Small area estimates of labour status in Dutch municipalities

Harm Jan Boonstra, Bart Buelens, Kasper Leufkens en Marc Smeets

Discussion paper (201102)

Statistics Netherlands

The Hague/Heerlen, 2011

Explanation of symbols

.	= data not available
*	= provisional figure
**	= revised provisional figure
x	= publication prohibited (confidential figure)
–	= nil or less than half of unit concerned
–	= (between two figures) inclusive
0 (0,0)	= less than half of unit concerned
blank	= not applicable
2010–2011	= 2010 to 2011 inclusive
2010/2011	= average of 2010 up to and including 2011
2010/’11	= crop year, financial year, school year etc. beginning in 2010 and ending in 2011
2008/’09–2010/’11	= crop year, financial year, etc. 2008/’09 to 2010/’11 inclusive

Due to rounding, some totals may not correspond with the sum of the separate figures.

Publisher
Statistics Netherlands
Henri Faasdreef 312
2492 JP The Hague

Prepress
Statistics Netherlands - Grafimedia

Cover
TelDesign, Rotterdam

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

Where to order
E-mail: verkoop@cbs.nl
Telefax +31 45 570 62 68

Internet
www.cbs.nl

ISSN: 1572-0314

60083201102 X-10

Small area estimates of labour status in Dutch municipalities

Harm Jan Boonstra, Bart Buelens, Kasper Leufkens and Marc Smeets

Within the ongoing redesign program of social surveys at Statistics Netherlands a small area estimation method for labour status has been developed. The model used is the basic unit-level model, which is a linear mixed model with random area effects, where the areas are municipalities. We discuss several issues concerning model choice, including the use of linear (mixed) models for binary variables, the use of posterior means instead of maximum likelihood estimates to prevent zero or too small estimates of between area variance and the use of covariates at both the unit and area level. Several model selection measures and graphical diagnostics have been applied to arrive at a set of covariates used in the model. We focus on the estimation of municipal unemployment fractions, but also discuss estimation of fractions employed and not belonging to the labour force. The municipal estimates are benchmarked such that they are consistent with regularly produced provincial estimates. The small area estimates thus obtained have smaller estimated mean squared errors than the current estimates based on the generalized regression estimator, and display a much more plausible development over time.

Key words: Small Area Estimation, Model selection, Labour Force Survey

1 Introduction

Within the ongoing redesign program of social surveys at Statistics Netherlands a small area estimation method for labour status has been developed and implemented. A model-based small area estimation method is applied to estimate labour status for the approximately 450 Dutch municipalities, at an annual frequency. These estimates are based on data from the Dutch Labour Force Survey (LFS).

Until now these estimates are produced by means of direct generalized regression (GREG) estimation for municipalities with at least 30 thousand inhabitants, using three-year averages for municipalities with 10 to 30 thousand inhabitants. No estimates are produced for municipalities with fewer than 10 thousand inhabitants. For some municipalities direct estimation would not even be possible since they are not observed under the LFS.

The Dutch LFS is described in Section 2. For model-based estimation of municipal labour status we use the basic unit-level model, a normal linear model with random municipality effects, formulated at the unit level. This model is described in Section 3. Section 4 discusses some issues with the LFS data concerning missing data, measurement error and repeated observations. An important issue for statistical offices is consistency of aggregates of small area estimates with regular (direct) estimates at the higher level. Provincial unemployment figures, for example, are based on a direct generalized regression estimator. To make the small area estimates consistent at this level, we use a linear adjustment method, that takes into account the estimated mean squared errors of the small area estimates. This is discussed in Section 5. Model evaluation and diagnostics are the topics of Section 6. Results based on LFS data from 2001 to 2009 are given in Section 7. Section 8 concludes this paper.

2 The Dutch Labour Force Survey

The objective of the LFS is to provide reliable information about the labour market. The population aged 15 through 64 is divided into three groups: the employed labour force, the unemployed labour force and those that do not belong to the labour force. The population fractions belonging to these groups are important parameters of the LFS. Another important parameter is the unemployment rate, which is defined as the ratio of the unemployed labour force and the labour force.

The Dutch LFS is conducted according to a rotating panel design, in which the respondents are interviewed five times at quarterly intervals. Each month a sample of addresses is selected through a stratified two-stage cluster design. Strata are formed by geographic regions. Municipalities are considered as primary sampling units and addresses as secondary sampling units. All households residing at an address, up to a maximum of three, are included in the sample. Addresses with only persons aged 65 years and over are undersampled, since most target parameters of the LFS concern people aged 15 through 64 years. Gross sample sizes used to be about 8,000 addresses monthly, but have declined to about 6,500 addresses in recent years.

In the first wave of the panel, data are collected by means of computer assisted personal interviewing (CAPI). In the four subsequent waves of the panel, data are collected by means of computer assisted telephone interviewing (CATI). During these re-interviews a reduced questionnaire is applied to establish changes in the labour market position of the household members aged 15 years and over. When a household member cannot be contacted, proxy interviewing is allowed by members of the same household in each wave.

Several registrations provide a wealth of auxiliary information from which predictors can be selected to be used for weighting and estimation. Most predictors are available at the unit level. Among these auxiliary variables is registered unemployment[1], a strong predictor for the unemployment variable of interest. Figure 1 shows the development over time of registered unemployment (RU) and LFS unemployment (LFSU) as fractions of the target population of persons aged 15 through 64. The figure shows unweighted sample means as well as the known RU population mean and the estimated LFSU population mean. Also shown are unweighted sample fractions of persons both registered and LFS unemployed. It is clear from the figure that the time-developments of RU and LFSU are quite similar. From the two lowest lying lines it can be seen that a significant fraction of LFS unemployed persons are registered as unemployed as well, so that the variables are relatively strongly correlated. Together with the underrepresentation of RU in the LFS sample for all years except 2001, this shows the importance of RU as auxiliary information. The anomalous pattern for 2001 is due to the fact that registered unemployed were oversampled until July 2001.

The weighting procedure of the LFS, used to produce quarterly figures about the labour market, is based on the GREG estimator (Särndal et al., 1992). The weighting scheme employed is based on a combination of different socio-demographic categorical variables, including gender, age, household type, and registered unemployment.

3 Model-based small area estimation

For the estimation of municipal employment and unemployment we use the basic unit level model, alternatively called nested error regression model or Battese-Harter-Fuller model (Battese et al., 1988). The model and derived expressions for small area estimates and mean squared errors are given in Appendix A. Additional information can be found in Datta and Ghosh (1991) and Rao (2003). The basic unit level model is a linear model that assumes normally distributed errors. However, LFS (un)employment data consist of binary variables at the unit level. Usually, such data are modeled using (non-normal) generalized linear (mixed) models, for example a binomial model with a logistic link function. There

[1]We use the term registered unemployment for the variable indicating whether a person is registered at the employment agency or not. This deviates from the definition used in published figures on registered unemployment, which also involves information observed in the LFS.

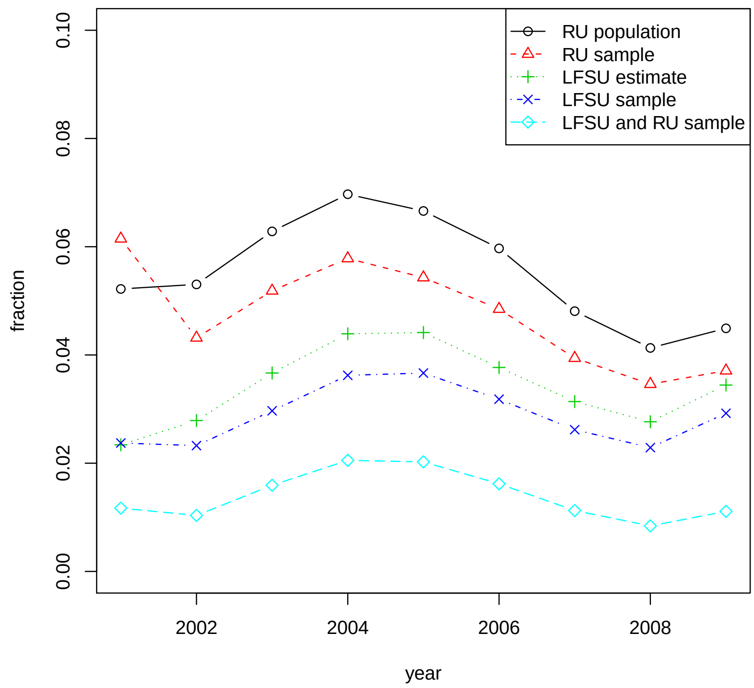

Figure 1. Fractions unemployed at the national level between 2001 and 2009. The sample means are based on first and fifth wave LFS data.

are, however, several reasons why we choose to use normal linear models:

- In a preliminary investigation including a simulation study it was found that logistic models with or without area effects do not improve upon the estimates based on the simpler normal linear models (Boonstra et al., 2007).

