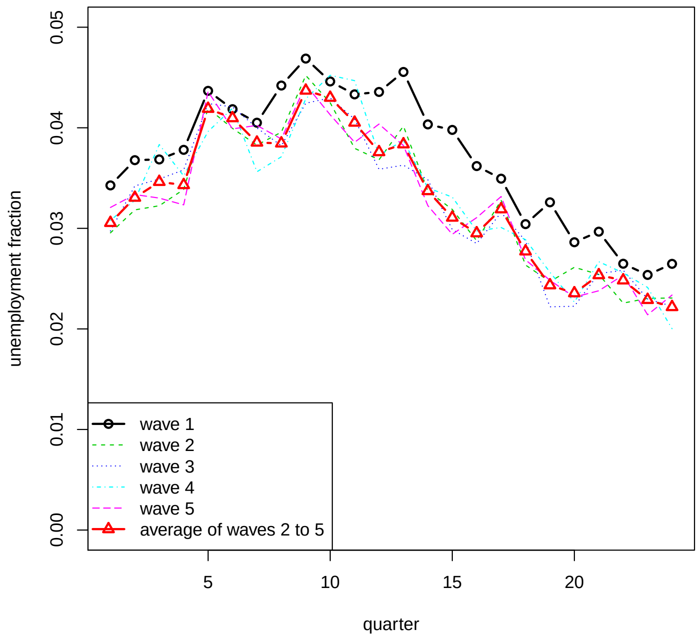

Figure 3 Systematic differences between survey regression estimates based on different waves over time. The estimates are aggregated over areas.

A complication is the fact that in some periods there are no observations within some small municipalities for one, several, or even all waves. This missing data problem is addressed by using, instead of survey regression estimates, synthetic estimates $\hat{\beta}'_{tp}\bar{X}_{it}$ for cases where $n_{itp} = 0$. To make sure that these estimates have negligible influence the corresponding variance estimates are set to relatively very large values. This way of dealing with missing data has the advantage of not requiring any adaptation of the model or fitting procedure.

The initial estimates $\hat{\hat{Y}}_{itp}$ are combined into a vector $\hat{\hat{Y}} = (\hat{\hat{Y}}_{111}, \hat{\hat{Y}}_{112}, \dots, \hat{\hat{Y}}_{115}, \hat{\hat{Y}}_{121}, \dots)'$, i.e., wave index runs faster than time index which runs faster than area index. The numbers of areas, periods and waves are denoted by m_A, m_T and m_P, respectively. The total length of $\hat{\hat{Y}}$ is therefore $m = m_A m_T m_P = 414(areas) * 24(quarters) * 5(waves) = 49680$. Similarly, the variance estimates $v(\hat{\hat{Y}}_{itp})$ are put in the same order along the diagonal of a $m \times m$ covariance matrix Φ.

Due to the rotating panel design, which induces several non-negligible correlations among initial estimates for the same municipality and different periods and waves, the covariance matrix Φ is not diagonal, although it is still very sparse. 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 municipality in quarters t_1, t_2 and based on waves p_1, p_2 whenever $t_2 - t_1 = p_2 - p_1 \leq 4$, see Fig. 1. The covariances between $\hat{\hat{Y}}_{it_1p_1}$ and $\hat{\hat{Y}}_{it_2p_2}$ are estimated as (see e.g. Kish (1965))

$$v(\hat{\hat{Y}}_{it_1p_1}, \hat{\hat{Y}}_{it_2p_2}) = \frac{n_{it_1p_1t_2p_2}}{\sqrt{n_{it_1p_1} n_{it_2p_2}}} \hat{\rho}_{t_1p_1t_2p_2} \sqrt{v(\hat{\hat{Y}}_{it_1p_1}) v(\hat{\hat{Y}}_{it_2p_2})}\,, \tag{8}$$

where $n_{it_1p_1t_2p_2}$ is the number of units in the overlap, i.e. the number of observations on the same units in area i between period and wave combinations (t_1, p_1) and (t_2, p_2). The estimated (auto)correlation coefficient $\hat{\rho}_{t_1p_1t_2p_2}$ is computed as the correlation between the residuals of the linear regression models underlying the survey regression estimators at (t_1, p_1) and (t_2, p_2), based on the overlap of both samples over all areas. So they are smoothed over areas, just as the $\widehat{\sigma^2_{tp}}$.

The covariance matrix Φ thus formed is a band matrix, and the ordering of the vector $\hat{\hat{Y}}$ is such that it achieves minimum possible bandwidth, which is advantageous from a numerical computational point of view. To illustrate, the pattern of non-vanishing elements of the top-left 150×150 block of Φ is displayed in Fig. 4. This part of the covariance matrix refers to all estimates for the first municipality, and a small number of estimates for the second municipality. The figure clearly shows the correlation structure induced by the rotating panel design, as well as independence between municipalities, resulting in a block diagonal matrix with a relatively small number of nonzero bands around the diagonal.

Fig. 5 shows the distribution of autocorrelations implied by Φ. The densities are shown separately for correlations between initial estimates 1 to 4 quarters apart. A larger separation in time generally means smaller correlation because of 1) smaller overlap $\frac{n_{it_1p_1t_2p_2}}{\sqrt{n_{it_1p_1} n_{it_2p_2}}}$ due to panel attrition and 2) smaller autocorrelation coefficient $\hat{\rho}_{t_1p_1t_2p_2}$, see equation (8).

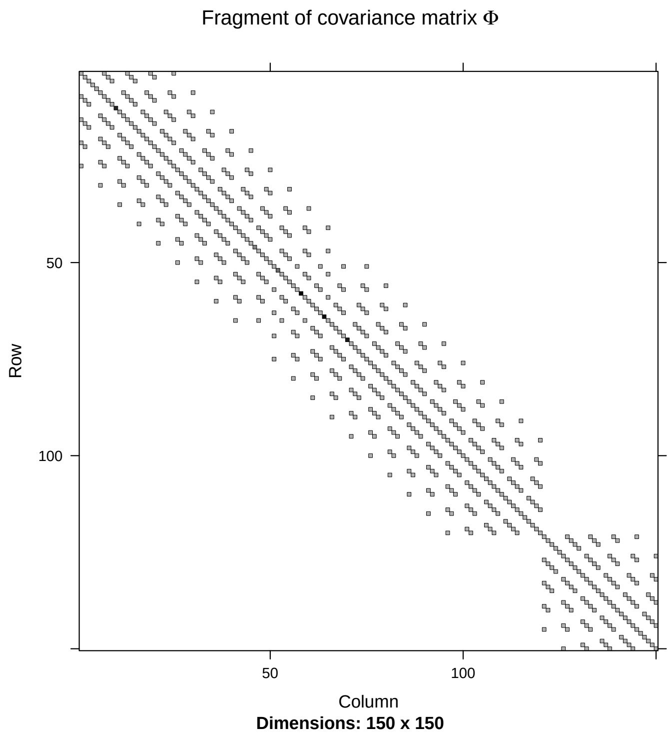

Figure 4 **The top-left** 150×150 **block of** Φ**.**

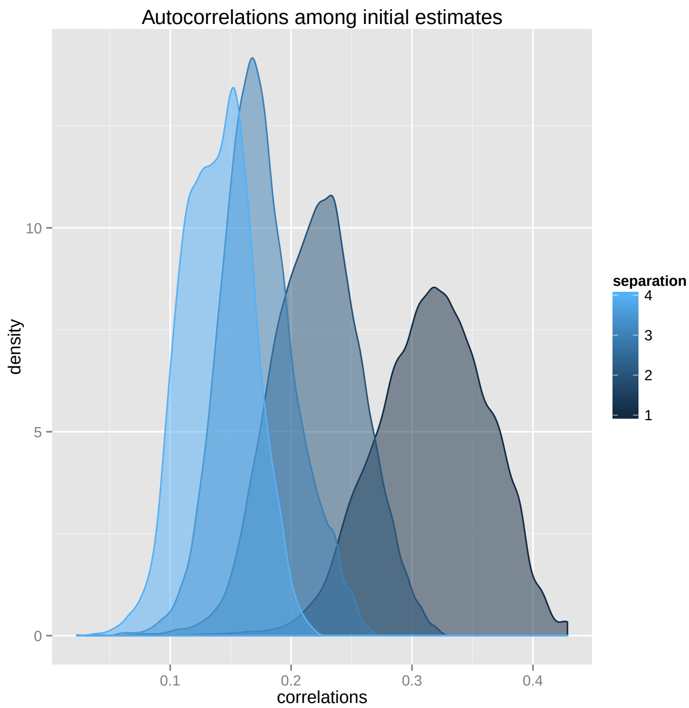

Figure 5 **Distribution of non-zero autocorrelations between initial estimates, grouped by the separation in time (quarters).**

4.2 Time-series multi-level model

The initial estimates $\hat{\hat{Y}}$ are subsequently modeled by a time-series multi-level model. The model used is an extension of the basic area-level model over time, and is given by

$$\begin{aligned} \hat{\hat{Y}}_{itp} &= \theta_{it} + \alpha' z_{itp} + e_{itp} , \\ \theta_{it} &= \beta' \bar{X}_{it} + \nu_{it} , \end{aligned} \tag{9}$$

where $\alpha' z_{itp}$ denotes a measurement bias term depending on known variables z_{itp} and unknown coefficients α. There are 5 measurements $\hat{\hat{Y}}_{itp}, p = 1, \dots, 5$ for each small area estimand θ_{it}. The sampling errors $e = (e_{111}, e_{112}, \dots, e_{115}, e_{121}, \dots)'$ are distributed as $e \sim N(0, \Phi)$ where Φ is the covariance matrix for the initial estimates as discussed in Section 4.1. The vector $\bar{X}_{it}$ of predictors consists of population means known from registrations. Finally, ν_{it} is an error term composed of multiple random effects.

The random effects term is taken to be a combination of some or all of 1) independent area effects v_i, 2) independent area-by-time effects w_{it}, and 3) area-by-time effects u_{it} modeled as random walks over time. So for the most comprehensive model, $\nu_{it} = v_i + w_{it} + u_{it}$ with

$$v_i \overset{iid}{\sim} N(0, \sigma_v^2) , \tag{10}$$

$$w_{it} \overset{iid}{\sim} N(0, \sigma_w^2) , \tag{11}$$

$$u_{it} \sim u_{i;t-1} + \varepsilon_{it} \;\; with \;\; \varepsilon_{it} \overset{ind}{\sim} N(0, \sigma_u^2) . \tag{12}$$

For each area i, the u_{it} follow a random walk model over time with a common variance σ_u^2. Each term may bring additional explanatory power. The combination of area and area-by-time effects, for example, would fit the data much better in case the variation over areas is of a different magnitude than the variation over area-by-time combinations. Note that the random effects for different areas are independent, i.e. no explicit spatial correlation is included in the models. Datta et al. (1999) and You (2008) use time-series models with area effects v_i and random walks u_{it}. In Rao and Yu (1994) a model is used with random area effects v_i and a stationary autoregressive AR(1) model instead of a random walk model for u_{it}. 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. Whereas in some respects our model is simpler than some of the models considered in the above-mentioned references, a novel feature of our model is that it includes the independent area-by-time effects w_{it}, in addition to the area-by-time random walk effects u_{it} and area effects v_i. As our application to LFS data is of a large scale, we use an efficient block Gibbs sampling algorithm that uses sparse matrix algebra to fit the model.

