Model evaluation for multivariate structural time series models for the Dutch Labour Force Survey

Sabine Krieg, Jan van den Brakel
Department of Statistical Methods, Statistics Netherlands, Heerlen, Netherlands

Abstract

In this paper different multivariate structural time series models are described and applied to estimate the monthly unemployment rate of the Dutch Labour Force Survey. The estimation results are compared in a model evaluation. Compared to the generalized regression estimator, the time series approach results in a substantial increase of the precision because this approach uses sample information observed in previous time periods and other domains to improve the monthly estimates.

KEY WORDS: Structural time series models, Small area estimation

1. Introduction

Generalized regression (GREG) estimators are widely applied by national statistical institutes since they are always approximately design unbiased. They have, however, relatively large design variances in the case of small sample sizes. In this situation, model based small area estimators can be used to improve the precision of the estimates, since they have smaller variance than the GREG estimator. Model misspecification, on the other hand, can result in seriously biased estimates. The application of model based procedures for official statistics therefore requires careful model selection and evaluation.

In this paper, structural time series models are used to produce model based estimates for the monthly unemployment rate of the Dutch Labour Force Survey (LFS). With these models, sample information from other time periods and from other domains is borrowed to improve the estimates. Key references to the use of structural time series models in the context of small area estimation are Pfeffermann and Burck (1990), and Pfeffermann and Bleuer (1993). The unemployment rate is defined as the ratio of the total unemployment and the total labour force.

In section 2 the Dutch LFS is summarized. In section 3 the multivariate structural time series model is described for the monthly unemployment rate for six demographic domains. In section 4, estimation results for 8 different models are described. Model evaluation techniques are used to compare the models in section 5. Some general remarks are made in section 6.

2. The Dutch Labour Force Survey

The objective of the Dutch LFS is to provide reliable information about the labour market. The LFS is based on a stratified two-stage cluster design of addresses. Strata are formed by geographical regions. Municipalities are considered as primary sampling units and addresses as secondary sampling units. All households residing on an address, up to a maximum of three, are included in the sample (there is generally one household per address in the Netherlands).

In this paper, the data of the LFS from January 1996 until December 2006 are used. Until September 1999, the LFS was a continuous survey. In October 1999, the LFS changed to a rotating panel design, where the respondents are re-interviewed four times at quarterly intervals. The data from these re-interviews are not used in this paper. A structural time series model that makes advantages of the rotating panel design, is described in Van den Brakel and Krieg (2007).

The weighting procedure of the LFS is based on the GREG estimator (Särndal et al., 1992). The weighting scheme consists of a combination of different social-demographical categorical variables. Because the monthly sample size of the Dutch LFS is too small to publish reliable monthly figures using the GREG estimator, moving averages over the preceding three months are published.

3. A structural times series model for six domains

Let θ_t denote the population parameter at time *t*, e.g. the true unemployment rate for month *t*. Direct estimators, like the GREG estimator, assume that θ_t is a fixed but unknown parameter. Under this design-based approach, an estimator for θ_t for cross-sectional surveys only uses the data observed at time *t*. Data from the past are only used in the case of partially overlapping samples in a panel design, but not in the case of repeatedly conducted cross-sectional designs. Scott and Smith (1974) proposed to consider the population parameter θ_t as a realization of a stochastic process that can be described with a time series model. Under this assumption, data observed in preceding periods *t*-1, *t*-2, ..., can be used to improve the estimator for θ_t. In the context of small area estimation this is called borrowing strength in time. Sample information from different domains can be used to further improve the domain estimates, which is known as borrowing strength in space. The common approach is to allow for random area and random time effects in a linear mixed model, and apply a composite estimator like the BLUP or EBLUP, see e.g. Rao and Yu (1994). Pfeffermann and Burck (1990) and Pfeffermann and Bleuer (1993) proposed to model the correlation between the model parameters of the domains in a multivariate structural time series model. Under this approach, it is also possible to borrow strength in space by assuming that model parameters for different domains are equal.