- The quantities of interest are area means, so we are not interested in predictions for individual units, but always aggregate such predictions to the area level. This makes it more reasonable to use normal linear models.

- The normal linear models result in analytical expressions (called Best Linear Unbiased Predictors or BLUPs) that can be interpreted as weighted averages of direct and synthetic estimates. For non-linear models such analytical expressions are not available, which makes them harder to interpret.

- Normal linear models are convenient from a computational point of view, because estimates and variances can be expressed analytically as functions of a single variance parameter. In a Bayesian approach only one-dimensional numerical integration is required to average over the posterior density of this variance parameter, which is a relatively easy task. For more complex models, e.g. logistic hierarchical models with multiple random effects, one would need Monte Carlo methods to approximate the multi-dimensional integrals. For the simple linear models, the advantage of much shorter computation time is important because it means that many different models can be tried within a reasonable time. We implemented the basic unit-level model in R, and optimized the numerical integrations by storing function evaluations for re-use in subsequent numerical integrations. In our application, small area estimates for all approximately 450 municipalities are obtained in seconds, whereas a Monte Carlo approach might take hours.

- Linear models require only area population *totals* for prediction. For example, using the basic unit level model with fixed effects $Ageclass + MaritalStatus + Ethnicity$ requires area population totals for each of these variables separately, whereas a non-linear model, such as logistic regression, would require area totals for the complete cross-classification $Ageclass \times MaritalStatus \times Ethnicity$. The reason is that for such models the predictions are non-linear functions of the auxiliary variables and therefore the sum of such predictions over unobserved units cannot be expressed in terms of sums of the auxiliary variables, unlike in the linear case. If continuously varying covariates are used, non-linear models need complete auxiliary information, i.e. the individual vectors of covariates for all units. This advantage of linear models can be important in official statistics, where often covariates from several registrations are used. Assembling a data set of complete auxiliary information would require linking of all registrations.

- The basic unit level model is a multi-level generalization of linear regression models underlying GREG estimators. The GREG, resulting from linear survey weighting, is widely applied at statistical bureaus to estimate population totals of any kind of variable, including binary variables.

The registered unemployment variable is a strong predictor for the unemployment variable of interest. Using this variable as a covariate improves the model fit to the point that problems associated with small or zero estimated variance parameters arise, see Buelens et al. (2009). Zero estimated between-area variance would mean that the estimates reduce to synthetic estimates, which is certainly undesirable for the large municipalities, which have quite large sample sizes. Moreover, the model-based error estimates become unrealistically small by ignoring between-area variance. The extent of these problems is reduced by adopting a Bayesian approach. This is illustrated in Figure 2. Although there are no zero estimates of the variance parameter in this case, the REML estimates are always smaller than the posterior means, and they also seem to fluctuate more. So EBLUP estimates based on the REML estimates are closer to synthetic estimates. Whereas the REML estimate might become zero in some future year, the posterior mean of the between-area variance will always be positive. A zero REML estimate actually occurs for the fit to 2005 unemployment data when using the extended set of covariates discussed in §7.3.

The additional computational burden for a Bayesian approach is small because we only need one-dimensional numerical integration to compute posterior means and variances, as explained in Appendix A. The approach adopted is also called the hierarchical Bayes (HB) approach (Rao, 2003, Chapter 9), after the hierarchical or multi-level structure of the model used.

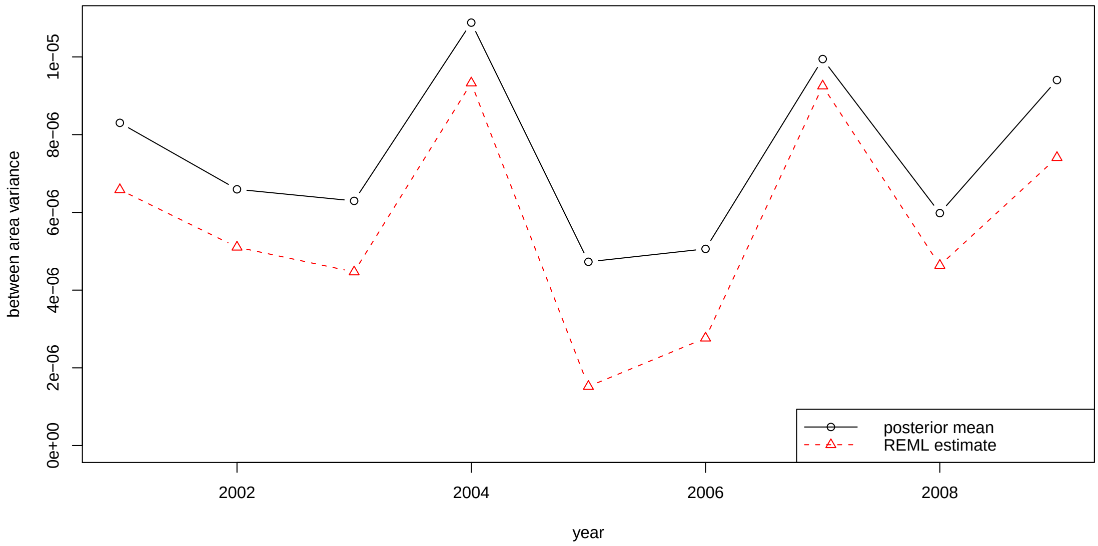

Figure 2. Posterior means and REML estimates for the between-area variance of the basic unit level model fitted to unemployment data of the years 2001 through 2009. The same predictors are used in each year: registered unemployment in combination with a number of demographic covariates.

4 Accounting for the LFS design, non-response and measurement error

In contrast to traditional design-based estimation, small area estimation based on the basic unit level model is a purely model-based approach and therefore does not start from inverse inclusion probability weighting. Instead, it relies on including covariates related to the variable of interest, especially if they are also related to the missing data mechanism. This is similar to the way a weighting model is constructed, except that more attention is paid to the variable being modeled. If inclusion probabilities differ by design, the covariates explaining these differences should be taken into account in the model; the sampling design is then said to be ignorable (Rubin, 1976). A minimal variant would be to use the inclusion probabilities themselves as a covariate in the mean and perhaps variance specifications of the model. Often, differences between inclusion probabilities are appropriately accounted for by the presence of stratum indicators in the model. Boonstra (2005) reviews how design-based GREG estimates for population totals are reproduced as model-based prediction estimates based on models that incorporate the inclusion probabilities in the mean and variance specification.

Addresses with persons aged between 14 and 26 as well as addresses with non-western immigrants are oversampled in the LFS, which means that age and ethnicity must be included as covariates in the model. Another feature of the LFS design is clustering. Municipal clustering is not very relevant for annual estimates, since almost all municipalities are included in the data of a single quarter. However, if household clustering is not taken into account, this may affect error estimates. In model-based estimation, clustering is normally taken into account by incorporating cluster random effects. However, this would complicate the unit level model used. Fortunately, for labour status (employed, unemployed, not in the labour force), household cluster effects turn out quite small: design-based variance estimates that do not account for household clustering are only a few percent lower than estimates that do account for it. Ignoring household clustering should therefore not be problematic in this application.

Measurement errors are another cause of concern in the LFS, especially for the estimation of unemployment. Panel and mode effects occur due to the repeated observation of households and the switch of observation mode after the first interview from CAPI to CATI. The total effect between the first and subsequent waves, called rotation group bias, is quite large (see van den Brakel and Krieg, 2009), but it is not clear how much of it can be attributed to non-response and panel attrition and how much to measurement errors. First wave observations are believed to be the most accurate, and therefore final estimates are currently adjusted using ratios of first-wave-only to all-wave estimates averaged over several years. In the time series models for monthly unemployment as described in van den Brakel and Krieg (2009) the effects relative to the first wave are modeled explicitly as measurement error terms and are thereby eliminated. For municipal small area estimation, we do not as yet borrow strength over time, but instead we borrow strength over space, i.e. over the municipalities, by means of the basic unit level model. Rotation group biases relative to the first wave can be eliminated by including dummies for follow-up waves as additional regressors, as explained below. But first we must address the issue of repeated measurements for the same persons.