Combining the measurement and structural parts of model (9), we have

$$\hat{\hat{Y}}_{itp} = \alpha' z_{itp} + \beta' \bar{X}_{it} + \nu_{it} + e_{itp} . \tag{13}$$

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 models considered belong to the class of Gaussian Markov Random Field (GMRF) models, and we make use of the sparse matrix and block sampling techniques described in Rue and Held (2005) for efficiently fitting such models to large datasets. See I for details.

The samples generated ultimately converge to draws from the posterior distribution for the model parameters of model (9). Specifications of the prior distributions for all model parameters can be found in I, along with derivations of the full conditional distributions required for the Gibbs sampler. For the standard deviation parameters σ_v, σ_w and σ_u we use priors from the folded-t family of distributions (Gelman, 2006). These priors are implied by a multiplicative redundant parameterization (Gelman et al., 2008), which helps the Gibbs sampler converge more quickly.

For each model considered, the Gibbs sampler is run in five 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 fifth draw to save memory while reducing the effect of autocorrelation between successive draws. This leaves $5 * 400 = 2000$ draws used to compute small area 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 diagnostics suggest that all chains converge to the posterior distribution well within 500 draws. Also, the estimated Monte Carlo simulation errors (accounting for the remaining autocorrelation in the chains) were small compared to the posterior standard errors for all parameters, indicating that the number of retained draws is sufficient for our purposes.

The model smooths the initial estimates and reduces rotation group bias, yielding estimates

$$\hat{\theta}_{it} = \hat{\beta}'\bar{X}_{it} + \hat{v}_{it} . \tag{14}$$

where the 'hats' denote taking the mean with respect to the posterior distribution of the model parameters. The posterior means are approximated by the means of the retained Monte Carlo samples. That is, with R the number of retained samples,

$$\hat{\theta}_{it} = \frac{1}{R}\sum_{r=1}^{R} \theta_{it,r} , \tag{15}$$

where $\theta_{it,r} = \beta_r'\bar{X}_{it} + v_{it,r}$ based on the rth sample of the model parameters. Similarly, the variance

$$\mathrm{mse}(\hat{\theta}_{it}) = \frac{1}{R-1}\sum_{r=1}^{R} (\theta_{it,r} - \hat{\theta}_{it})^2 , \tag{16}$$

of the Monte Carlo samples approximates the posterior variance for θ_{it}. We use the square root of this quantity as measure of uncertainty about θ_{it}.

5 Selecting predictors and random effects for the time-series model

Using the Gibbs sampler, the time-series multi-level model (9) has been fit to the vector of initial estimates for various choices of the measurement-related variables z, predictors $\bar{X}$ and random effects v. The different choices of z and $\bar{X}$ are given in Table 2. Variable wave (5 categories corresponding to wave number p) is associated with rotation group bias, and terms involving this variable (the first term in each model) correpond to z. The other variables are used as predictors and correspond to $\bar{X}$. Note that for the purpose of model fitting there is no need to distinguish between z and $\bar{X}$. In addition to wave and the variables mentioned in Table 1 a few temporal predictors are used in the time-series models: (seasonal) quarter (4 categories, 1 to 4), year (6 categories, 2003 to 2008) and T1, T2, T3 which are standardized linear, quadratic and cubic monomials in time (measured in quarters). Where the models in Table 2 contain interaction terms it is understood that all lower order interactions and/or main effects are included as well. We will refer to the models defined in Table 2 as predictor models, even though the terms involving wave do not predict the true population means but rather the measurement error in the corresponding direct estimates.

Table 2 Fixed effects corresponding to z and $\bar{X}$ used in the time-series multi-level model.

Model nr.	Variables z and $\bar{X}$ used in the model
1	wave + quarter + sex + age5 + ethn + hhtype + prov
2	wave + quarter + sex + age5 + ethn + hhtype + prov + ru5
3	wave + quarter + sex + age5 + ethn + hhtype + prov + T1 + T2 + T3
4	wave + quarter × year + sex + age5 + ethn + hhtype + prov + ru5
5	wave × year + quarter × year + sex + age5 + ethn + hhtype + prov + ru5
6	wave × year + quarter × year × ru + sex + age5 + hhtype + ethn × (ru + age3) + prov + ru5
7	wave × quarter × year + quarter × year × ru + sex + age5 + hhtype + ethn × (ru + age3) + prov + ru5
8	wave × quarter × year + quarter × year × (ru + sex + age5 + ethn)

In Table 3 and Table 4 the results based on the time-series multi-level model with different combinations of fixed and random effects are shown. The first column refers to the predictors included as fixed effects, i.e. the predictor model as defined in Table 2. The second column lists the random effect terms included in the model, and columns three to five contain the posterior means of the corresponding standard deviation parameters. The next two columns contain the effective number of model parameters p_{eff} and the Deviance Information Criterion (DIC), as defined in I.4. Lower DIC values correspond to a better compromise between model fit and model complexity. For easier comparison, all DIC values have been shifted downward by the DIC value of the first model. The last two columns contain information about the (posterior) relative root mean squared errors of the vector θ of interest. The mean relative root mean squared error (RRMSE) is defined as the average of

$$\frac{\sqrt{\text{mse}(\hat{\theta}_{it})}}{\hat{\theta}_{it}} \tag{17}$$

over all areas and quarters. The last column lists the fractions of relative errors (17) less than 0.2. This is the fraction of the $m_A m_T = 9936$ estimates that would get published under a publication strategy saying that only figures with relative errors less than 0.2 are to be published.

Table 3 Time-series multi-level model results for different combinations of fixed and random effects, part 1: predictor models 1-4.

predictor model	random effects	σ_v ×100	σ_u ×100	σ_w ×100	p_{eff}	DIC	mean RRMSE	fraction with RRMSE $<$ 0.2
1		0	0	0	28	0	0.03	1.00
1	v	0.51	0	0	279	-666	0.11	0.95
1	u	0	0.32	0	1042	-1303	0.16	0.76
1	w	0	0	0.49	1623	-689	0.16	0.79
1	vu	0.37	0.31	0	1206	-1686	0.19	0.63
1	vw	0.47	0	0.41	1452	-1031	0.17	0.71
1	wu	0	0.31	0.17	1212	-1315	0.17	0.74
1	vuw	0.37	0.31	0.05	1229	-1681	0.19	0.63
2		0	0	0	32	-2052	0.03	1.00
2	v	0.21	0	0	138	-2133	0.07	1.00
2	u	0	0.13	0	326	-2171	0.09	0.99
2	w	0	0	0.26	603	-2145	0.09	1.00
2	vu	0.20	0.13	0	413	-2240	0.11	0.97
2	vw	0.20	0	0.24	613	-2204	0.11	0.99
2	wu	0	0.12	0.20	602	-2197	0.11	0.98
2	vuw	0.20	0.12	0.17	626	-2257	0.12	0.95
3		0	0	0	31	-1634	0.03	1.00
3	v	0.35	0	0	220	-2010	0.10	0.97
3	u	0	0.14	0	340	-1775	0.09	0.96
3	w	0	0	0.30	754	-1777	0.11	0.96
3	vu	0.35	0.14	0	527	-2144	0.13	0.88
3	vw	0.35	0	0.23	664	-2067	0.12	0.92
3	wu	0	0.12	0.25	764	-1835	0.12	0.92
3	vuw	0.35	0.13	0.15	690	-2153	0.14	0.86
4		0	0	0	52	-2265	0.03	1.00
4	v	0.18	0	0	136	-2312	0.07	1.00
4	u	0	0.10	0	237	-2312	0.08	1.00
4	w	0	0	0.21	458	-2310	0.08	1.00
4	vu	0.17	0.10	0	308	-2353	0.09	0.99
4	vw	0.17	0	0.20	475	-2345	0.09	1.00
4	wu	0	0.09	0.18	470	-2331	0.09	0.99
4	vuw	0.17	0.08	0.16	500	-2365	0.10	0.98

Table 4 Time-series multi-level model results for different combinations of fixed and random effects, part II: predictor models 5-8.

predictor model	random effects	σ_v ×100	σ_u ×100	σ_w ×100	p_{eff}	DIC	mean RRMSE	fraction with RRMSE $<$ 0.2
5		0	0	0	72	-2293	0.04	1.00
5	v	0.18	0	0	157	-2339	0.07	1.00
5	u	0	0.10	0	255	-2339	0.08	1.00
5	w	0	0	0.22	480	-2340	0.08	1.00
5	vu	0.17	0.10	0	326	-2382	0.09	0.98
5	vw	0.17	0	0.20	499	-2374	0.10	1.00
5	wu	0	0.08	0.18	487	-2356	0.09	0.99
5	vuw	0.17	0.08	0.16	522	-2389	0.11	0.97
6		0	0	0	101	-2398	0.04	1.00
6	v	0.15	0	0	167	-2424	0.07	1.00
6	u	0	0.06	0	190	-2408	0.06	1.00
6	w	0	0	0.16	343	-2411	0.07	1.00
6	vu	0.15	0.06	0	248	-2433	0.08	1.00
6	vw	0.15	0	0.14	356	-2432	0.08	1.00
6	wu	0	0.05	0.14	355	-2415	0.08	1.00
6	vuw	0.15	0.05	0.12	383	-2435	0.09	1.00
7		0	0	0	173	-2349	0.05	1.00
7	v	0.15	0	0	238	-2376	0.07	1.00
7	u	0	0.06	0	262	-2358	0.07	1.00
7	w	0	0	0.16	415	-2363	0.08	1.00
7	vu	0.15	0.06	0	323	-2381	0.09	1.00
7	vw	0.15	0	0.15	440	-2386	0.09	1.00
7	wu	0	0.05	0.14	418	-2365	0.08	1.00
7	vuw	0.15	0.05	0.12	449	-2385	0.10	0.99
8		0	0	0	312	-2068	0.06	1.00
8	v	0.23	0	0	436	-2184	0.10	0.99
8	u	0	0.08	0	440	-2090	0.09	0.99
8	w	0	0	0.18	596	-2088	0.09	0.99
8	vu	0.23	0.08	0	570	-2211	0.11	0.96
8	vw	0.23	0	0.13	596	-2190	0.11	0.98
8	wu	0	0.07	0.14	607	-2096	0.10	0.99
8	vuw	0.23	0.08	0.08	638	-2211	0.12	0.96

The predictor models 1-8 range from parsimonious to comprehensive. For the models with no random effect terms the number p_{eff} equals the number of predictors. According to the DIC model selection criterion, the best models of those listed in Table 3 and Table 4 are the models based on predictor model 6. The simpler predictor models (1-5) seem to be worse than model 6 in terms of model fit. Predictor models 7 and 8, on the other hand, seem to overfit the data.

Each predictor model has been combined with all 8 possible combinations of random effects v, u and w. Recall that v denote independent area effects, u denote area-by-time effects that are random walks over time, and w are independent area-by-time effects. The results show that it is important to include area effects v in combination with at least one of the area-by-time components u and w; such models have the lowest DIC values. For predictor model 1, the best combination of random components is $v + u$; adding w leads to a slightly larger DIC value. For the other predictor models, it is better to include all three random components, although for the largest predictor models 6, 7 and 8 similar DIC values are obtained with the combinations $v + w$ and/or $v + u$. As the predictor model gets stronger the random effect terms become less important, as is to be expected. The sizes of all random components decrease as the predictor model is expanded, as can be seen from the σ_v, σ_u and σ_w columns. In predictor model 8 the sizes of v and (increments of) u are larger than in predictor model 6, but this may be because of overfitting: the model is fitting part of the noise as well.