GREG estimates $Y_{t,d}$ for the true unemployment rate $\theta_{t,d}$ of domain d and month t based on monthly samples are produced for the following six domains: (1) Men, 15-24 year, (2) Women, 15-24 year, (3) Men, 25-44 year, (4) Women, 25-44 year, (5) Men, 45-64 year, (6) Women, 45-64 year. So each month a vector $\mathbf{Y}_t = (Y_{t,1}\ Y_{t,2}\ Y_{t,3}\ Y_{t,4}\ Y_{t,5}\ Y_{t,6})^T$ is observed, which can be modelled as

$$\mathbf{Y}_t = \boldsymbol{\theta}_t + \mathbf{e}_t\ , \tag{3.1}$$

with $\boldsymbol{\theta}_t = (\theta_{t,1}\ \theta_{t,2}\ \theta_{t,3}\ \theta_{t,4}\ \theta_{t,5}\ \theta_{t,6})^T$ a vector with true monthly unemployment rates for the six domains and $\mathbf{e}_t = (e_{t,1}\ e_{t,2}\ e_{t,3}\ e_{t,4}\ e_{t,5}\ e_{t,6})^T$ the corresponding survey errors for each domain estimate.

With a structural time series model, the population parameter can be decomposed in a trend, a seasonal, and an irregular component, i.e.:

$$\boldsymbol{\theta}_t = \mathbf{L}_t + \mathbf{S}_t + \boldsymbol{\varepsilon}_t\ , \tag{3.2}$$

where $\mathbf{L}_t = (L_{t,1}\ L_{t,2}\ L_{t,3}\ L_{t,4}\ L_{t,5}\ L_{t,6})^T$ denotes the trend, $\mathbf{S}_t = (S_{t,1}\ S_{t,2}\ S_{t,3}\ S_{t,4}\ S_{t,5}\ S_{t,6})^T$ the seasonal, and $\boldsymbol{\varepsilon}_t = (\varepsilon_{t,1}\ \varepsilon_{t,2}\ \varepsilon_{t,3}\ \varepsilon_{t,4}\ \varepsilon_{t,5}\ \varepsilon_{t,6})^T$ the irregular components. For the trend components the smooth trend model is assumed, which is defined by the following set of equations:

$$\begin{aligned} L_{t,d} &= L_{t-1,d} + R_{t-1,d}\ , \\ R_{t,d} &= R_{t-1,d} + \eta_{R,t,d}\ , \\ E(\eta_{R,t,d}) &= 0, \\ Cov(\eta_{R,t,d}, \eta_{R,t',d'}) &= \begin{cases} \sigma^2_{R,d} & \text{if} \quad t = t' \quad \text{and} \quad d = d' \\ \varsigma_{R,d,d'} & \text{if} \quad t = t' \quad \text{and} \quad d \neq d' \\ 0 & \text{if} \quad t \neq t'. \end{cases} \end{aligned} \tag{3.3}$$

The parameters $L_{t,d}$ and $R_{t,d}$ are referred to as the trend and the slope parameter respectively. The seasonal components are modelled as

$$\begin{aligned} \sum_{j=0}^{11} S_{t-j,d} &= \eta_{S,t,d}\ , \\ E(\eta_{S,t,d}) &= 0, \\ Cov(\eta_{S,t,d}, \eta_{S,t',d'}) &= \begin{cases} \sigma^2_{S,d} & \text{if} \quad t = t' \quad \text{and} \quad d = d' \\ \varsigma_{S,d,d'} & \text{if} \quad t = t' \quad \text{and} \quad d \neq d' \\ 0 & \text{if} \quad t \neq t'. \end{cases} \end{aligned} \tag{3.4}$$