4.1 Dealing with repeated measurements for annual estimates

To make annual estimates using all LFS data observed in a year, one has to deal with multiple observations for the same person. This is a consequence of the rotating panel design.

Figure 3 shows the annual data subdivided into the different waves (let t correspond to the fourth quarter). First and subsequent waves are coloured differently to emphasize the different observation modes (CAPI vs. CATI). Blocks that have one or more blocks above them comprise repeated measurements of persons that were already in the panel.

We briefly discuss three different ways of using these data to make annual small area estimates based on the area or unit level models.

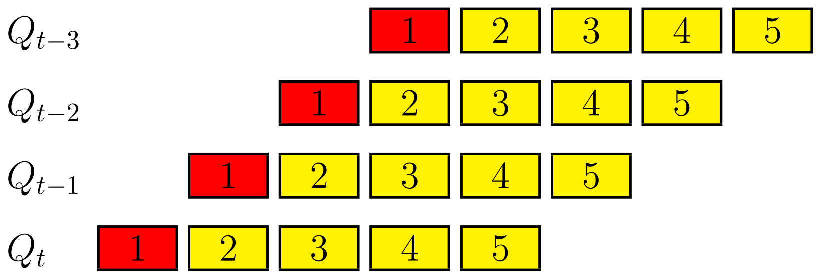

Figure 3. LFS rotating panel design

1. *Using first wave (CAPI) data only.* Besides that this implies that there are no repeated measurements, this has the advantage of avoiding the rotation group bias. As a consequence, calibration of the small area estimates to provincial GREG-based estimates requires smaller adjustments, since the latter are adjusted for rotation group bias. However, this approach uses less than half of all available data. This is hardly acceptable, especially for a small area estimation application.

2. *Using one observation for all unique cases.* This approach uses more data, but not all. Only a single observation for each unique person is retained; repeated observations are discarded. Referring to Figure 3, we see that this is achieved by choosing one block in each of the 8 columns. By using all first and fifth waves, the data is approximately balanced over the quarters of the year. The relatively small number of respondents in waves 2 to 4 of the first quarter that drop out of the panel before their fifth interview, may also be included.

3. *Using all information.* This approach uses all the data, including multiple measurements for most persons. The basic unit level model is not directly suitable for this kind of longitudinal data, and extending the model to include a longitudinal correlation structure in addition to the within area correlations would be complicating. One way to circumvent this problem is to use average measurements for each person. If a person is measured twice as being unemployed and once as being employed, the value of the averaged unemployment variable is 2/3, etc. The averaged variable is then directly modeled. One issue with this approach is how to account for the number of observations underlying each average. An effective number of observations based on the estimated autocorrelation (for quarterly intervals) of the unemployment variable could be used in the variance structure, thus giving more weight to averages based on more measurements. This approach must deal also with the rotation group biases between all different waves. Another disadvantage is that this approach can be laborious since for time-dependent covariates, including registered unemployment, one needs averaged versions as well.

An advantage of the second and third methods is that the sample overlap between consecutive years leads to better estimates of change. However, the second and third methods are affected by the rotation group bias. One way to deal with the rotation group bias is to ignore the differences until the end when the estimates are calibrated to direct estimates, which themselves have been calibrated to the first-wave level. Another, more refined way, is to adapt the model to allow for measurement errors. For the unit level model, a simple way to do this is to include wave number as an additional categorical covariate. Setting associated area population means to zero for all but the first wave, effectively removes the bias of the other waves relative to the first wave. Such a model assumes constant measurement errors for all but the first wave, which is not realistic, but presumably better than ignoring measurement error altogether. The effect is that, globally, the estimated level is shifted to approximately the first-wave level, while the individual area estimates still benefit from the additional data in the subsequent waves as expressed by smaller error estimates. It would also be possible to allow for differences in rotation group bias between different subpopulations by including interactions, but we have not found substantial interactions of this kind.

Based on the above discussion we have chosen to base the small area estimates on first and fifth wave data, i.e. approach 2, as a compromise between practical considerations and the wish to use all available information. The rotation group bias is modeled as the regression coefficient of a dummy variable for the fifth wave. Small area estimates are computed by prediction using the population means of regressors for each area. Setting the population means corresponding to the fifth wave dummy variable to zero effectively removes the relative bias of the fifth wave data to the first wave data. This is explained further in Appendix A.2.

5 Calibration to direct estimates at the level of provinces

Provincial annual estimates are currently produced using GREG estimation, since annual sample sizes at the provincial level are considered large enough for this purpose. The small area estimates at the municipal level can be aggregated to the provincial level, yielding another set of provincial estimates. However, it is undesirable to have inconsistencies among the sets of estimates. To remove these inconsistencies the small area estimates can be minimally adjusted such that they aggregate to the direct provincial estimates. This calibration or benchmarking procedure may also provide some protection against possible model defects.

For notational simplicity, we use the notation a for the vector of (uncalibrated) small area estimates, b for the sought vector of calibrated small area estimates, and $V(a)$ for the posterior covariance matrix of the small area estimands. The consistency restrictions are denoted $Rb = r$, with R a $P \times M$ aggregation matrix, and r is a P-vector of direct estimates, in our case a vector of 12 provincial estimates. If a and r estimate population means, the (p, i) element of R equals N_i/N_p^{prov} if municipality i lies in province p, and 0 otherwise, N_i and N_p^{prov} being known municipal and provincial population sizes. A simple linear adjustment is obtained by minimizing the quadratic form

$$(b-a)'V(a)^{-1}(b-a) \tag{1}$$

as a function of b subject to $Rb = r$. The result, which can be derived using Lagrange multipliers, is

$$b = a + C(r - Ra)\,, \qquad C = V(a)R'(RV(a)R')^{-1}\,. \tag{2}$$

Pfeffermann and Barnard (1991) proposed this adjustment in the special case of a single restriction with regard to the overall population mean.

If r is known exactly, (2) is optimal in the sense that it has minimum error variance among all linear estimates of the form $a + C(r - Ra)$ for any matrix C compatible with the constraints, i.e. satisfying $RC = I$. However, in a typical small area application, r is a vector of (direct) estimates, coming with its own uncertainty. In that case, there is no solution that simultaneously minimizes the error variances of all components of b simultaneously, see Pfeffermann and Barnard (1991), Knottnerus (2002) Chapter 12, Boonstra (2004) and Wang et al. (2008). Such an optimal result would only be possible by relaxing the constraints $Rb = r$ thereby implicitly adjusting the direct provincial estimates as well.

6 Model selection and evaluation

In order to evaluate the models, and in particular the set of predictors used in the models, we looked at several model selection measures and diagnostics. In previous work, we compared the unit level model with the Fay-Herriot model, a model formulated at the area level. The area level model was seen to be more limited in its use of auxiliary information. Also, a comparison based on a cross-validation measure favoured the unit level model.

6.1 Model selection measures

Along with the small area estimates we compute a number of model selection measures: Bayesian Information Criterion (BIC), Akaike Information Criterion (AIC), conditional AIC (cAIC) and a Cross Validation (CV) measure. These measures weigh goodness-of-fit against model complexity, and are defined as follows:

1. $BIC = -2llh + log(n)d$
2. $AIC = -2llh + 2d$
3. $cAIC = -2llh_c + 2d_c$
4. $CV = L(y_s, \hat{y}_s^-)$

where llh is the log-likelihood, llh_c the conditional log-likelihood, d and d_c the number of model parameters (see below for details), n the sample size, y_s the vector of observed sample values, $\hat{y}_s^-$ the vector of leave-one-out predicted values $(\hat{y}_1^{-(1)}, ..., \hat{y}_n^{-(n)})$, and L a loss function.

The likelihood used in llh is $p(y_s|\beta) = \int p(y_s|\beta, v)\mathrm{d}v$, where the integration is over the vector v of random effects. The conditioning on variance parameters is omitted for notational convenience. When used in BIC and AIC, the likelihood is evaluated at the Maximum Likelihood (ML) estimates of the coefficients of the fixed effects β and of the variance parameters. In the cAIC, the conditional likelihood is used instead, $p(y_s|\hat{\beta}, \hat{v})$, evaluated at the estimated fixed effects $\hat{\beta}$ and random effects $\hat{v}$.