Predictor model 1 is the only model without the strong registered unemployment predictor. As a consequence, it yields much higher DIC values than all other predictor models. Predictor model 2 contains registered unemployment, but like model 1 it does not contain temporal fixed effects other than seasonal quarter. Predictor model 3 contains cubic polynomial time effects and all other predictor models contain $quarter \times year$, i.e. a fixed effect for each time period. From this point of view it is not surprising that the random walk effects u, which capture smooth time dependence, are most valuable in conjunction with predictor model 1, followed at some distance by predictor model 2.

Concerning mean squared error, the RRMSE columns in Table 3 and Table 4 show that the inclusion of random effects is important to prevent over-optimistic assessment of uncertainty. In particular, models with fixed effects only yield the smallest mean squared errors. But their fit to the data is much worse and they ignore any variation over areas and time not explained by the predictors. So the MSEs are (severe) underestimates of uncertainty in these cases. The unstructured area-by-time effects w capture residual variation over areas and time. So it may be wise to include this component in addition to area and random walk effects, in order to prevent underestimation of MSEs.

It can be argued also from a substantive point of view that area effects v and at least one of the area-by-time effect terms u and w should be included in the model, because estimation is at the level of area-by-time and one cannot expect the predictors to explain all variation over areas and time. Models including independent effects w instead of random walk effects u have the advantage of being simpler and faster to fit. However, our results, in particular the DIC values, indicate that it is generally safe to include both u and w, besides area effects v; this does not seem to cause overfitting (except perhaps with predictor model 1), and both u and w can contribute to model fit. An effect of including u not shown in the tables is that it rightly causes MSEs at the beginning and end of the time-series to be somewhat larger than in the middle, because less strength can be borrowed at both ends. However, this effect is very small for predictor model 6 because of the small size of the u effects. Based on the DIC criterion and the substantive arguments mentioned, we choose model '6 vuw' as the model to compare against the cross-sectional models in the next section.

6 Comparison of estimates

Here we compare estimates based on the time-series multi-level model with direct survey regression estimates and estimates based on cross-sectional basic area and unit-level models. The time-series model used in the comparison is the one selected in the previous paragraph, i.e. model '6 vuw' that includes predictor model 6 and all three random effect terms. The predictors used in the cross-sectional models have been selected with the help of the conditional AIC value (Vaida and Blanchard, 2005), a criterion very similar to DIC. The survey regression and basic unit-level model estimates both use the following set of predictors:

$$ru \times (sex + age3 + ethn + prov) + ru5 + sex \times age5 + hhtype + ru\_area + wave\,. \tag{18}$$

This is the same predictor model as used for the initial estimates for the time-series model, see (6), except for the last two terms ru_area and $wave$. Of these, ru_area is an area-level predictor defined for each unit as the area population fraction registered unemployed of the area the unit belongs to. As before, $wave$ is not a predictor for the variable of interest but rather for the (average rotation group) bias with which it is measured. In the notation of equation (4) for the basic unit-level model, it is the variable z, whereas the other predictors consitute x.

The basic area-level model uses the survey regression estimates and their smoothed variance estimates as input. The predictors used in the basic area-level model (3) itself are the area population means of

$$ru + sex + age5 + ethn\,. \tag{19}$$

As the number of areas is much smaller than the number of units, this predictor model is necessarily much more parsimonious than (18). Even though the time-series model is also specified at the area level, it can use more predictors than the basic area-level model, because it uses data over 24 quarters. This explains why the predictor model selected for the time-series model, model 6 in Table 2, can be more comprehensive than (19).

Fig. 6 and Fig. 7 show the estimated unemployment fractions and their standard errors for a selection of 8 municipalities over all 24 quarters. The figures compare survey regression estimates and estimates based on the basic unit and area-level models as well as the time-series model. The selected municipalities range from very large (Rotterdam, 403322 persons on average in the target population) to very small (Lith, 4550 persons on average in the target population). For the largest municipalities, the

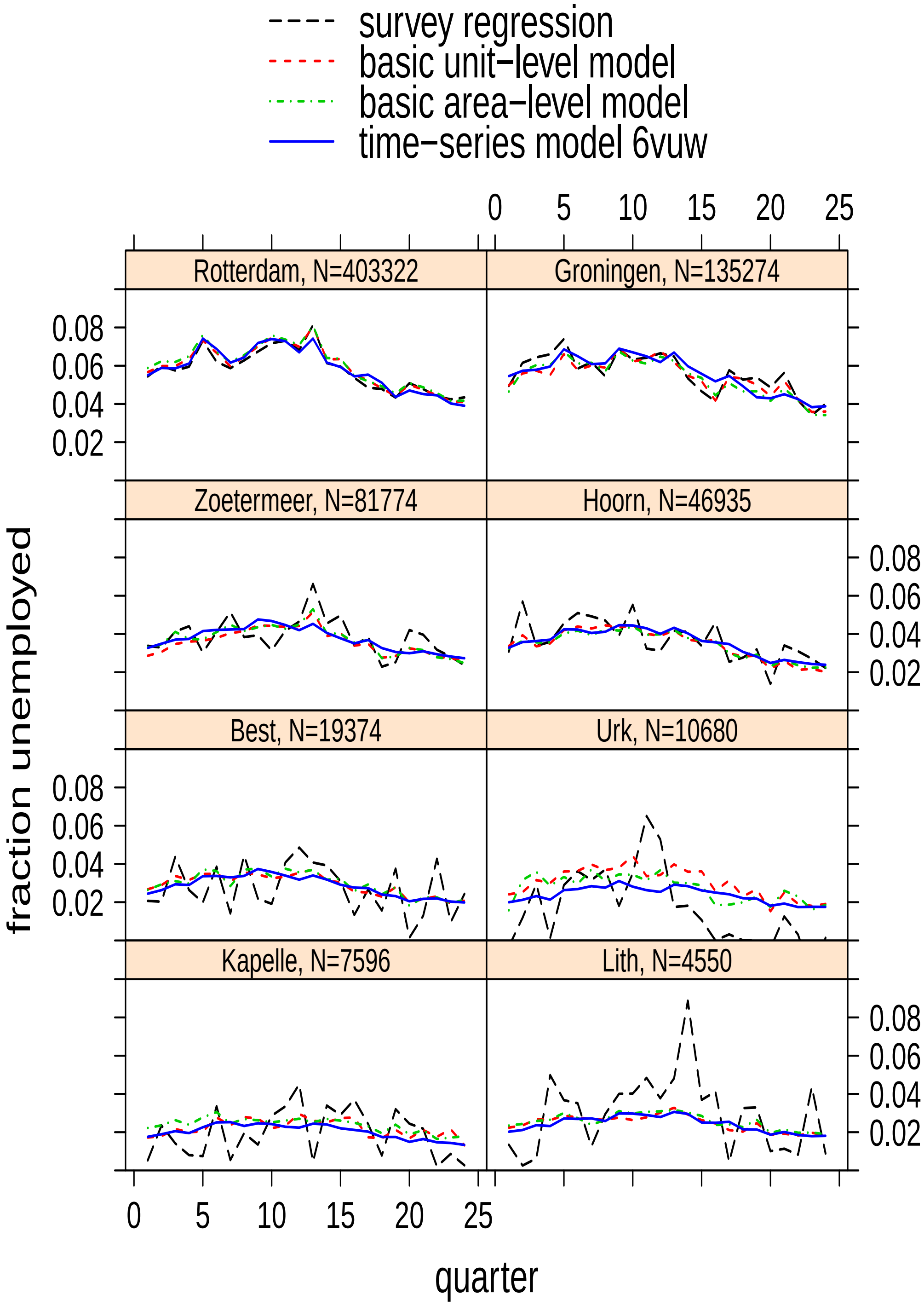

Figure 6 Comparison of estimates of unemployment fraction for 8 selected municipalities over time.

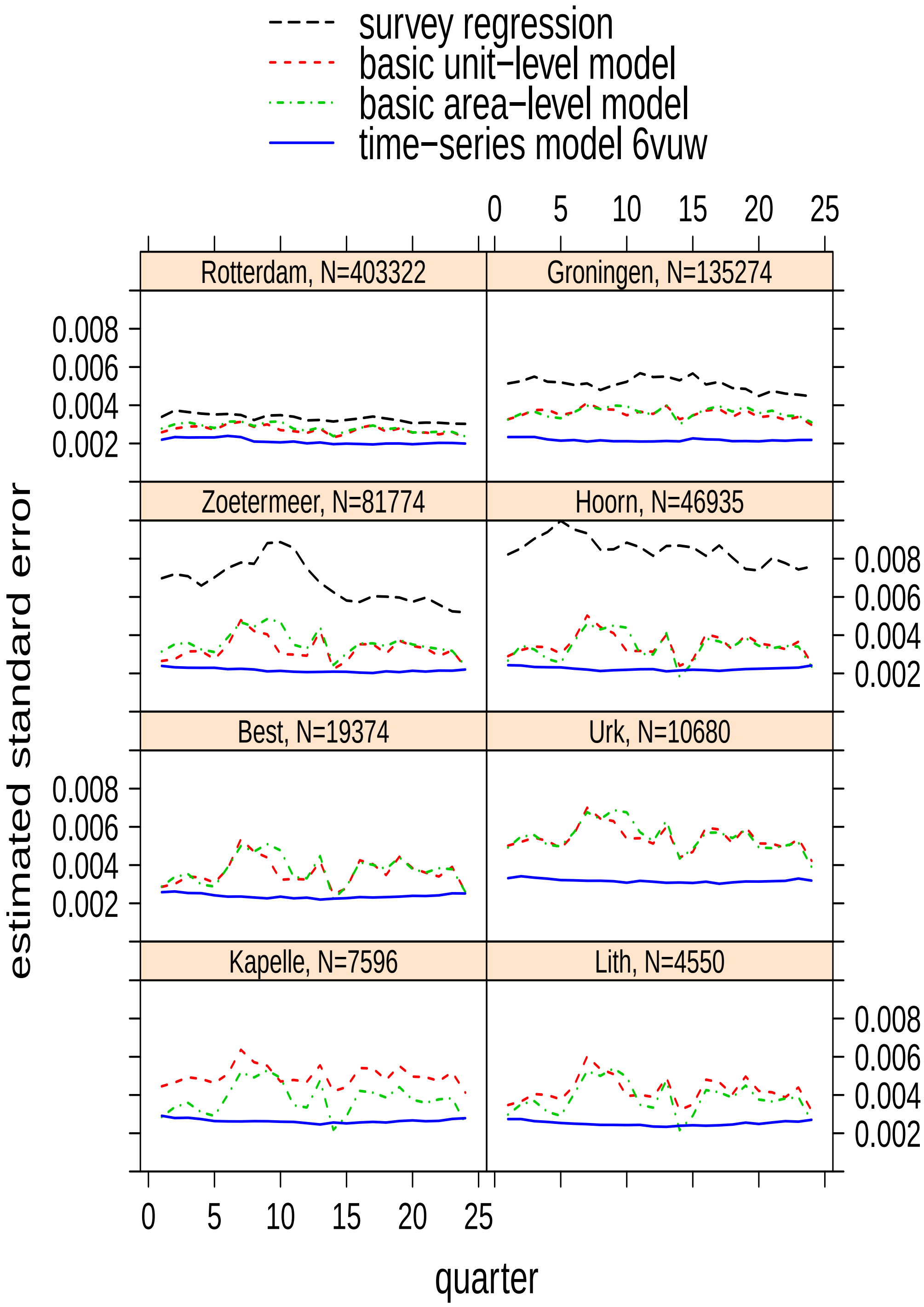

Figure 7 Comparison of standard errors for unemployed fraction for 8 selected municipalities over time.

differences between all four estimation methods are small. The smaller (the sample size within) the municipality, the more unstable the direct survey regression estimates become. Differences between the three model-based estimators are not so large, however. This indicates that much strength is already borrowed cross-sectionally. The time-series estimates are slightly less volatile than the cross-sectional model-based estimates. For the municipality Urk the cross-sectional model-based estimates are systematically above the time-series estimates in the first 15 quarters. This might be due to overshrinkage of the cross-sectional estimates. The time-series estimates, which borrow strength over time for the estimation of area effects, are more in line with the average level of the survey regression estimates over this period.