In the models considered in this paper, it is assumed that the seasonal effects are uncorrelated, i.e. $\varsigma_{S,d,d'} = 0$. The irregular components $\varepsilon_{t,d}$ contain the unexplained variation and are modelled as independent white noise processes. Combining equation (3.1) and equation (3.2) yields:

$$\mathbf{Y}_t = \boldsymbol{\theta}_t + \mathbf{e}_t = \mathbf{L}_t + \mathbf{S}_t + \mathbf{e}_t + \boldsymbol{\varepsilon}_t = \mathbf{L}_t + \mathbf{S}_t + \mathbf{v}_t\ ,$$

with $\mathbf{v}_t = (\nu_{t,1}\ \nu_{t,2}\ \nu_{t,3}\ \nu_{t,4}\ \nu_{t,5}\ \nu_{t,6})^T$ the sum of the survey errors and the irregular component of the true population parameter. The components $\nu_{t,d}$ are modelled as white noise processes:

$$\begin{aligned} E(\nu_{t,d}) &= 0, \\ Cov(\nu_{t,d}, \nu_{t',d'}) &= \begin{cases} \sigma^2_{\nu,d} & \text{if} \quad t = t' \quad \text{and} \quad d = d' \\ 0 & \text{if} \quad t \neq t' \quad \text{or} \quad d \neq d'. \end{cases} \end{aligned}$$

This structural time series model can be put in state space representation. Subsequently the Kalman filter can be applied to obtain optimal estimates for the model parameters, see Durbin and Koopman (2001). The analysis is conducted with Ssfpack beta 3 (Koopman e.a., 1999) in combination with Ox (Doornik, 1998). Note that a more recent version of Ssfpack is used than the 2.2 version described in Koopman et al. (1999).

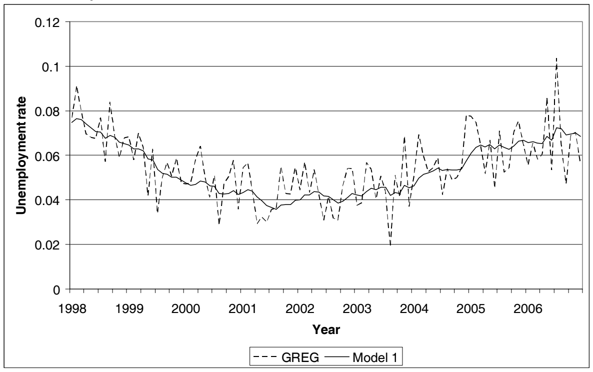
4. Estimation results for different models

In this section different models are considered, which are special cases of the general model formulation given in section 3.

Model 1 assumes a univariate trend model for each domain, i.e. $S_{t,d} = 0$ for all t and d and $\varsigma_{R,d,d'} = 0$ for all $d \neq d'$.

In Figure 1 the filtered estimates of model 1 are compared with the GREG estimates for domain 6. The filtered estimates are used because they are based on the complete set of information that would be available if the model were used to produce an estimate for month t for regular publication purposes. The irregularities in the series of the GREG estimates are considered as survey errors under the time series model, and are therefore flattened out in the model estimates. Some of these irregularities, however, are seasonal effects. This implies that the model is misspecified and the estimates are biased. The standard errors of these estimates, plotted in Figure 3, do not reflect this bias, and are therefore not a good measure for the accuracy of this model.

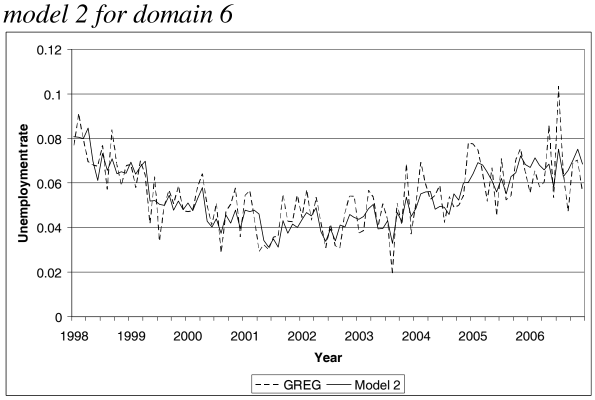
Figure 1: GREG estimates and filtered estimates of model 1 for domain 6



Model 2 assumes a univariate model for each domain that contains a trend and a seasonal component. Thus $\varsigma_{R,d,d'} = 0$ for all $d \neq d'$.