In the AIC and BIC d is the number of parameters in the model. In the cAIC, d_c is the number of model parameters taking into account the random effects as well. The value of d_c is taken to be the trace of the hat matrix H, the matrix that takes the data to the fitted values, $\hat{y}_s = Hy_s$ (Spiegelhalter et al., 2002; Ch. 7 of Hastie et al., 2003). Vaida and Blanchard (2005) argue that cAIC is more appropriate than AIC for use in small area settings. In particular, advantages include the ability to compare models with and without random effects, and to compare models fitted using ML, REML or any other method.

In the expression for CV, the loss function L is commonly taken to be quadratic, in which case $CV = \frac{1}{n}\sum_i (y_i - \hat{y}_i^{-(i)})^2$ can be viewed as an empirical prediction MSE. The leave-one-out predicted value $\hat{y}_i^{-(i)}$ is the prediction for the ith unit from the model fitted using the sample excluding the ith observation.

For a more detailed description of BIC, AIC, cAIC and CV, we refer to Boonstra et al. (2009) and references therein.

These model selection measures do not take into account differences between the response and the population. Auxiliary variables that contribute relatively little to the model fit may still be important to include in the model if the response and population means of these variables are very different.

6.2 Model diagnostics

Since the models we use do not borrow strength over time, assessment of the plausibility of time series of both small area estimates and of estimated model coefficients is a useful model diagnostic. Such an assessment is typically conducted graphically, by studying time series plots. Temporal instabilities in these plots may flag issues with the chosen approach.

Diagnostics based on analyses of residuals are often used in regression problems. Since we are interested in estimation at the area level we are primarily concerned with residuals at the area level, that is the random area effects. One such diagnostic is the graphical representation of the area random effects versus (the logarithm of) the area population size. There should be no obvious patterns visible in such a plot.

Finally, the deviation of the model-based estimates from design-based estimates at high aggregation levels can be seen as a diagnostic. If this deviation is large, and calibration (see Section 5) results in considerable adjustments to the estimates, the model is likely to be misspecified.

7 Results

7.1 The selected model

A detailed analysis of model selection measures associated with a range of potential models has resulted in the following model for unemployment:

`unemployed` $\sim$ `gender` $\times$ `age3` $\times$ `ru1 + ru1` $\times$ `ethnicity3 + ethnicity2D + ru + age + typehh + prov + ru-area + wave`.

The variables used in this model are described in detail in Appendix B.

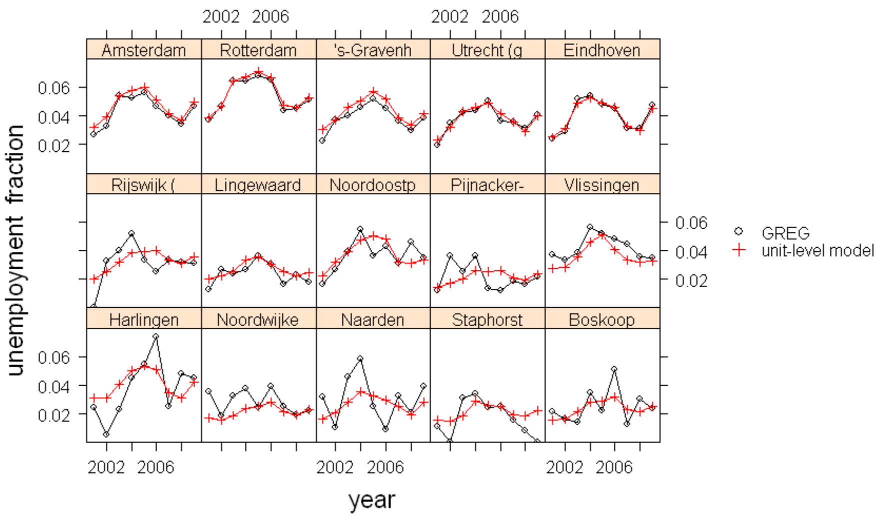

Figure 4. Direct (survey regression) and unit level model estimates of unemployment fractions for the years 2001 to 2009. The top row contains the five largest municipalities, the middle row the largest municipalities with less than 30,000 persons, and the bottom row the largest municipalities with less than 10,000 persons in the target population.

We present some model diagnostics for this particular model. Figure 4 shows an example displaying unemployment estimates for selected municipalities for nine consecutive years, from 2001 to 2009. For the largest municipalities (top row) the model-based estimates follow the direct GREG estimates closely. For smaller municipalities (middle and especially bottom row) the model-based estimates are clearly much more plausible than the direct estimates.

Figure 5 shows a plot of some estimated model coefficients (fixed effects) over time. Error bounds of plus or minus two standard errors are also shown. Some of the coefficient

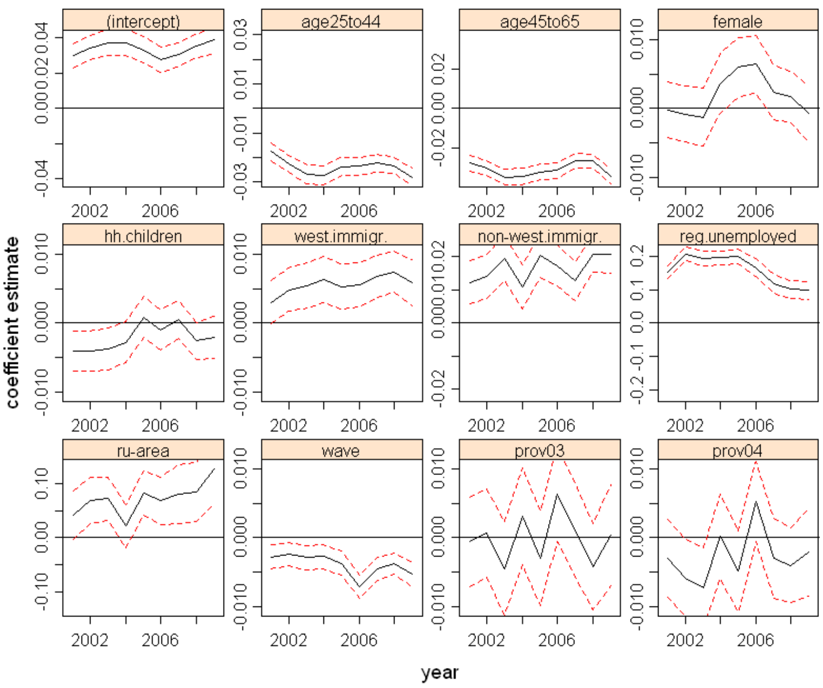

Figure 5. Selection of estimated coefficients in the unit level model fit to unemployment data for 2001-2009, with error bounds of plus or minus two standard errors. Note the different scales on the y-axis.

estimates vary smoothly over time. Among them, the coefficient of registered unemployment has a large value and very narrow error bounds. Other coefficients vary much more over time, especially the province indicators. They are mostly not significantly different from 0, but have been included because the estimates are calibrated to provincial direct estimates, see Section 5.

A clear pattern that emerged is that among registered unemployed, natives are more often LFS unemployed than non-western immigrants, whereas among persons not registered as unemployed, non-western immigrants are more often LFS unemployed. This is the reason why the interaction term $ru1 \times ethnicity3$ is included in the model. Indeed, Figure 5 shows that the coefficient of non-western immigrants is significantly positive for all years, which was not the case without the interaction term.

Figure 6 shows a plot of estimated random municipality effects versus the logarithm of area population size for 2009. Plots for other years look similar. The spread of random effects seems to grow slowly with the municipal population size. It turns out that this pattern largely disappears when the variances of the random effects are taken to be inversely proportional to the square root of the population sizes. Fortunately though, this change in variance specification hardly affects the small area estimates, which means that we can still use the simpler model with constant variances. Note that there are a few exactly zero random effects, corresponding to areas without observations.