Fig. 7 shows that standard errors for the cross-sectional and time-series model-based estimates are smaller than those for the survey regression estimates, and the difference increases with decreasing sample size. For the lower four municipalities the standard errors of the survey regression estimates are not even visible as they fall outside the plotting area. The figure also shows that the standard errors based on the time-series model are almost always smaller than those based on the area and unit-level models. The latter are also much more volatile.

Fig. 8 shows for each quarter the relative standard errors averaged over all (equally weighted) municipalities for the three sets of model-based estimates. The survey regression estimates have been left out because their standard errors are so much larger on average and because some of the survey regression estimates themselves are near zero or even negative so that the average relative standard error makes not much sense. In almost all periods the time-series model yields smaller relative standard errors than the cross-sectional models. It is again apparent that the standard errors of the cross-sectional model estimates vary quite a lot from period to period. This variation is also visible in Fig. 9, which shows the fraction of municipalities for which the relative standard error is less than 20%. The 20% figure may be a bound imposed by a publication strategy, so that estimates above the bound do not get published. According to such a strategy, the number of publishable cross-sectional model-based estimates would vary greatly over time, whereas almost all time-series estimates would be publishable. Survey regression estimates would only be publishable for the 10% to 20% largest municipalities.

Fig. 10 compares the national estimates of unemployment obtained by aggregating the municipal estimates. The survey regression and cross-sectional model-based estimates show almost identical aggregate figures. But the estimates based on the time-series model clearly deviate from the other three series. The deviation is quite large over some periods, and turns out to be due to the different way the rotation group bias is modeled. The survey regression and cross-sectional model-based estimates account for the rotation group biases cross-sectionally, i.e. on a quarterly basis, while the time series model '6 vuw' contains wave × year effects and therefore accounts for the bias on a yearly basis. If instead we use a time-series model with quarterly wave effects, such as model 7 of Table 2, we see that the differences almost disappear. This is shown in Fig. 11, which is the same as Fig. 10 except that it displays the time-series estimates based on model '7 vuw'. Although according to the DIC values of Table 4 model 7 does not fit the data as well as model 6, it is perhaps more natural to account for rotation group bias per quarter than per year. Fig. 3 showed that rotation group bias varies quite

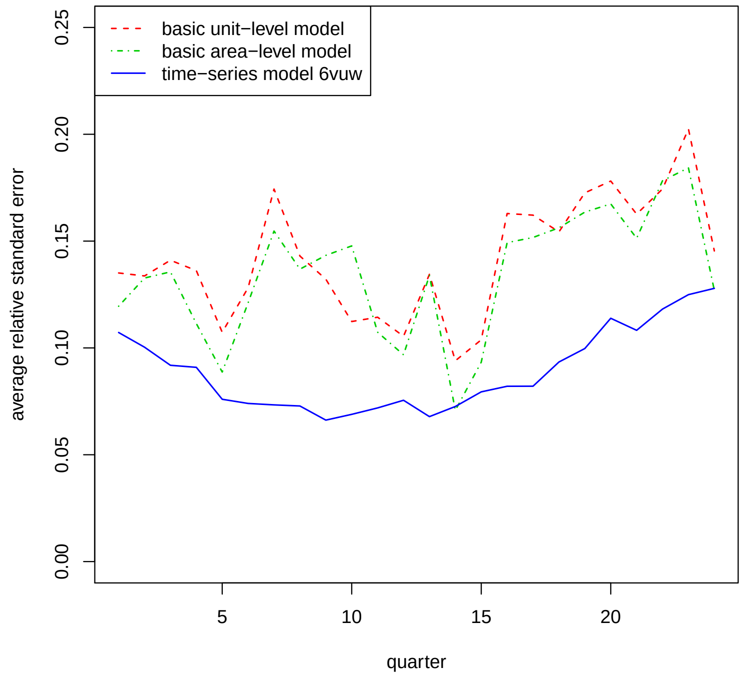

Figure 8 Comparison of relative standard errors averaged over all municipalities.

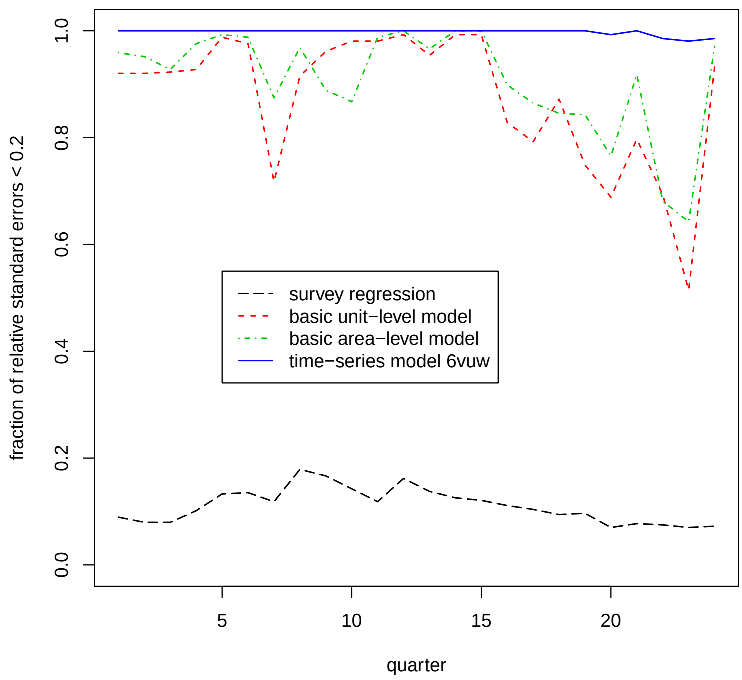

Figure 9 Fractions of estimates with relative error less than 20%.

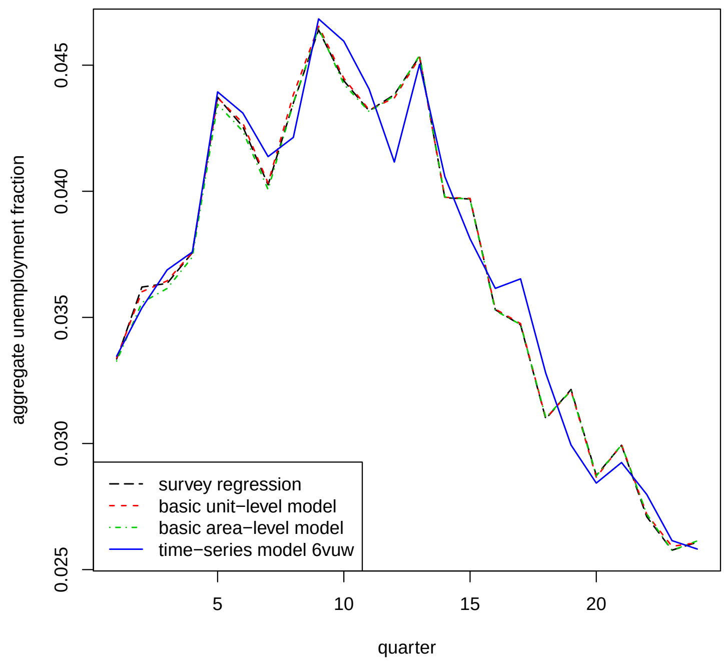

Figure 10 Aggregates of the municipal estimates to the national mean.

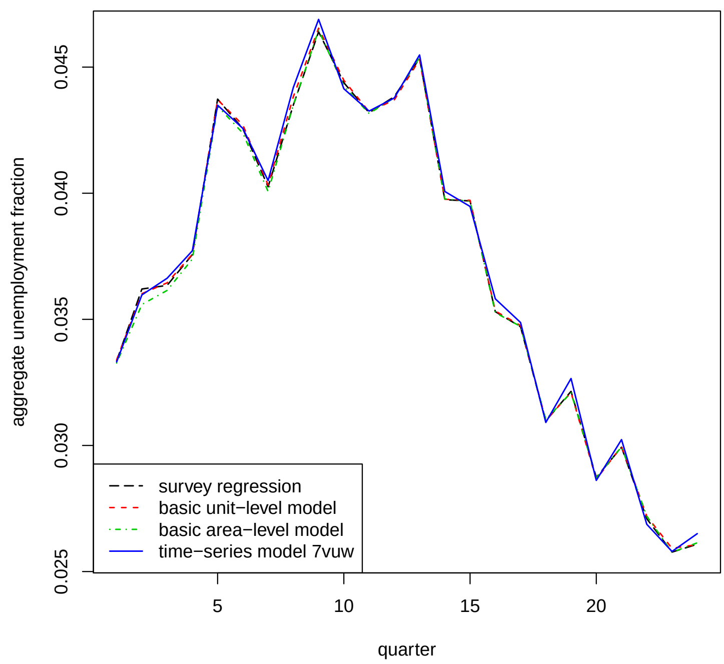

Figure 11 Aggregates of the municipal estimates to the national mean. Here time-series model `7 vuw' is used in the comparison.

a lot over time. Based on these observations it would probably be better to model the rotation group biases as random effects varying over time, perhaps as random walks as in van den Brakel and Krieg (2009).

For completeness the figures displaying the comparisons with cross-sectional estimates have been repeated for model '7 vuw' in II. Differences between the estimates based on models 6 and 7 are small. Standard errors for model 7 are slightly higher than those based on model 6, but they are still almost always smaller than the standard errors for the cross-sectional estimates.

7 Discussion

A time-series small area estimation model has been applied to a large amount of survey data, comprising 6 years of Dutch LFS data, to estimate quarterly unemployment fractions for more than 400 municipalities over this period. Estimation is carried out in two stages in order to limit the computational burden and model complexity. First, the unit-level data is aggregated to the quarter-times-municipality level by means of a survey regression estimator that reduces non-response bias. The Dutch LFS is a rotating panel survey with five waves differing according to their time-in-panel, and initial survey regression estimates are computed for each wave. The resulting set of initial estimates together with estimates of their standard errors and correlations are input for the time-series SAE model, which borrows strength over municipalities and over time to yield improved estimates. Besides, the model accounts for rotation group bias relative to the first wave by modeling it as a measurement error. A Bayesian approach is taken, and the model is fit using a Gibbs sampler. Sparse matrices are used for fast generation of draws from the full conditional distributions, while multiplicative redundant reparameterizations are used to speed up convergence of the sampler.

The time-series model includes fixed effects corresponding to a set of predictors known from registrations, as well as a selection of three random effect terms: independent municipality effects, independent municipality-by-quarter effects and municipality-by-quarter effects modeled as random walks over time. A model comparison using DIC as the main criterion shows that models with municipality effects and at least one of the municipality-by-quarter effects perform best. Models that contain strong predictors benefit as much from the independent municipality-by-quarter effects as from the random walk effects. For weaker predictor models the random walk effects are more important. In the end a model with all three random effect terms has been selected, as it carries a smaller risk of underestimating standard errors, and it does not overfit the data in our case.