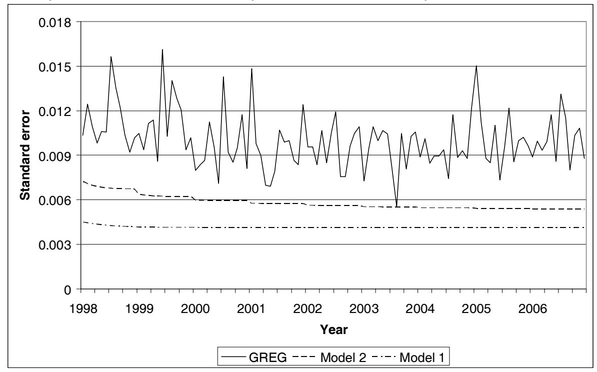
In Figure 2 the filtered estimates of model 2 are compared with the GREG estimates for domain 6. The filtered estimates partly follow the fluctuations in the GREG series, since they are considered as time dependent seasonal effects under model 2. Nevertheless a substantial part of the irregularities in the series of the GREG estimates are flattened out, since they are considered as survey errors under the time series model.

Figure 2: GREG estimates and filtered estimates of model 2 for domain 6



In Figure 3 the standard errors of the filtered estimates of model 1 and 2 are compared with the standard errors of the GREG estimates. The standard errors of the filtered estimates are substantially smaller, since the time series models borrow strength from the past. The standard errors of model 1 are smaller than the standard errors of model 2. Nevertheless model 2 has to be preferred because a model with seasonal components is more realistic. Table 1 shows the means of the standard errors for the other domains.

Figure 3: The standard errors of the GREG estimates and filtered estimates of model 1 and 2 for domain 6



The domains can be classified in two groups with more or less equal seasonal effects. The seasonals of domain 1 and 2 follow a similar pattern; the smoothed estimates, obtained with the fixed interval smoother, are shown in Figure 4. The difference in February, however, is quite large. The seasonals of the other 4 domains also follow a similar pattern; the smoothed estimates are shown in Figure 5.

Figure 4: Seasonal effects of domain 1 and domain 2, smoothed estimates of model 2

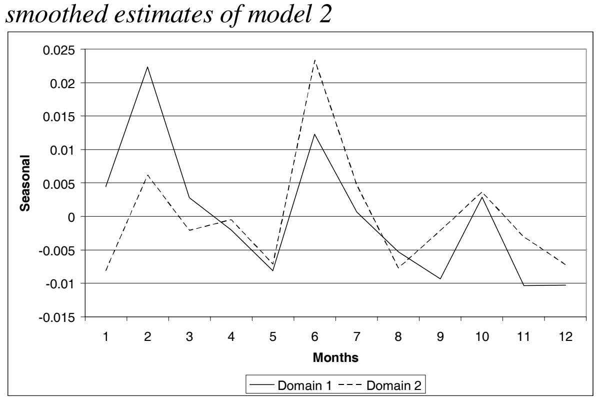
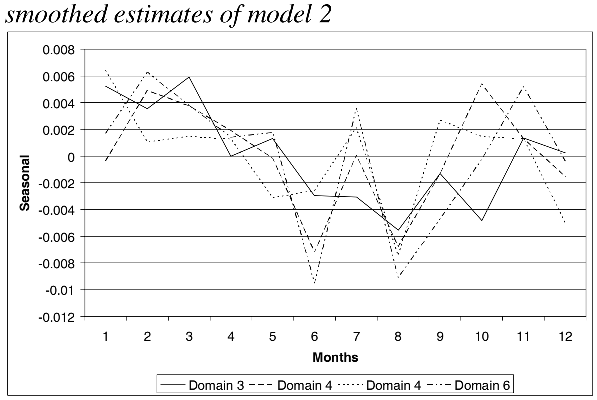


Figure 5: Seasonal effects of domain 3, 4, 5, and 6, smoothed estimates of model 2