Next, the effect of calibration is considered. We have used (2) to obtain small area

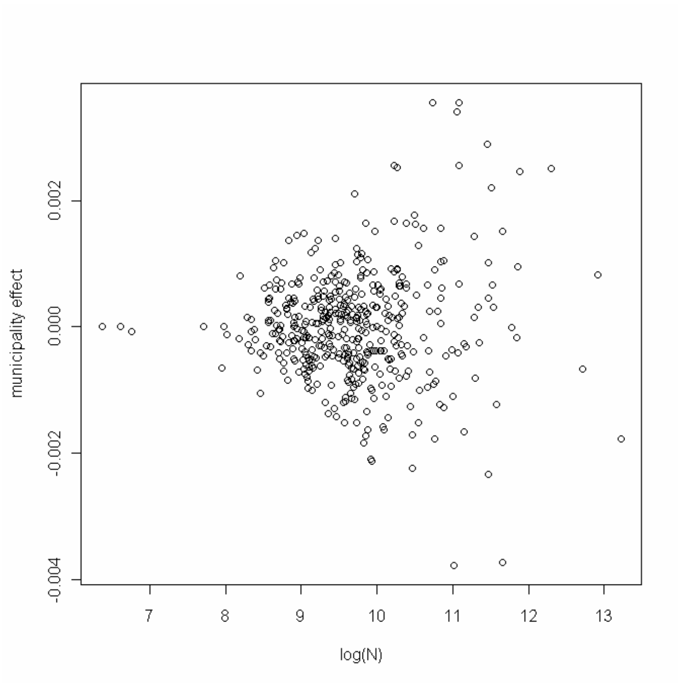

Figure 6. Estimated municipality effects versus the logarithm of the size of the target population in each municipality.

estimates consistent with direct provincial estimates. Although it is possible to adjust also the MSEs of the small area estimates, we did not do so. In any case, it would not be justified to adjust the MSEs acting as if r, the vector of direct provincial estimates, were known without error, when in fact it is a vector of uncertain estimates. In our application we expect that $V(a)$ is still an appropriate measure of the uncertainty in the adjusted estimates. In using (2), we initially ignored off-diagonal elements of $V(a)$ as an approximation. However, this led to too much of the required adjustments being allocated to the largest municipalities. Therefore we use the full covariance matrix $V(a)$ in the adjustment procedure.

Ideally, calibration adjustments are small "cosmetic" adjustments. This is normally the case when most auxiliary information used in the regular LFS weighting is also incorporated in the model used to make small area estimates. In order to reduce the size of the calibration adjustments, we also include province indicators as covariates in the model.

Figure 7 shows the difference between including and not including province indicators as covariates. The estimates and standard errors are obtained by aggregation of the municipal estimates and covariances. As expected, estimates based on the model including province are on average closer to the direct provincial estimates. More notable is that the provincial standard errors are, on average, almost doubled as a result of including province. This is not the case for the underlying municipal estimates; the estimated standard errors at that level are not much affected by the province indicators. The larger standard errors at the provincial level are due to more positive correlations between the municipal estimates within the same province as a result of the uncertainty in the province coefficients. Another way to think of this is that by including province indicators, the provincial standard errors get closer to the standard errors of the direct provincial estimates. This confirms the importance of including indicators for subpopulations of interest, not only to improve the subpopulation estimates, but also to prevent over-optimistic standard errors.

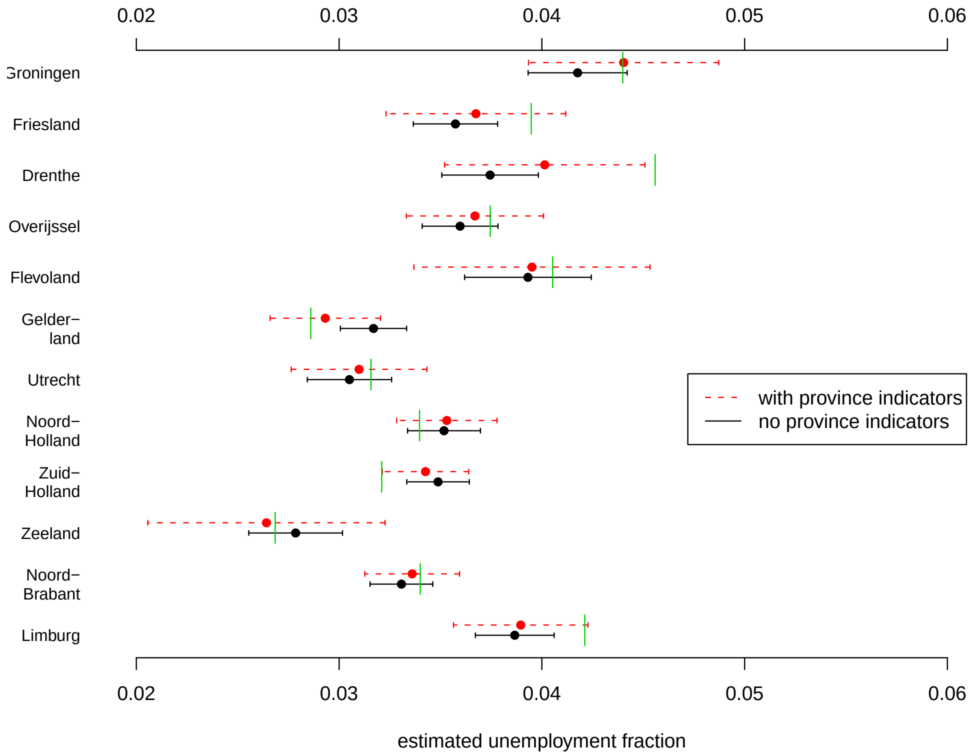

Figure 7. Municipal unemployment estimates aggregated to the provincial level based on a model with or without province indicators and 2009 data. The lines are approximate 95% intervals. The vertical bars denote the direct provincial estimates.

Finally, we discuss the benefit of including the term $ru - area$, the area level registered unemployment fraction. As explained in Bafumi and Gelman (2006) this removes the correlation between the unit-level predictors and the area random effects, thus improving the validity of the model assumptions. This is illustrated in Figure 8, in which the correlation between estimated municipality effects and municipal population means is displayed for a selection of covariates. The two models considered are the selected model discussed previously, and the same model without $ru - area$. The figure is based on 2009 data, but the same pattern shows in all years. Not only does the area-level covariate $ru - area$ remove the correlation of the random effects with ru population means, but it also reduces most other such correlations. Note that because municipality is nested within province, the province indicators are area-level covariates, and so are uncorrelated with the area level errors by construction.

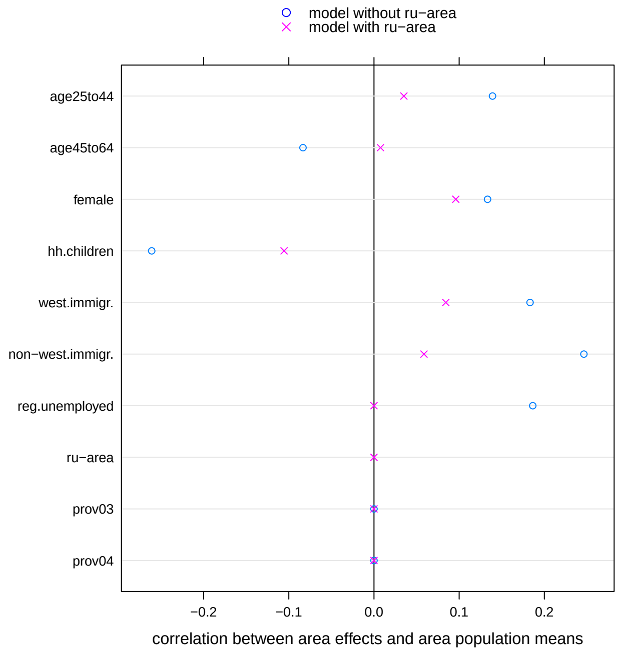

Figure 8. Correlation between estimated area (municipality) effects and population area means with or without including registered unemployment as an additional area level covariate.

7.2 Municipal small area estimates for unemployment

Using the model discussed in the previous section, municipal estimates of unemployment are obtained for all municipalities. Each of these estimates has standard errors associated with it. We will now consider these results and compare them to the GREG estimates.

Table 1 lists the mean coefficients of variation (cv) for the GREG estimates for 2009. At present, GREG estimates of annual unemployment fractions are published for municipalities with more than 30,000 inhabitants. The cv can be used as a quality criterion. As an example we have chosen a threshold of 20%, which means that we accept a standard error which is less than 1/5th of the point estimate, and consider estimates with larger standard errors as poor. This choice is somewhat arbitrary (see below). From Table 1 we see that only 57% of the 79 municipalities with more than 30,000 inhabitants have a cv of less than 20%. By this criterion, the quality of 43% of the published municipal estimates is not good.