The estimates based on the time-series SAE model have been compared with estimates based on cross-sectional basic area-level and basic unit-level models as well as to survey regression estimates. As expected, the survey regression estimates cannot compete with the model-based estimates because of their much larger standard errors for small to medium-sized municipalities. The cross-sectional model-based estimates borrow strength only over areas but they are generally not very different from the time-series model estimates, presumably because a lot of strength is already borrowed over space and from the predictors. Still, the time-series model estimates are slightly more smooth over time and have smaller standard errors. In other applications, differences between time-series and cross-sectional estimates can be much larger, and sometimes more strength may be gained over time than over space, as evidenced for example in EURAREA (2004). For a few small municipalities our time-series estimates deviate from the cross-sectional estimates over an extended period where they are closer to the direct survey regression estimates. In these cases additional strength is borrowed over time. More benefits of the time-series model can be expected for areas for which data are entirely missing for one or a few periods, especially when random walk effects are included. This is similar to the benefit of explicit spatial modeling for estimates for out-of-sample areas (Gómez-Rubio et al., 2010).

The time-series model estimates have smaller standard errors than those based on the cross-sectional models. However, a regular production process will normally produce estimates for the current, latest period only. In that case the reduction of standard errors brought about by the time-series model is smaller, because strength over time is only borrowed from the past. The standard errors based on the time-series model are much more stable over time than the standard errors based on the cross-sectional models. The large variation of the latter seems to be caused by their sensitivity to the value of the between-area variance and the relatively large uncertainty about that value (Boonstra et al., 2011). Another advantage of the time-series model approach, one that we have not employed in this paper, is that standard errors of estimates of change can be easily computed, especially when the model fit is in the form of a MCMC simulation.

The time-series standard errors are very smooth, and a more thorough model evaluation is necessary to find out whether that is appropriate or whether the time-series model underfits the unemployment data or is open to improvement in other ways. There are many ways in which the time-series SAE model may be extended to further improve the estimates and standard errors. For example, it may be an improvement to use a logarithmic link function in the model formulation as in You (2008). Effects would then be multiplicative instead of additive. Another improvement would come from modeling the sampling variances (You and Chapman, 2006; You, 2008; Gómez-Rubio et al., 2010) instead of treating them as known. In order to decrease the risk of underfitting it may be important to replace some effects now treated as fixed (unmodeled) by random (modeled) effects varying over time and/or areas. For example, the model might improve by allowing the effects for registered unemployment to vary not only over time but also by area (this would also be a possible improvement for the cross-sectional models). In Datta et al. (1999) similar effects associated with unemployment insurance are modeled as varying over areas, although not over time. Another extension is to model the wave effects as random walk components over time, as in van den Brakel and Krieg (2009). We have seen that some results are quite sensitive to the way these effects are modeled. Modeling them as random effects would probably be an improvement over the models with fixed wave effects varying by year or by quarter that we have considered here.

Finally, before a time-series multi-level model as discussed in this paper can be employed for the production of official figures on municipal unemployment, a number of practical issues have to be considered. The first is that at the beginning of almost every year several administrative changes in the municipal structure are carried through. In the Netherlands these changes have solely been mergers over the recent past. A relatively easy way to deal with such changes is to keep the model specification at the original, most detailed municipal level. Initial estimates are then always computed at that level and do not need to be recomputed when the data for a new period are included. After the model is fit all estimates can be aggregated to the current municipal arrangement. Another issue is that survey redesigns can occur. The Dutch LFS has undergone redesigns in 2010 and 2012, which have led to changes in non-sampling error and consequent level changes in the time-series of direct unemployment estimates. To account for such changes, intervention variables may be added (van den Brakel and Krieg, 2012) as additional components in the measurement part of model (9). A last practical issue is that of benchmarking the detailed estimates obtained from the multi-level time-series model so as to agree with estimates at a higher level of aggregation. The estimates at a high level of aggregation are often estimated prior to the detailed figures, using different methods; in the case of the Dutch LFS, national unemployment figures are currently estimated using a structural time-series model. To obtain a consistent set of published estimates the detailed figures can be minimally adjusted so as to agree with these national estimates. If estimates for multiple related target variables are based on separate models, more consistency relations may have to be imposed. For the purpose of benchmarking and imposing more general consistency relations various methods are available, see for example Bell et al. (2013) and references therein.

Acknowledgements

I'd like to thank Jan van den Brakel, Bart Buelens and John Michiels for useful discussions and comments on earlier versions of this document.

References

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

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

Besag, J. and Kooperberg, C. (1995). On Conditional and Intrinsic Autoregression. *Biometrika* 82 (4), 733--746.

Boonstra, H., Buelens, B., Leufkens, K., and Smeets, M. (2011). *Small area estimates of labour status in Dutch municipalities*. Technical Report 201102, http://www.cbs.nl/nl-NL/menu/methoden/onderzoek-methoden/discussionpapers/archief/2011, Statistics Netherlands.

Boonstra, H., Buelens, B., and Smeets, M. (2011). *Borrowing strength over time for the estimation of unemployment at the municipal level*. Technical Report DMH-2011-12-5, Statistics Netherlands.

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

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

Boonstra, H. J. (2012). *hbsae: Hierarchical Bayesian Small Area Estimation*. R package version 1.0.

Datta, G. and Ghosh, M. (1991). Bayesian Prediction in Linear Models: Applications to Small Area Estimation. *The Annals of Statistics* 19 (4), 1748--1770.

Datta, G., Lahiri, P., Maiti, T., and Lu, K. (1999). Hierarchical Bayes Estimation of Unemployment Rates for the States of the U.S. *Journal of the American Statistical Association* 94 (448), 1074--1082.

EURAREA (2004). *Project reference volume, deliverable D7.1.4*. Technical report, EURAREA consortium.

Fay, R. and Herriot, R. (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 Smith, A. (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., Carlin, J., Stern, H., and Rubin, D. (2003). *Bayesian Data Analysis* (2 ed.). Chapman & Hall/CRC.

Gelman, A., Dyk, D. V., Huang, Z., and Boscardin, W. (2008). Using Redundant Parameterizations to Fit Hierarchical Models. *Journal of Computational and Graphical Statistics* 17 (1), 95--122.

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

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

Gómez-Rubio, V., Best, N., Richardson, S., Li, G., and Clarke, P. (2010). *Bayesian Statistics for Small Area Estimation*. Technical Report http://www.bias-project.org.uk/research.htm.

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

Plummer, M., Best, N., Cowles, K., and Vines, K. (2010). *coda: Output analysis and diagnostics for MCMC*. R package version 0.13-5.

R Development Core Team (2009). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Rao, J. (2003). *Small Area Estimation*. Wiley-Interscience.

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

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

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

Spiegelhalter, D., Best, N., Carlin, B., and van der Linde, A. (2002). Bayesian Measures of Model Complexity and Fit. *Journal of the Royal Statistical Society B* 64 (4), 583--639.

Vaida, F. and Blanchard, S. (2005). Conditional Akaike information for mixed-effects models. *Biometrika* 95 (2), 351--370.

van den Brakel, J. and Krieg, S. (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 Krieg, S. (2012). *Dealing with small sample sizes, rotation group bias and discontinuities in a rotating panel design*. Technical Report PPM-2012-11-13-JBRL, Statistics Netherlands.

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

You, Y. and Chapman, B. (2006). Small Area Estimation Using Area Level Models and Estimated Sampling Variances. *Survey Methodology* 32 (1), 97--103.

You, Y., Rao, J., and Gambino, J. (2003). Model-Based Unemployment Rate Estimation for the Canadian Labour Force Survey: A Hierarchical Bayes Approach. *Survey Methodology* 29 (1), 25--32.

Appendices

I Gibbs sampler for the time-series model

In order to write model (9) and (13) in vector form, we first introduce some notation. For area i ($i = 1, \dots, m_A$) let $\theta_i = \ominus_{t=1}^{m_T} \theta_{it} \equiv (\theta_{i1}, \dots, \theta_{im_T})'$ denote the vector of interest. Here m_T is the number of periods, and $\ominus$ denotes the operation of vertical stacking. Similarly, for the area-by-time random effects we write $u_i = \ominus_t u_{it}$ and $w_i = \ominus_t w_{it}$. Due to the rotating panel structure there are $m_P = 5$ observations for each θ_{it}, one for each wave. We write $\hat{\hat{Y}}_i = \ominus_{t=1}^{m_T} \ominus_{p=1}^{m_P} \hat{\hat{Y}}_{itp}$ for the observations (survey regression estimates) and $e_i = \ominus_t \ominus_p e_{itp}$ for the vector of survey errors associated with area i. The survey errors for different areas are independent, but due to the rotating panel structure they are correlated over time for the same area, as explained in Section 4.1. The full sampling error covariance matrix for $e = \ominus_i e_i$ can thus be written $\Phi = \oplus_i \Phi_i$ where Φ_i are the estimated $m_T \times m_T$ covariance matrices, which are assumed known in the model. We also define $X_i = \ominus_t \ominus_p (x'_{it}, z'_{itp})$, combining the predictors x and measurement-related covariates z into a single predictor matrix for area i. The corresponding effects are denoted $B = (\beta', \alpha')'$.

The model relating the data $\hat{\hat{Y}}_i$ to the unknowns can now be written as ($i = 1, \dots, m_A$)

$$\hat{\hat{Y}}_i = X_i B + R_{TP} v_i + R_{TP}^T (u_i + w_i) + e_i \,, \qquad e_i \overset{ind}{\sim} N(0, \Phi_i)\,, \tag{20}$$

where $R_{TP} = \iota_{m_T} \otimes \iota_{m_P} = \iota_{m_T m_P}$ and $R_{TP}^T = I_{m_T} \otimes \iota_{m_P}$ are replication matrices, which expand v_i, u_i and w_i to vectors of length $m_T m_P$. Here ι and I denote a vector of ones and an identity matrix of appropriate dimensions, respectively.

The model as written above includes all random effect terms considered, although we also fit models including only a subset of these. The independent random effect terms are modeled as $v_i \sim N(0, \sigma_v^2)$ and $w_i \sim N(0, \sigma_w^2 I_{m_T})$. All random terms are mutually independent. For the area-by-time random effects vector we have

$$u_i \overset{iid}{\sim} N(0, \sigma_u^2 A)\,.$$

If initial values u_{i1} are given, A is the $m_T \times m_T$ covariance matrix with elements $A_{st} = min(s, t)$. Instead, if we assign a non-informative, uniform, prior to u_{1i} for all areas, the matrix A becomes unbounded. However, the precision matrix A^{-1}, which is the matrix that appears in the Gibbs sampling algorithm below, has a well-defined value in this non-informative limit. It is the tri-diagonal matrix with diagonal $(1, 2, 2, \dots, 2, 1)$ and with all elements just above and below the diagonal equal to -1. This is a singular precision matrix of rank $m_T - 1$ as $A^{-1}\iota_{m_T} = 0$. The resulting random walk model is also known as an intrinsic autoregressive model (Besag and Kooperberg, 1995; Rue and Held, 2005). In order to guarantee identifiability of parameters, sum-to-zero constraints $\iota'_{m_T} u_i = 0$ are imposed, for $i = 1, \dots, m_A$. This ensures that only the fixed effects (notably the intercept) and the area effect v_i account for the time-average level of θ_{it}.

To complete the model specification, we assign a normal prior

$$p(B) = N_B(B_0, \Omega_B)$$

to the coefficients B, where we set $B_0 = 0$ and take Ω_B to be diagonal with relatively very large variances, expressing negligible prior knowledge. The choice of prior for the variance parameters σ_v^2, σ_u^2 and σ_w^2 is explained below in I.2.