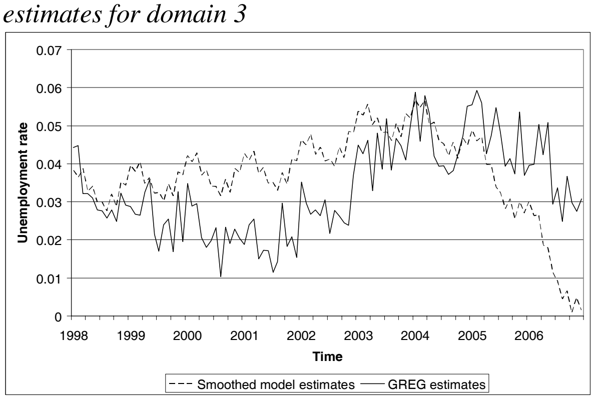


The difference between domain 3 and domain 4 for October is quite large. The model finds constant seasonal effects for domain 1, 2, and 3, i.e. $\eta_{S,t,d} = 0$. Though the seasonals change gradually over time for domain 4, 5 and 6, only the last year is presented in Figure 5 in order to make the figure readable.

Model 1 and 2 only borrow strength in time. Now models are considered that also borrow strength in space by allowing non-zero correlations between the slope parameters of the smooth trend models for the different domains.

Model 3 allows for separate correlations between the slope parameters of the six domains. This model does not produce reliable estimates for the monthly unemployment rates. Figure 6 shows that the smoothed estimates of the unemployment rate of e.g. domain 3 is very different from the GREG estimates. Extremely biased estimates are obtained with this model. Similar results are obtained for the other domains. The numerical estimation procedure fails to find a valid maximum likelihood estimate for the covariance matrix of the slopes, since this covariance matrix is not positive semidefinite. As a result the estimated correlations vary too much and are conflicting with each other. We do not further elaborate on model 3 and also exclude this model from the model evaluation in section 5.

Figure 6: Smoothed estimates of model 3 and GREG estimates for domain 3



To avoid these problems, simpler models are considered were the correlations between the slope parameters are restricted.

Model 4 assumes that the correlations between the slopes of all domains are equal. The correlation is estimated as $\varsigma_{R,d,d'} = 0.86$ for all *d* and *d'*. Figure 7 shows the influence of this correlation on the estimates of domain 6. The correlation between the slope parameters results in a small adjustment of the trend.

Model 5 allows for different correlations. The estimated correlations are bounded to take values between bounds b_1 and b_2, $0 < b_1 < b_2 < 1$, which are chosen in such a way that the estimated covariance matrix for the slope parameters is positive semidefinite. The lower and upper boundaries equal 0.805 and 0.915 and are derived through a grid-search. Figure 8 shows that the estimation results of model 4 and model 5 are very similar.

Model 6 allows for non-zero correlations between domain 1 and 2, between 3 and 4, and between 5 and 6. This implies that three bivariate models are assumed for the two gender classes within each age-class. Conflicting correlations are avoided because a two dimensional covariance matrix is always positive semidefinite. The correlations are estimated as $\varsigma_{R,1,2} = 0.65$, $\varsigma_{R,3,4} = 0.95$, $\varsigma_{R,5,6} = 0.79$.

The filtered estimates obtained under the models 4, 5 and 6 for domain 6 are compared in Figures 8 and 9.