Table 2 is the equivalent of Table 1 for the SAE estimates of municipal unemployment for 2009. It is immediately obvious that the mean cv values of the SAE estimates are much smaller than those of the GREG estimates. Using the same 20% criterion, all large municipalities pass, almost all of the medium sized municipalities do, and with 85% even most of the smallest municipalities have a cv of less than 20%. This is a large improvement in precision compared to the GREG estimates. Important to note is that SAE methods are able to provide reliable estimates for two of the four municipalities that were out of

Table 1. Mean coefficients of variation of municipal GREG estimates of unemployment based on 2009 LFS data

Municipalities	N	mean cv	cv<0.2	cv<0.2 (%)
>30K inhabitants	79	0.2097	45	57%
10-30K inhabitants	233	0.8277	5	2%
<10K inhabitants	125	2.6941	1	1%
all (in sample)	437	1.2498	51	12%
out of sample	4	-	-	-

sample. There are estimates for the other two as well, but with large standard errors.

Table 2. Mean coefficients of variation of municipal SAE estimates of unemployment based on 2009 LFS data

Municipalities	N	mean cv	cv<0.2	cv<0.2 (%)
>30K inhabitants	79	0.0908	79	100%
10-30K inhabitants	233	0.1316	229	98%
<10K inhabitants	125	0.1583	106	85%
all (in sample)	437	0.1319	414	95%
out of sample	4	0.2223	2	50%
all	441	0.1327	416	94%

Depending on the particular sample obtained each year, and the model fit in that year, the number of municipalities with a cv of less than 20% varies over time. In addition, varying the threshold level alters the number of municipalities having a cv not exceeding that level. This is shown in Table 3. From 2003 onwards, SAE estimates are reliable for the vast majority of municipalities according to the 20% criterion. As mentioned earlier, the threshold of 20% is chosen in the tables presented above, but other choices are equally valid. Table 3 gives an indication of how the percentage of municipalities not exceeding the threshold varies with the level of the threshold. Clearly, the smaller the threshold, the fewer municipalities do not exceed it. A criterion for the cv value could be used in deciding whether or not to publish a particular municipal estimate.

7.3 Municipal estimates for the employed labour force

Besides the unemployed labour force, we also need to estimate the employed labour force. Only then is it possible to calculate the unemployment rate per municipality, defined as the fraction of unemployed within the total labour force consisting of both employed and unemployed persons. The search for auxiliary variables to estimate the employed labour force turns out to be quite involved.

Table 3. Percentage of municipalities not exceeding various cv thresholds

Year	cv<0.10	cv<0.15	cv<0.20	cv<0.25 (%)
2001	4%	28%	61%	86%
2002	14%	54%	84%	97%
2003	36%	81%	96%	99%
2004	43%	89%	99%	100%
2005	65%	96%	99%	100%
2006	48%	95%	99%	100%
2007	15%	61%	93%	99%
2008	17%	67%	95%	99%
2009	21%	70%	94%	99%

All auxiliary variables that are used in estimating the unemployed labour force, can be used to estimate the employed labour force. Demographic characteristics are relevant and registered unemployment should correlate negatively with being employed. The employed labour force also has to be calibrated on a provincial level, hence province needs to be included in the model. However, we also need auxiliary variables that specifically correlate with the employed, just like registered unemployment does for the unemployed.

The Polis register contains income data for all persons working as an employee in the Netherlands. This is monthly data which becomes available with a couple of months delay. Polis data of sufficient quality is available from 2008 onwards. We link this data to the persons living in a municipality. Consequently it is possible to use this data for SAE estimation.

The timeliness of data becomes a problem for two other auxiliary variables that are needed. Since the Polis register only contains data on people that are an employee in the Netherlands, we would underestimate the employed labour force in municipalities close to the border. Especially for municipalities in the north (Groningen) and south (Limburg) this turns out to be the case. Therefore we need an auxiliary variable that indicates whether a person receives income from labour in a foreign country. This auxiliary variable becomes available with a delay of approximately nine months. Therefore we have to use the information of one year earlier when estimating a certain year.

The last group of employed for which data is needed are the self-employed. This group is not spread evenly across the country and therefore the estimation results would be skewed without including an auxiliary variable for them. Data for self-employed becomes available with a delay of two to three years. Therefore we have to do with the most recent data available.

We estimate the employed labour force for each municipality in 2008. The model of Section 7.1 is augmented with auxiliary variables on income from labour. Polis data of

2008 is used in combination with income from abroad data of 2007 and data on self-employment from 2005. For each municipality the estimated coefficient of variation for the employed labour force lies below 0.05.

For earlier years auxiliary variables can be obtained from the social statistical file (SSB). However, it is not possible to make a series of estimates with the same model, as is done for the unemployed labour force from 2001 to 2009. Furthermore, the exact model specification has to be decided on a yearly basis since the quality of the auxiliary variables changes.

The auxiliary variables relating to income from labour can also be used in estimating the unemployed labour force. The correlation between these auxiliary variables and unemployment should be negative. For 2008, including the auxiliary variables into the model of Section 7.1 results in lower standard errors for almost all municipalities. The percentage of municipalities where the estimate for the unemployed labour force does not exceed the coefficients of variation threshold of 0.20 increases from 95% to 98%.

7.4 Municipal estimates for the inactive population

The auxiliary variables that are used to estimate the employed and unemployed labour force, can also be used to estimate the inactive population. Furthermore, the Polis register contains data on the social benefits that people receive. This data can be used as auxiliary variables when estimating the inactive population. For each municipality the SAE estimate of the inactive population has a coefficient of variation below 0.10. The sum of the employed, unemployed and inactive must equal the population in each municipality. This leads to an extra set of calibration requirements.

A much more simple method to determine the inactive population per municipality, is by subtracting the total labour force from the population. The difference in outcome between this approach and SAE estimation is small. Therefore we decided that the municipal inactive population is determined by subtracting the total labour force from the population. This also means that calibration is only required for the employed and unemployed labour force on a provincial level.

8 Conclusions and further work

This paper discusses small area estimation of labour status in municipalities in the Netherlands. The estimates are based on data from the Labour Force Survey and make use of a number of covariates obtained from administrative registers. The basic unit-level model is used, which is a linear mixed model with random area effects. This allows for more efficient use of the auxiliary data than the direct estimation methods that are currently used in regular official production. The municipal small area estimates for unemployment are more plausible than the direct estimates, in particular with respect to stability over time. Small and medium sized municipalities benefit more from the model-based approach than large municipalities, for which direct estimates are often good because of the larger sample sizes. The estimated MSE of the small area estimates is much smaller than that of the direct estimates. When using this as a measure of quality, small area methods are capable of providing good estimates for 94% of all municipalities, while good direct estimates are available for only 12% of all municipalities.

Further studies of the application of the small area estimation model will be conducted along a number of lines. First, we intend to study how the calibration to direct estimates at the provincial level affects MSEs for the small area estimates. For unemployment, some of the relative calibration adjustments to the municipal estimates are larger than 10%, so one cannot expect the influence on MSEs to be negligible. To clarify this issue involves obtaining a better understanding of the differences between the direct and model-based estimates at the provincial level.

Eventually, municipal unemployment estimates should be produced for 6 age-gender combinations. The basic unit-level model with municipalities as area effects can also be used to estimate the unemployment means for the age-gender combinations for all municipalities, since it allows prediction at the unit level. This requires that all population information be specified at the municipality $\times$ gender $\times$ age level. For this purpose, it may be beneficial to include additional interaction terms with age and gender in the model. Small area methods will also be applied to the estimation of municipal educational levels.

Although the model proposed produces estimates that are superior to direct estimates, further improvements to the model are imaginable. Such improvements may be sought in the use of geographical positions of municipalities, borrowing of strength over time and a study of the best variance structures to be used in the model. Further diagnostic checks may lead the way.

Finally, changes in the LFS design carried out in this and coming years may require changes to the model. In particular, labour status and educational level will be observed in a basic questionnaire that precedes all social surveys. This data should be used in addition to that observed in the LFS. In order to combine data observed under different (mixed-mode) designs it is important to further study ways to deal with possible differences in measurement errors.

References

Bafumi, J. and A. Gelman (2006). Fitting Multilevel Models When Predictors and Group Effects Correlate, prepared for the 2006 Annual Meeting of the Midwest Political Science Association, Chicago (unpublished).

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

Bell, W.R. (1999). Accounting for Uncertainty About Variances in Small Area Estimation. Bulletin of the International Statistical Institute.

Boonstra, H.J. (2004). Calibration of Tables of Estimates. BPA-no. M336-04-TMO, Statistics Netherlands, Heerlen.