If we write $\hat{\hat{Y}} = \ominus_i \hat{\hat{Y}}_i$, $\theta = \ominus_i \theta_i$, $v = \ominus_i v_i$, $u = \ominus_i u_i$, $w = \ominus_i w_i$, as well as $Z = \ominus_i \ominus_t \ominus_p z'_{itp}$ and $X = \ominus_i \ominus_t x'_{it}$, then the combined model over all areas becomes

$$\hat{\hat{Y}} = R^{AT}_{ATP}\theta + Z\alpha + e\,, \quad e \sim N(0, \Phi)\,,$$
$$\theta = X\beta + R^{A}_{AT}v + u + w\,, \qquad (21)$$

where $R^{AT}_{ATP} = I_{m_A} \otimes I_{m_T} \otimes \iota_{m_P}$ and $R^{A}_{AT} = I_{m_A} \otimes \iota_{M_T}$ are replication matrices.

I.1 Gibbs sampling

Inference about θ is based on its posterior distribution $p(\theta|\hat{\hat{Y}})$. 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 $\psi = (B, v, u, w, \sigma_v^2, \sigma_u^2, \sigma_w^2)$, the vector of model parameters. The joint posterior is determined by the model and prior specifications up to a normalization constant, as described in I.3. Using the 2nd line of (21), which expresses θ as a function of ψ, we obtain samples from the posterior distribution for θ.

The Gibbs sampler does not directly sample from the joint posterior. Instead, it iteratively samples from the full conditional distributions $p(\psi_g|\hat{\hat{Y}}, \psi_1, \dots, \psi_{g-1}, \psi_{g+1}, \dots, \psi_G)$, for a suitable decomposition of ψ in blocks ψ_g, $g = 1, \dots, G$. The full conditionals are generally easy to sample from; in particular, for model (21) all full conditionals are normal or inverse chi-squared densities. 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 $p(\psi_g|\hat{\hat{Y}}, \psi_1^{(k)}, \dots, \psi_{g-1}^{(k)}, \psi_{g+1}^{(k-1)}, \dots, \psi_G^{(k-1)})$

After convergence, samples can be considered draws from $p(\psi|\hat{\hat{Y}})$.

I.2 Redundant multiplicative reparameterization

It turns out that convergence of the Gibbs sampler can be very slow for hierarchical models with small variance components, as is the case in our application to unemployment data. To speed up convergence we use the redundant multiplicative reparameterization as described in Gelman (2006); Gelman et al. (2008). Thereto multiplicative parameters ξ_v, ξ_u and ξ_w are introduced, and scaled random effects $\tilde{v}_i$, $\tilde{u}_i$ and $\tilde{w}_i$ are defined by

$$v_i = \xi_v \tilde{v}_i\,, \qquad u_i = \xi_u \tilde{u}_i\,, \qquad w_i = \xi_w \tilde{w}_i\,,$$

with $\tilde{v}_i \sim N(0, \tilde{\sigma}_v^2)$, $\tilde{u}_i \sim N(0, \tilde{\sigma}_u^2 A)$ and $\tilde{w}_i \sim N(0, \tilde{\sigma}_w^2 I_{m_T})$ so that, given ξ_v, ξ_u and ξ_w,

$$\sigma_v = |\xi_v|\tilde{\sigma}_v\,, \qquad \sigma_u = |\xi_u|\tilde{\sigma}_u\,, \qquad \sigma_w = |\xi_w|\tilde{\sigma}_w$$

Default inverse chi-squared prior distributions are assigned to $\tilde{\sigma}_v^2$, $\tilde{\sigma}_u^2$ and $\tilde{\sigma}_w^2$:

$$\tilde{\sigma}_v^2 \sim Inv\!-\!\chi^2(\nu_v, s_v^2)\,, \quad and \quad \tilde{\sigma}_u^2 \sim Inv\!-\!\chi^2(\nu_u, s_u^2)\,.$$

The multiplicative parameters are assigned normal priors:

$$\xi_v \sim N(\alpha_v\,, \gamma_v)\,, \qquad \xi_u \sim N(\alpha_u\,, \gamma_u)\,, \qquad \xi_w \sim N(\alpha_w\,, \gamma_w)\,.$$

Together, the priors on the ξ and $\tilde{\sigma}$ parameters imply priors for σ_v, σ_u and σ_w from the folded t-family (Gelman, 2006). In particular, if we take $\alpha_v = 0$ and $\gamma_v = 1$

$$p(\sigma_v) \propto \left(1 + \frac{\sigma_v^2}{\nu_v s_v^2}\right)^{-(\nu_v+1)/2}\,,$$

and similarly for σ_u and σ_w. As argued in Gelman (2006), these priors are better non-informative priors for standard deviation parameters in hierarchical models than the ones implied by the conventional inverse chi-squared priors. The latter priors can be obtained by setting $\alpha_v = 1$ and $\gamma_v = 0$, which fixes the multiplicative parameter ξ_v to 1 and reproduces the original parameterization.

In terms of the redundant reparameterization (20) becomes

$$\hat{\hat{Y}}_i = X_i B + \xi_v R_{TP} \tilde{v}_i + R_{TP}^T(\xi_u \tilde{u}_i + \xi_w \tilde{w}_i) + e_i\,. \tag{22}$$

The individual parameters ξ_v and $\tilde{v}_i$, say, are not well identified, but this does not matter since we are only interested in their product, v_i, as a component of the vector of interest θ_i.

I.3 Derivation of the full conditionals

The expanded set of model parameters is written as $\psi = (B, \tilde{v}, \tilde{u}, \tilde{\sigma}_v^2, \tilde{\sigma}_u^2, \xi_v, \xi_u)$. This also defines the subdivision of ψ into blocks for which the full conditional posterior distributions are derived. The redundant multiplicative reparameterization does not really complicate the application of the Gibbs sampler, since all full conditionals are still either normal or inverse chi-squared distributions.

The full conditional distributions can be derived from the joint distribution of the model parameters given the data

$$p(\psi|\hat{\hat{Y}}) \propto p(\psi)p(\hat{\hat{Y}}|\psi)\,,$$

where the likelihood implied by the model specification (22) is

$$p(\hat{\hat{Y}}|\psi) = \prod_i N_{\hat{\hat{Y}}_i}(X_i B + \xi_v R_{TP} \tilde{v}_i + R_{TP}^T(\xi_u \tilde{u}_i + \xi_w \tilde{w}_i), \Phi_i)\,,$$

and the prior is decomposed as

$$\begin{aligned} p(\psi) = {} & N_B(B_0, \Omega_B)\left(\prod_i N_{\tilde{v}_i}(0, \tilde{\sigma}_v^2)\right)\left(\prod_i N_{\tilde{u}_i}(0, \tilde{\sigma}_u^2 A)\right)\left(\prod_i N_{\tilde{w}_i}(0, \tilde{\sigma}_w^2 I_{m_T})\right) \\ & \times Inv\!-\!\chi^2_{\tilde{\sigma}_v^2}(\nu_v, s_v^2)\; Inv\!-\!\chi^2_{\tilde{\sigma}_u^2}(\nu_u, s_u^2)\; Inv\!-\!\chi^2_{\tilde{\sigma}_w^2}(\nu_w, s_w^2) \\ & \times N_{\xi_v}(\alpha_v, \gamma_v)\; N_{\xi_u}(\alpha_u, \gamma_u)\; N_{\xi_w}(\alpha_w, \gamma_w)\,. \end{aligned}$$

The full conditionals can be derived using the fact that the conditional posterior density for a specific component ψ_k is proportional to the joint posterior, $p(\psi_k|\psi^{(-k)}, \hat{\hat{Y}}) \propto p(\psi|\hat{\hat{Y}})$, where $\psi^{(-k)}$ denotes the parameter vector with the kth component (block) deleted. So one can drop all factors of the joint posterior not involving ψ_k and then reexpress the remaining factors as a (unnormalized) known density in ψ_k. The following results concerning conjugate priors are useful for this purpose, see for example Gelman et al. (2003). If for a vector x there is prior information $p(x) = N_x(\mu_0, \Sigma_0)$ and data d observed on linear combinations Dx with normal measurement errors such that $p(d|x) = N_d(Dx, \Omega)$, then the posterior density for x is $p(x|d) = N_x(\mu, \Sigma)$ with

$$\mu = \left(D'\Omega^{-1}D + \Sigma_0^{-1}\right)^{-1} \left(D'\Omega^{-1}d + \Sigma_0^{-1}\mu_0\right) ,$$
$$\Sigma = \left(D'\Omega^{-1}D + \Sigma_0^{-1}\right)^{-1} .$$

Besides this result for combining normal densities we use a similar result for combining normal and inverse chi-squared densities as functions of the variance parameter: a prior $p(\sigma^2) = Inv{-}\chi^2_{\sigma^2}(\nu, s^2)$ and likelihood $p(x|\sigma^2) = N_x(\mu_0, \sigma^2\Sigma_0)$ combine into a posterior inverse chi-squared density

$$p(\sigma^2|x) = Inv{-}\chi^2_{\sigma^2}\left(\nu + q, \frac{1}{\nu + q}\left(\nu s^2 + (x-\mu_0)'\Sigma_0^{-1}(x-\mu_0)\right)\right) ,$$

where q is the dimension of x. Using these results we now derive the full conditionals for each component of the parameter vector ψ.

The full conditional for B is

$$p(B|\psi^{(-B)}, \hat{\hat{Y}}) \propto N_B(B_0, \Omega_B) \prod_i N_{\hat{\hat{Y}}_i}(X_iB + \xi_v R_{TP}\tilde{v}_i + R_{TP}^T(\xi_u\tilde{u}_i + \xi_w\tilde{w}_i), \Phi_i) ,$$
$$\Rightarrow p(B|\psi^{(-B)}, \hat{\hat{Y}}) = N_B(E_B, V_B) , \quad with$$
$$E_B = \left(X'\Phi^{-1}X + \Omega_B^{-1}\right)^{-1} \left(X'\Phi^{-1}\left(\hat{\hat{Y}} - \xi_v R_{ATP}^A\tilde{v} - R_{ATP}^{AT}(\xi_u\tilde{u} + \xi_w\tilde{w})\right) + \Omega_B^{-1}B_0\right) ,$$
$$V_B = \left(X'\Phi^{-1}X + \Omega_B^{-1}\right)^{-1} ,$$

where $R_{ATP}^A = R_{ATP}^{AT}R_{AT}^A = I_{m_A} \otimes \iota_{m_T m_P}$, see (21).

The full conditional for $\tilde{v}$ is

$$p(\tilde{v}|\psi^{(-\tilde{v})}, \hat{\hat{Y}}) \propto \prod_i \left(N_{\tilde{v}_i}(0, \tilde{\sigma}_v^2)\, N_{\hat{\hat{Y}}_i}(X_iB + \xi_v R_{TP}\tilde{v}_i + R_{TP}^T(\xi_u\tilde{u}_i + \xi_w\tilde{w}_i), \Phi_i)\right) ,$$
$$\Rightarrow p(\tilde{v}|\psi^{(-\tilde{v})}, \hat{\hat{Y}}) = \prod_i N_{\tilde{v}_i}(E_{\tilde{v}_i}, V_{\tilde{v}_i}) , \quad with$$
$$E_{\tilde{v}_i} = \left(\xi_v^2 R'_{TP}\Phi_i^{-1}R_{TP} + 1/\tilde{\sigma}_v^2\right)^{-1} \xi_v R'_{TP}\Phi_i^{-1}(\hat{\hat{Y}}_i - X_iB - R_{TP}^T(\xi_u\tilde{u}_i + \xi_w\tilde{w}_i)) ,$$
$$V_{\tilde{v}_i} = \left(\xi_v^2 R'_{TP}\Phi_i^{-1}R_{TP} + 1/\tilde{\sigma}_v^2\right)^{-1} ,$$

the components $\tilde{v}_i$ being conditionally independent.