Figure 7: Filtered estimates of model 2 and model 4 for domain 6

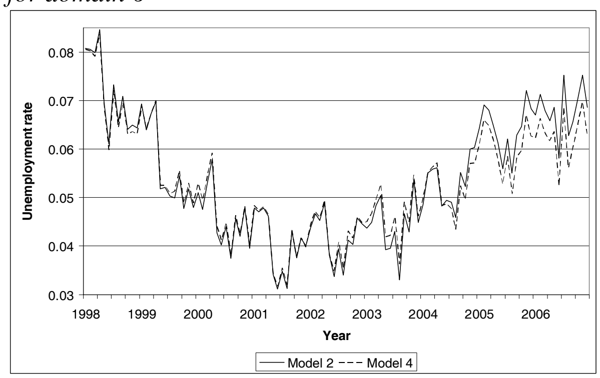


Figure 8: Filtered estimates of model 4 and model 5 for domain 6

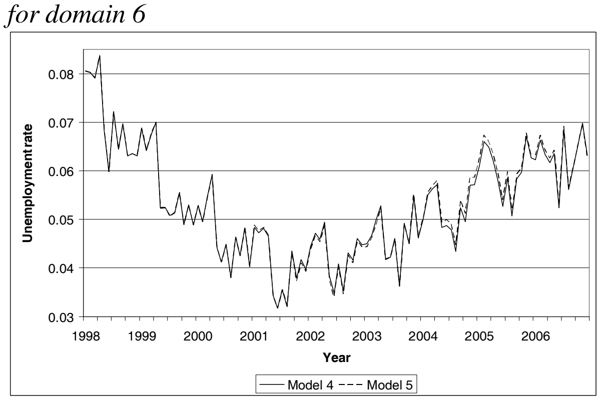
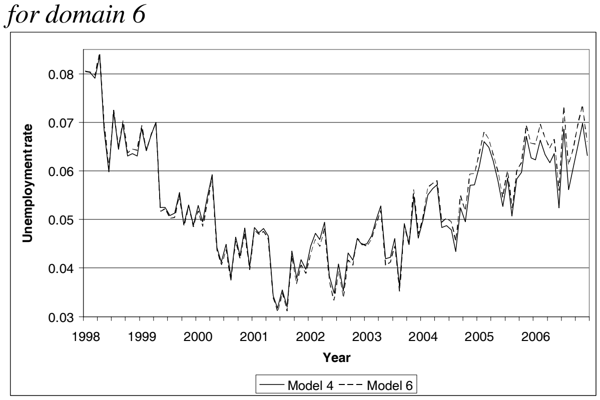


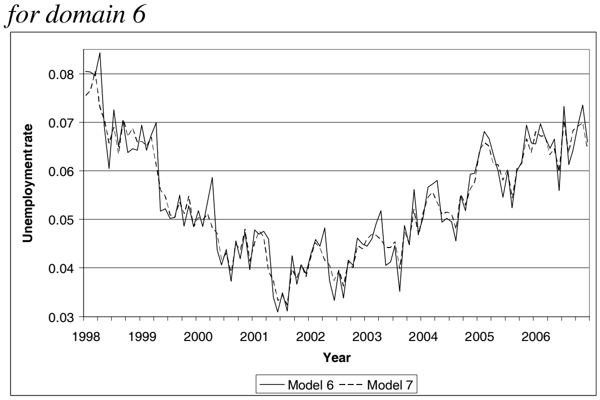
Figure 9: Filtered estimates of model 4 and model 6 for domain 6



The estimates obtained under model 4 and 5 are very similar. The estimates of the last years, obtained under model 6, are slightly larger than the estimates obtained under model 4. The estimates obtained with model 6 are very similar to the univariate model 2.

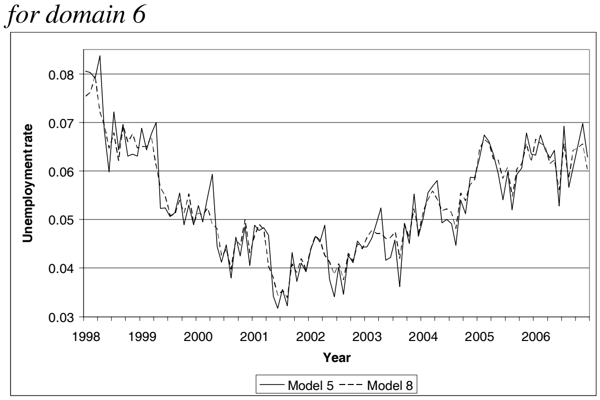
Model 7 allows for non-zero correlations between domain 1 and 2, between 3 and 4 and between 5 and 6, as defined in model 6. Based on the estimates for the seasonal effects obtained under model 2, it is also assumed that the seasonal effects for domain 1 and 2 and for domain 3, 4, 5 and 6 are equal, i.e. $S_{t,1} = S_{t,2}$ and $S_{t,3} = S_{t,4} = S_{t,5} = S_{t,6}$. Figure 10 compares the estimates of model 6 and model 7. Due to the assumption of equal seasonal effects, the estimates of these effects are more stable under model 7. Therefore, the series of model 7 follows a slightly smoother pattern. If the assumption of equal seasonal effects is not true, however, the estimates under model 7 are slightly biased.