Boonstra, H.J. (2005). Model-based estimation of a finite population total: a Bayesian approach. BPA-no. TMO-R&D-2005-08-16-HBTA, Statistics Netherlands, Heerlen.

Boonstra, H.J., J.A. Van den Brakel, B. Buelens, S. Krieg and M. Smeets (2008). Towards small area estimation at Statistics Netherlands, Metron Vol. LXVI, Nr. 1, p. 21−49.

Boonstra, H.J., B. Buelens and M. Smeets (2007). Estimation of municipal unemployment fractions - a simulation study comparing different small area estimators. Internal report, BPA-no. DMK-DMH-2007-04-20-HBTA, Statistics Netherlands, Heerlen.

Boonstra, H.J., B. Buelens, Chi Ma and M. Smeets (2009). Model selection for small area estimation, BPA-no. DMH-2009-01-22-HBTA, Statistics Netherlands, Heerlen.

van den Brakel, J. and S. Krieg (2009). Structural Time Series Modelling of the Monthly Unemployment Rate in a Rotating Panel, Survey Methodology Vol. 35, No. 2, p. 177-190.

Buelens, B., H.J. Boonstra and M. Smeets (2007). Approaches to model selection in small area estimation. Internal report, BPA-no. DMH-2007-06-22-BBUS, Statistics Netherlands, Heerlen.

Buelens, B., H.J. Boonstra and M. Smeets (2009). Estimation of the variance components of area level models for small area estimation. DMH-2009-01-06-BBUS, Statistics Netherlands, Heerlen.

Datta, G.S. and M. Ghosh (1991). Bayesian Prediction in Linear Models: Applications to Small Area Estimation, The Annals of Statistics, Vol. 19, No. 4, p. 1748−1770.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models, Bayesian Analysis 1, No. 3, p. 515−533.

Knottnerus, P. (2002). Sample survey theory: some Pythagorean perspectives, Springer-Verlag, New York.

McCulloch, C.E. and S.R. Searle (2001). Generalized, Linear and Mixed Models. John Wiley, New York.

Neuhaus, J.M. and C.E. McCulloch (2006). Separating between- and within-cluster covariate effects by using conditional and partitioning methods, Journal of the Royal Statistical Society B, 68, Part 5, p. 859−872.

Pfeffermann, D. and C.H. Barnard (1991). Some new estimators for small-area means with application to the assessment of Farmland values, Journal of Business & Economic Statistics, p. 73−84.

Rao, J.N.K. (2003). Small area estimation. John Wiley, New York.

Rubin, D.B. (1976). Inference and missing data, Biometrika 63, p. 581−592.

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

Wang, J., W.A. Fuller and Y. Qu (2008). Small area estimation under a restriction, Survey Methodology Vol. 34, No. 1, p. 29−36

Appendix A The basic unit level model

Let us denote by θ_i the quantity of interest for area i, for example the fraction of employed persons in the target population. If y is the underlying variable at the unit, i.e. person level, e.g. the binary 0/1 variable employed, then

$$\theta_i = \frac{1}{N_i}\sum_{j=1}^{N_i} y_{ij}\,,$$

where y_{ij} is the variable of interest for unit j in area i, and N_i is the population size of area i. The number of areas in the population will be denoted by M and are sample sizes by n_i. Not all areas are necessarily represented in the available sample data, and m is used to denote the number of areas for which data are available. Within a year, the Dutch LFS collects data in all municipalities, except for about three to five small municipalities (mainly islands). Estimates for the missing municipalities can still be made using the model, since auxiliary information is available for all M municipalities.

The basic linear unit level model is given by

$$y_{ij} \overset{\text{ind}}{\sim} \mathcal{N}(\beta' x_{ij} + v_i, \sigma_e^2)\,, \qquad i = 1 \ldots M\,,\quad j = 1 \ldots N_i\,, \tag{A.1}$$

$$v_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_v^2)\,, \qquad i = 1 \ldots M\,. \tag{A.2}$$

Here y_{ij} is the variable of interest for unit j within area i, x_{ij} is a corresponding p-vector of auxiliary characteristics, β are associated fixed effects, and σ_e^2 is a residual variance parameter. The area effects v_i are modelled as normal with zero mean and variance σ_v^2. BLUP predictors based on this linear mixed model are (see e.g. Rao, 2003)

$$\begin{aligned}
\tilde{\theta}_i &= \frac{n_i}{N_i}\bar{y}_i + \tilde{\beta}'\left(\bar{X}_i - \frac{n_i}{N_i}\bar{x}_i\right) + (N_i - n_i)\tilde{v}_i & \text{(A.3)}\\
&= \gamma_i\left(\bar{y}_i + \tilde{\beta}'\left(\bar{X}_i - \bar{x}_i\right)\right) + (1-\gamma_i)\left(\frac{n_i}{N_i}\bar{y}_i + \tilde{\beta}'\left(\bar{X}_i - \frac{n_i}{N_i}\bar{x}_i\right)\right)\,, & \\
\tilde{\beta} &= \left(X'\tilde{\Sigma}^{-1}X\right)^{-1} X'\tilde{\Sigma}^{-1} y_s\,, \qquad \tilde{v}_i = \gamma_i(\bar{y}_i - \tilde{\beta}'\bar{x}_i)\,, & \text{(A.4)}\\
\gamma_i &= \frac{\sigma_v^2}{\sigma_v^2 + \sigma_e^2/n_i}\,. & \text{(A.5)}
\end{aligned}$$

Here $\bar{x}_i$ is a p-vector of sample means for area i, $\bar{X}_i$ is the corresponding vector of population means, X is the $n \times p$ sample matrix of covariates, $\bar{y}_i$ are area sample means of y, y_s is the n-vector of response values, and $\tilde{\Sigma} = \text{cov}(y_s) = \sigma_e^2 I_n + \sigma_v^2 \oplus_{i=1}^m J_{n_i}$, where I_n is the n-dimensional identity matrix, J_{n_i} the $n_i \times n_i$ matrix with all elements 1, and $\oplus_{i=1}^m J_{n_i}$ the block diagonal matrix with blocks J_{n_i}. The BLUP $\tilde{\theta}_i$ is a weighted average of a survey regression estimate with weight γ_i and a regression synthetic estimate $\tilde{\beta}'\bar{X}_i$ modified by a finite population correction term. In the LFS the finite population corrections are negligible since in all municipalities only a small fraction of the population is observed.

In frequentist approaches, σ_e^2 and σ_v^2 are estimated, typically as the maximizers of the restricted likelihood, and plugged into the BLUP $\tilde{\theta}_i$ to yield EBLUP small area estimates. In the Bayesian approach, as detailed in Datta and Ghosh (1991), one averages over the joint posterior density for σ_e^2 and σ_v^2. After a transformation $(\sigma_e^2, \sigma_v^2) \rightarrow (\sigma_e^2,\ \lambda \equiv \sigma_v^2/\sigma_e^2)$, the integration over σ_e^2 can be carried out analytically for suitable choices of prior densities, and only the one-dimensional integration over λ remains to be carried out numerically. Note that the BLUP predictors $\tilde{\theta}_i$ given above are functions of λ only; they are the posterior means of θ_i given λ.

The remaining integration over the variance ratio λ must be carried out numerically. The prior we use is simply

$$p(\beta, \sigma_e^2, \lambda) \propto 1\,, \tag{A.6}$$

which differs from the priors used in Datta and Ghosh (1991). In that paper, independent inverse gamma priors for the variance parameters σ_e^2 and σ_v^2 were used. In our application such priors give results that can be quite sensitive to the specific choice of parameters of the σ_v^2 prior.[2] Note that the prior (A.6) corresponds to $p(\beta, \sigma_e^2, \sigma_v^2) \propto 1/\sigma_e^2$ in the original parametrization.

The hierarchical Bayes (HB) small area estimates are now obtained as

$$\hat{\theta}_i = E(\theta_i|y_s) = E(E(\theta_i|y_s, \lambda)|y_s) = \int_0^\infty \tilde{\theta}_i\, p(\lambda|y_s)\mathrm{d}\lambda\,, \tag{A.7}$$

where all expectations are posterior expectations that condition on the observed data vector denoted by y_s, and $p(\lambda|y_s)$ is the posterior density for λ. Up to a constant of proportionality, this density is

$$p(\lambda|y_s) \propto |\Sigma|^{-1/2}|X'\Sigma^{-1}X|^{-1/2}\left((y_s - X\tilde{\beta})'\Sigma^{-1}(y_s - X\tilde{\beta})\right)^{-(n-p)/2}, \tag{A.8}$$

where $\Sigma = \tilde{\Sigma}/\sigma_e^2 = I_n + \lambda \oplus_{i=1}^m J_{n_i}$.