The full conditional for $\tilde{u}$ is

$$p(\tilde{u}|\psi^{(-\tilde{u})}, \hat{\hat{Y}}) \propto \left(\prod_i N_{\tilde{u}_i}(0, \tilde{\sigma}_u^2 A)\, N_{\hat{\hat{Y}}_i}(X_iB + \xi_v R_{TP}\tilde{v}_i + R_{TP}^T(\xi_u\tilde{u}_i + \xi_w\tilde{w}_i), \Phi_i)\right) ,$$
$$\Rightarrow p(\tilde{u}|\psi^{(-\tilde{u})}, \hat{\hat{Y}}) = \prod_i N_{\tilde{u}_i}(E_{\tilde{u}_i}, V_{\tilde{u}_i}) , \quad with \qquad (23)$$
$$E_{\tilde{u}_i} = \left(\xi_u^2 R_{TP}^{T'}\Phi_i^{-1}R_{TP}^T + A^{-1}/\tilde{\sigma}_u^2\right)^{-1} \xi_u R_{TP}^{T'}\Phi_i^{-1}(\hat{\hat{Y}}_i - X_iB - \xi_v R_{TP}\tilde{v}_i - \xi_w R_{TP}^T\tilde{w}_i) ,$$
$$V_{\tilde{u}_i} = \left(\xi_u^2 R_{TP}^{T'}\Phi_i^{-1}R_{TP}^T + A^{-1}/\tilde{\sigma}_u^2\right)^{-1} ,$$

the components $\tilde{u}_i$ being conditionally independent. For reasons of identifiability, constraints $\iota'_{m_T}u_i = 0$ are imposed, or in terms of $\tilde{u}$, $\iota'_{m_T}\tilde{u}_i = 0$. These constraints are most easily imposed by drawing $\tilde{u}_i$ from the unconstrained full conditional (24) and subsequently adjusting to

$$\tilde{u}_i \to \left(I_{m_T} - \frac{V_{\tilde{u}_i} \iota_{m_T} \iota'_{m_T}}{\iota'_{m_T} V_{\tilde{u}_i} \iota_{m_T}} \right) \tilde{u}_i \,,$$

see Rue and Held (2005).

The full conditional for $\tilde{w}$ is

$$p(\tilde{w}|\psi^{(-\tilde{w})}, \hat{\hat{Y}}) \propto \left(\prod_i N_{\tilde{w}_i}(0, \tilde{\sigma}_w^2 I_{m_T})\, N_{\hat{\hat{Y}}_i}(X_i B + \xi_v R_{TP} \tilde{v}_i + R_{TP}^T(\xi_u \tilde{u}_i + \xi_w \tilde{w}_i), \Phi_i) \right) ,$$

$$\Rightarrow p(\tilde{w}|\psi^{(-\tilde{w})}, \hat{\hat{Y}}) = \prod_i N_{\tilde{w}_i}(E_{\tilde{w}_i}, V_{\tilde{w}_i}) \,, \quad with \tag{24}$$

$$E_{\tilde{w}_i} = \left(\xi_w^2 R_{TP}^{T'} \Phi_i^{-1} R_{TP}^T + I_{m_T}/\tilde{\sigma}_w^2 \right)^{-1} \xi_w R_{TP}^{T'} \Phi_i^{-1} (\hat{\hat{Y}}_i - X_i B - \xi_v R_{TP} \tilde{v}_i - \xi_u R_{TP}^T \tilde{u}_i) \,,$$

$$V_{\tilde{w}_i} = \left(\xi_w^2 R_{TP}^{T'} \Phi_i^{-1} R_{TP}^T + I_{m_T}/\tilde{\sigma}_w^2 \right)^{-1} \,,$$

the components $\tilde{w}_i$ being conditionally independent.

The full conditional for $\tilde{\sigma}_v^2$ is

$$p(\tilde{\sigma}_v^2|\psi^{(-\tilde{\sigma}_v^2)}, \hat{\hat{Y}}) \propto \left(\prod_i N_{\tilde{v}_i}(0, \tilde{\sigma}_v^2) \right) Inv{-}\chi^2_{\tilde{\sigma}_v^2}(\nu_v, s_v^2) \,,$$

$$\Rightarrow p(\tilde{\sigma}_v^2|\psi^{(-\tilde{\sigma}_v^2)}, \hat{\hat{Y}}) = Inv{-}\chi^2_{\tilde{\sigma}_v^2}\left(\nu_v + m_A, \frac{\nu_v s_v^2 + \sum_i \tilde{v}_i^2}{\nu_v + m_A} \right) .$$

The full conditional for $\tilde{\sigma}_u^2$ is

$$p(\tilde{\sigma}_u^2|\psi^{(-\tilde{\sigma}_u^2)}, \hat{\hat{Y}}) \propto \left(\prod_i N_{\tilde{u}_i}(0, \tilde{\sigma}_u^2 A) \right) Inv{-}\chi^2_{\tilde{\sigma}_u^2}(\nu_u, s_u^2) \,,$$

$$\Rightarrow p(\tilde{\sigma}_u^2|\psi^{(-\tilde{\sigma}_u^2)}, \hat{\hat{Y}}) = Inv{-}\chi^2_{\tilde{\sigma}_u^2}\left(\nu_u + m_A(m_T - 1), \frac{\nu_u s_u^2 + \sum_i \tilde{u}_i' A^{-1} \tilde{u}_i}{\nu_u + m_A(m_T - 1)} \right) .$$

As a consequence of the sum-to-zero constraints on $\tilde{u}$, the number of additional degrees of freedom is not $m_A m_T$ but $m_A(m_T - 1)$.

The full conditional for $\tilde{\sigma}_w^2$ is

$$p(\tilde{\sigma}_w^2|\psi^{(-\tilde{\sigma}_w^2)}, \hat{\hat{Y}}) \propto \left(\prod_i N_{\tilde{w}_i}(0, \tilde{\sigma}_w^2 I_{m_T}) \right) Inv{-}\chi^2_{\tilde{\sigma}_w^2}(\nu_w, s_w^2) \,,$$

$$\Rightarrow p(\tilde{\sigma}_w^2|\psi^{(-\tilde{\sigma}_w^2)}, \hat{\hat{Y}}) = Inv{-}\chi^2_{\tilde{\sigma}_w^2}\left(\nu_w + m_A m_T, \frac{\nu_w s_w^2 + \sum_i \tilde{w}_i' \tilde{w}_i}{\nu_w + m_A m_T} \right) .$$

The full conditional for ξ_v is

$$p(\xi_v|\psi^{(-\xi_v)}, \hat{\hat{Y}}) \propto N_{\xi_v}(\alpha_v, \gamma_v) \prod_i N_{\hat{\hat{Y}}_i}(X_i B + \xi_v R_{TP} \tilde{v}_i + R_{TP}^T(\xi_u \tilde{u}_i + \xi_w \tilde{w}_i), \Phi_i) \,,$$

$$\Rightarrow p(\xi_v|\psi^{(-\xi_v)}, \hat{\hat{Y}}) = N_{\xi_v}(E_{\xi_v}, V_{\xi_v}) \,, \quad with$$

$$E_{\xi_v} = \left(\tilde{v}' R_{ATP}^{A'} \Phi^{-1} R_{ATP}^A \tilde{v} + 1/\gamma_v \right)^{-1} \left(\tilde{v}' R_{ATP}^{A'} \Phi^{-1} (\hat{\hat{Y}} - XB - \xi_u \tilde{u} - \xi_w \tilde{w}) + \alpha_v/\gamma_v \right) ,$$

$$V_{\xi_v} = \left(\tilde{v}' R_{ATP}^{A'} \Phi^{-1} R_{ATP}^A \tilde{v} + 1/\gamma_v \right)^{-1} .$$

The full conditional for ξ_u is

$$p(\xi_u|\psi^{(-\xi_u)},\hat{\hat{Y}}) \propto N_{\xi_u}(\alpha_u,\gamma_u)\prod_i N_{\hat{\hat{Y}}_i}(X_iB+\xi_v R_{TP}\tilde{v}_i + R_{TP}^T(\xi_u\tilde{u}_i+\xi_w\tilde{w}_i),\Phi_i)\,,$$

$$\Rightarrow p(\xi_u|\psi^{(-\xi_u)},\hat{\hat{Y}}) = N_{\xi_u}(E_{\xi_u},V_{\xi_u})\,,\quad with$$

$$E_{\xi_u} = \left(\tilde{u}'R_{ATP}^{AT'}\Phi^{-1}R_{ATP}^{AT}\tilde{u}+1/\gamma_u\right)^{-1}$$
$$\times\left(\tilde{u}'R_{ATP}^{AT'}\Phi^{-1}(\hat{\hat{Y}}-XB-\xi_v R_{ATP}^{A}\tilde{v}-\xi_w R_{ATP}^{AT}\tilde{w})+\alpha_u/\gamma_u\right)\,,$$
$$V_{\xi_u} = \left(\tilde{u}'R_{ATP}^{AT'}\Phi^{-1}R_{ATP}^{AT}\tilde{u}+1/\gamma_u\right)^{-1}\,.$$

The full conditional for ξ_w is

$$p(\xi_w|\psi^{(-\xi_w)},\hat{\hat{Y}}) \propto N_{\xi_w}(\alpha_w,\gamma_w)\prod_i N_{\hat{\hat{Y}}_i}(X_iB+\xi_v R_{TP}\tilde{v}_i + R_{TP}^T(\xi_u\tilde{u}_i+\xi_w\tilde{w}_i),\Phi_i)\,,$$

$$\Rightarrow p(\xi_w|\psi^{(-\xi_w)},\hat{\hat{Y}}) = N_{\xi_w}(E_{\xi_w},V_{\xi_w})\,,\quad with$$

$$E_{\xi_w} = \left(\tilde{w}'R_{ATP}^{AT'}\Phi^{-1}R_{ATP}^{AT}\tilde{w}+1/\gamma_w\right)^{-1}$$
$$\times\left(\tilde{w}'R_{ATP}^{AT'}\Phi^{-1}(\hat{\hat{Y}}-XB-\xi_v R_{ATP}^{A}\tilde{v}-\xi_u R_{ATP}^{AT}\tilde{u})+\alpha_w/\gamma_w\right)\,,$$
$$V_{\xi_w} = \left(\tilde{w}'R_{ATP}^{AT'}\Phi^{-1}R_{ATP}^{AT}\tilde{w}+1/\gamma_w\right)^{-1}\,.$$

I.4 The Deviance Information Criterion

We use the deviance information criterion (DIC) as a model comparison measure (Spiegelhalter et al., 2002). It is easily computed from the MCMC output. The deviance is

$$D(\psi) = -2\log p(\hat{\hat{Y}}|\psi) = m\log 2\pi + \sum_i \log|\Phi_i| + \sum_i E_i'\Phi_i^{-1}E_i\,.$$

where $E_i = \hat{\hat{Y}}_i - X_iB - R_{TP}v_i - R_{TP}^T u_i$. The model fit is represented by the value of the deviance at the posterior mean of ψ, $D(E(\psi|\hat{\hat{Y}}))$, whereas model complexity is measured by $p_{\text{eff}} = E(D(\psi)|\hat{\hat{Y}}) - D(E(\psi|\hat{\hat{Y}}))$, i.e. the posterior mean of the deviance minus the deviance at the posterior mean. This quantity is also known as the effective number of model parameters. DIC is then defined, analogously to AIC, as

$$DIC = D(E(\psi|\hat{\hat{Y}})) + 2p_{\text{eff}}\,.$$

Models with lower DIC values are preferred.

I.5 Implementation

The above Gibbs sampler has been implemented in R (R Development Core Team, 2009), using the sparse matrix facilities of package `Matrix` (Bates and Maechler, 2010). For convergence diagnostics, package coda (Plummer et al., 2010) has been used.

II Results for time-series Model `7 vuw'

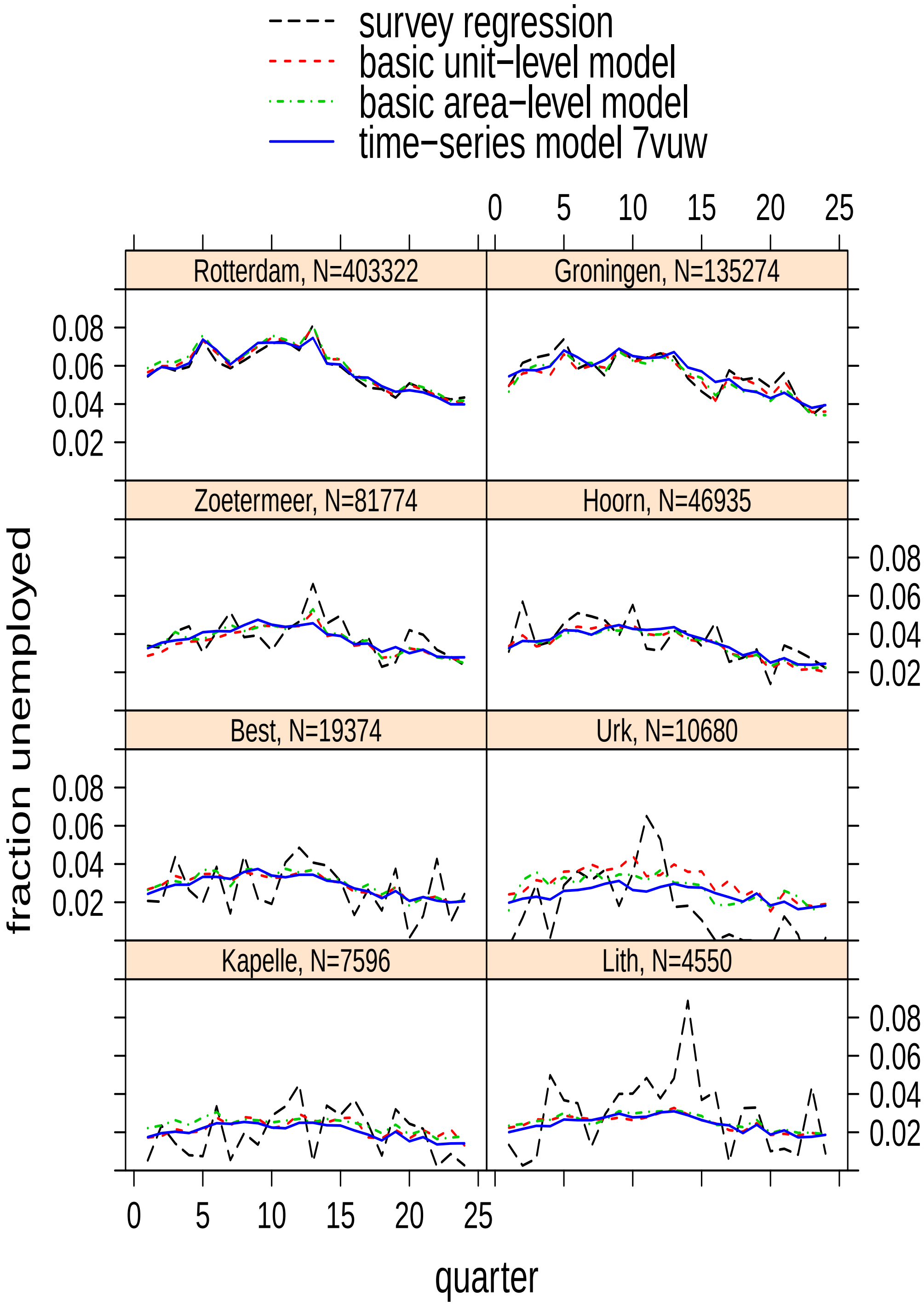

Figure 12 Comparison of estimates of unemployment fraction for 8 selected municipalities over time.

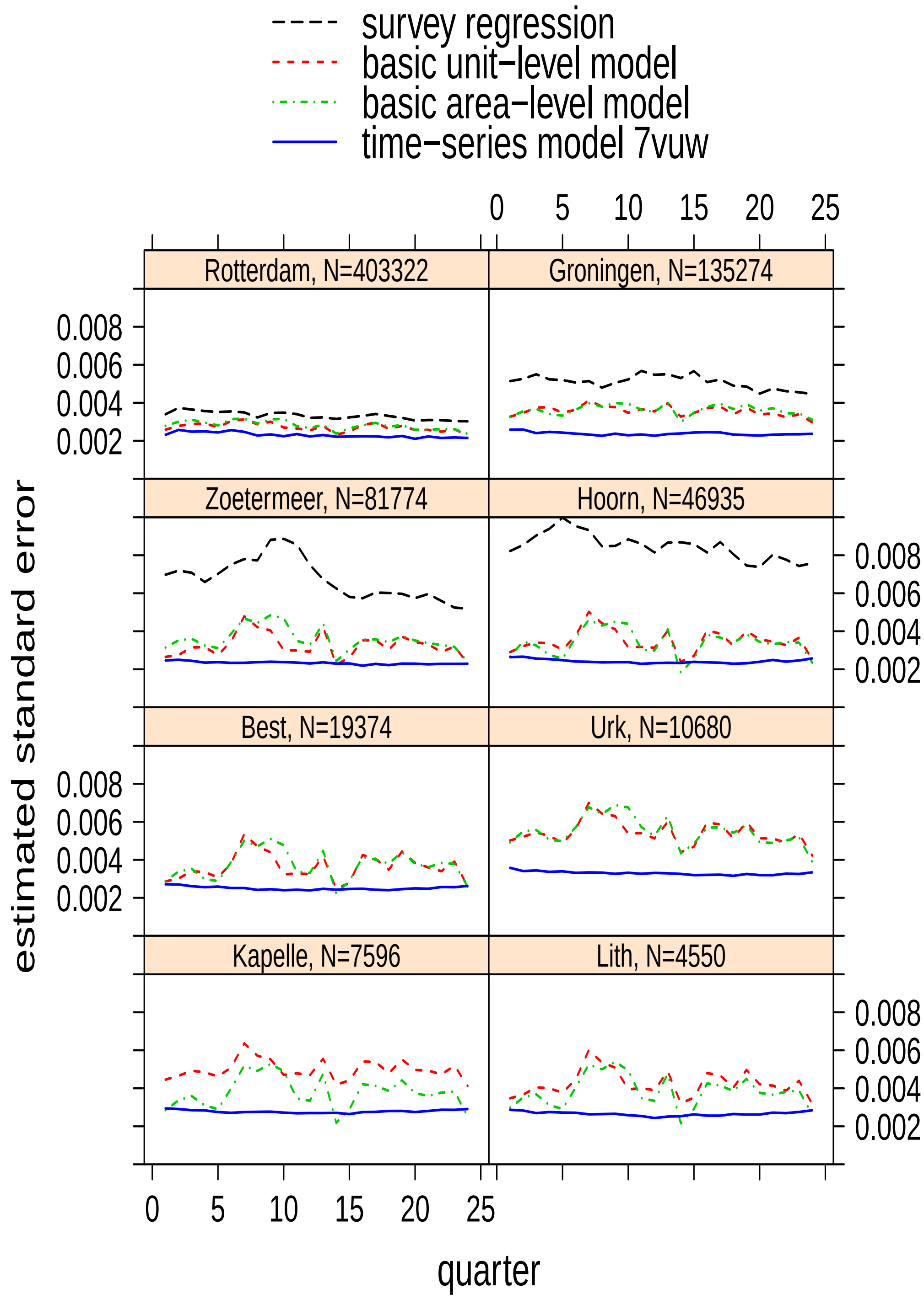

Figure 13 Comparison of standard errors for unemployed fraction for 8 selected municipalities over time.

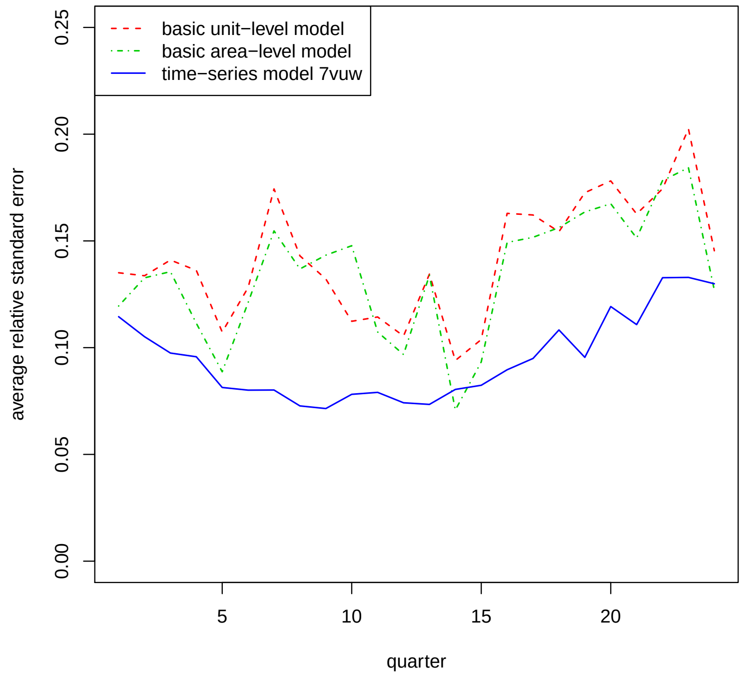

Figure 14 Comparison of relative standard errors averaged over all municipalities.

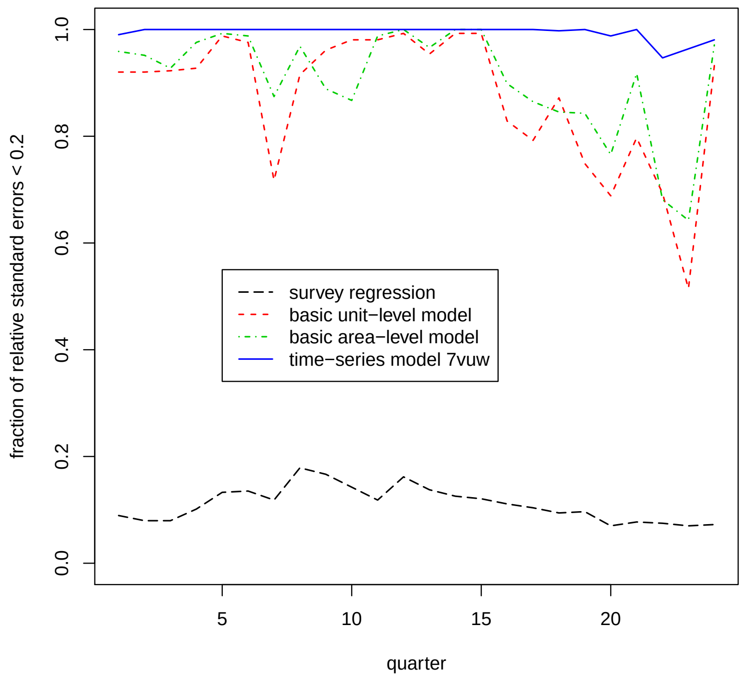

Figure 15 Fractions of estimates with relative error less than 20%**.**