For the mean squared errors, we have

$$\begin{aligned} V_i &= V(\theta_i|y_s) = E(V(\theta_i|y_s,\lambda)|y_s) + V(E(\theta_i|y_s,\lambda)|y_s) \\ &= \int p(\lambda|y_s)\left(\tilde{V}_i + \left(\tilde{\theta}_i - \hat{\theta}_i\right)^2\right)\mathrm{d}\lambda\,, \end{aligned} \tag{A.9}$$

where $\tilde{V}_i$ is the posterior variance of θ_i given λ. These are the diagonal elements of the $M \times M$ posterior covariance matrix given λ, whose general (i_1, i_2) element is,

$$\begin{aligned} V(\theta_{i_1}, \theta_{i_2}|y_s, \lambda) = \tilde{\sigma}_e^2 \Bigg\{ & \left(\frac{1-f_{i_1}}{N_{i_1}} + (1-f_{i_1})^2\frac{\gamma_{i_1}}{n_{i_1}}\right)\delta_{i_1 i_2} \\ & + (\bar{X}_{i_1} - f_{i_1}\bar{x}_{i_1} - (1-f_{i_1})\gamma_{i_1}\bar{x}_{i_1})'\left(X'\Sigma^{-1}X\right)^{-1}(\bar{X}_{i_2} - f_{i_2}\bar{x}_{i_2} - (1-f_{i_2})\gamma_{i_2}\bar{x}_{i_2})\Bigg\}\,, \end{aligned}$$

where $f_i = n_i/N_i$ is the sampling fraction in area i, $\delta_{i_1 i_2}$ is Kronecker's delta equaling one if $i_1 = i_2$ and zero otherwise, and

$$\tilde{\sigma}_e^2 = \frac{1}{n-p-2}(y_s - X\tilde{\beta})'\Sigma^{-1}(y_s - X\tilde{\beta})\,, \tag{A.10}$$

[2]These parameters must be small for the prior to be relatively uninformative, but cannot both be zero, since that would yield an improper posterior density. See Gelman (2006) for a discussion of these and other issues concerning priors for variance parameters in hierarchical models.

is the posterior mean for σ_e^2 given λ. The estimates and the posterior variances can thus be computed using one-dimensional numerical integration. Note that in addition to the $2M$ integrals to compute the posterior means and variances, one needs an additional one-dimensional numerical integration to compute the normalization constant in (A.8).

For aggregation of small area estimates, and in particular for benchmarking of the estimates to known or fixed aggregates, the complete posterior covariance matrix for θ is used. The posterior covariance matrix can be obtained by integrating over the posterior density for λ as in (A.9). Although this could be done, it would require an additional $\frac{1}{2}M(M-1)$ one-dimensional numerical integrations, and so we prefer to approximate the posterior covariance matrix by a plug-in estimate $D\tilde{V}(\hat{\lambda})D$ where $D = \text{diag}(V_1/\tilde{V}_1(\hat{\lambda}), \ldots, V_M/\tilde{V}_M(\hat{\lambda}))$, where $\hat{\lambda}$ is the posterior mean for λ. Thus we only carry out the integrations for the diagonal elements, and use plug-in estimates for the correlations.

All formulas given also hold in the presence of areas without observations after taking the limit $n_i \to 0$ in such areas. In this limit $\gamma_i \to 0$, so that, as can be seen from (A.3)-(A.5), $\hat{\theta}_i$ reduces to the regression prediction $\hat{\beta}'\bar{X}_i$ with $\hat{\beta}$ the posterior mean of β, i.e. the average of $\tilde{\beta}$ over the posterior density for λ.

A.1 A hybrid approach

In Bell (1999) and Buelens et al. (2009), a hybrid Bayesian/EBLUP approach is discussed, in which the posterior mean for the between area variance parameter is used as a plug-in estimate in the BLUPs based on an area level model. The same hybrid approach can be applied to the unit level model by using the posterior mean for λ as a plug-in estimate. This approach only requires two one-dimensional numerical integrations to obtain the posterior mean of the variance parameter. The $2M$ numerical integrations to obtain the estimates and MSEs are no longer needed. Like the full HB approach, the hybrid approach does not suffer from zero estimated variance parameter; its posterior mean is guaranteed to be positive. In the LFS application the hybrid approach leads to very good approximations of the full HB estimates. There is hardly any difference in point estimates, a consequence of the large overall LFS sample size. The only notable difference is that for a few areas with large area effects, full HB posterior variances can be somewhat (up to about 25%) larger than the hybrid plug-in variances, even though on average over all areas they are almost equal. The differences are due to uncertainty about the variance parameter, which is not taken into account in the hybrid approach (Bell, 1999).

A.2 Inclusion of measurement errors

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

$$y_{ij} \overset{\text{ind}}{\sim} \mathcal{N}(Y_{ij} + \alpha' z_{ij}, \sigma_m^2)\,, \qquad i = 1 \ldots m\,, \quad j = 1 \ldots n_i\,, \tag{A.11}$$

$$
\begin{aligned}
Y_{ij} &\overset{\text{ind}}{\sim} \mathcal{N}(\tilde{\beta}'\tilde{x}_{ij} + v_i, \sigma_e^2)\,, \qquad i = 1 \ldots M\,, \quad j = 1 \ldots N_i\,, \\
v_i &\overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_v^2)\,, \qquad i = 1 \ldots M\,,
\end{aligned} \tag{A.12}
$$

where the first line represents the measurement part of the model. In our application we simplify this model by assuming $\sigma_m^2 = 0$, so that the measurement error is a fixed bias of size $\alpha' z_{ij}$. In our case z is a single dummy variable taking the value 1 for 5th wave observations, and 0 for 1st wave observations, and α is the corresponding coefficient, which can be interpreted as the rotation group bias, controlling for the covariates $\tilde{x}$. Strictly speaking, this model is inappropriate for binary variables with classification errors, but nevertheless we believe it works well for prediction at aggregate levels, in which we are interested. We also note that the above model specification assumes that z is not predictive of Y, but only of measurement bias in the observed values y. The validity of this assumption is improved by incorporating into x variables associated with Y and with response probabilities for wave 1 and 5 respondents.

From (A.11) and (A.12) we have

$$
y_{ij} \overset{\text{ind}}{\sim} \mathcal{N}(Y_{ij} + \alpha z_{ij} + \tilde{\beta}'\tilde{x}_{ij} + v_i, \sigma_e^2)\,, \qquad i = 1 \ldots m\,, \quad j = 1 \ldots n_i\,.
$$

This shows that the likelihood function of the model parameters is exactly the same as for the model without measurement error but with covariate vector $x = (z, \tilde{x}')'$ and coefficients $\beta = (\alpha, \tilde{\beta}')'$. Provided α is assigned a uniform prior as well, inference about model parameters goes through in the same way as before. Prediction is based on x only, not on z, so the coefficient α is not involved directly in prediction, only through its indirect effects on the partial regression coefficients $\tilde{\beta}$. Inferences about small area means can still be based on the model without measurement error and covariate vector $x = (z, \tilde{x}')'$ by associating zero area population means with z. This removes the bias effect of the fifth relative to the first wave from the predictions.

Appendix B Description of variables

The following variables are used in the model:

- `unemployed`: binary indicator variable for unemployment (survey variable)
- `gender`: male or female
- `age3`: age in three categories (15-25, 25-45, 45-65)
- `ru1`: binary indicator for registered unemployment
- `ethnicity3`: ethnicity in three categories (native, native from other Western country, non-Western)
- `ethnicity2D`: ethnicity in seven categories (native, native from other Western country, Turkish, Moroccan, Surinamer, Antilles, native from other non-Western county)

- `ru`: registered unemployment in five categories (not-registered, registered but working, not working and registered since 0, 1-4, or 4+ years)
- `age`: age in five categories (15-25, 25-35, 35-45, 45-55, 55-65)
- `typehh`: household type in three categories (Single person, household with children, other)
- `prov`: province (there are 12 in the Netherlands)
- `ru-area`: area means of registered unemployment
- `wave`: wave of the LFS when this person was observed (1st or 5th).