Figure 10: Filtered estimates of model 6 and model 7 for domain 6



Model 8 allows for separate correlations between the slopes which are forced to lie between upper and lower boundaries in the same way as described in model 5. As in model 7, it is assumed that $S_{t,1} = S_{t,2}$ and $S_{t,3} = S_{t,4} = S_{t,5} = S_{t,6}$. Figure 11 compares the estimates of model 5 and model 8.

Figure 11: Filtered estimates of model 5 and model 8 for domain 6



The standard errors of the GREG estimates and the filtered estimates for the different models are compared in Table 1.

The first result is that the standard errors of the filtered estimates under model 1 and 2 are much smaller than the standard errors of the GREG estimates. This illustrates that borrowing strength from other time periods can increase the precision of the estimates substantially. The standard error does not reflect the bias due to model misspecification. Model 1 has the smallest standard errors, but as we will see in section 5.1, this model will result in severely biased estimates since it ignores seasonal effects.

The standard errors can be further reduced by using information from other domains (compare model 2 with models 4 through 8 in Table 1). The additional gain, however, is relatively small compared to the reduction in the standard errors that is obtained with the models that borrow strength from the past.

Model 6, where three bivariate models are assumed, has larger survey errors than model 4 and 5, where all domains are correlated. Models that allow for more flexible correlation patterns result in smaller standard errors. The reduction of the standard error under model 4 and 6 is smaller compared to the reduction obtained under model 5. The correlations under model 5 are bounded in a rather artificial way, to avoid the severely biased estimates obtained with model 3.

Borrowing information from other domains by assuming that the seasonal patterns are equal for different domains, yields a further reduction of the standard error (compare model 6 with model 7 and model 5 with model 8).

Table 1: Standard error of filtered estimates (x 1000), mean over 2006

		Model						
Domain	*GREG*	*1*	*2*	*4*	*5*	*6*	*7*	*8*
1	18.9	8.1	8.7	8.0	7.8	8.5	8.0	7.2
2	23.9	9.1	10.5	9.2	9.0	10.1	8.9	7.1
3	6.0	3.0	3.0	2.8	2.7	2.8	2.8	2.6
4	8.0	3.5	4.2	3.9	3.8	3.9	3.2	3.0
5	7.0	2.7	3.3	3.0	2.9	3.2	2.8	2.3
6	10.1	4.1	5.4	5.1	5.0	5.3	3.9	3.5

5. Model evaluation and model selection

In Brown e.a. (2001), model diagnostics for small area estimation models are proposed. These diagnostics are appropriate for linear mixed models. Some of the diagnostics can also be used for the multivariate time series models that are applied in this paper. They are used together with diagnostics which are specific for time series, see e.g. Durbin and Koopman (2001). All tests are done at a 5% significance level.

The model estimates for the first two years are excluded from the model evaluation because the model estimates from the first periods cannot be used due to the diffuse initialisation of the model.

5.1 Coverage

For the GREG-estimates, 95% confidence intervals can be computed. If the model estimates are similar to the true population value, the interval should not cover the model estimates in around 5% of the cases. A non-coverage rate of (much) more than 5% indicates that the model estimates are not very similar to the true population values, e.g. because they are biased. If (much) less than 5% of the model estimates are outside the confidence interval, then they are actually too close to the GREG estimates.
The smoothed estimates are used to calculate the coverage rates, because they are the most accurate estimates under the time series model. Table 2 shows the non-coverage rate for all domains and all models.

Table 2: Monthly non-coverage rate (in %) for smoothed estimates

	Model						
Domain	*1*	*2*	*4*	*5*	*6*	*7*	*8*
1	4.6	1.9	2.8	2.8	2.8	3.7	4.6
2	8.3	6.5	7.4	6.5	7.4	6.5	5.6
3	3.7	1.9	0.9	0.9	0.9	1.9	2.8
4	5.6	4.6	4.6	4.6	4.6	5.6	5.6
5	8.3	3.7	3.7	3.7	3.7	5.6	5.6
6	6.5	5.6	4.6	4.6	4.6	4.6	4.6
Total	6.2	4.0	4.0	3.9	4.0	4.6	4.8

Total refers to the rates averaged over the six domains.

Model 1 is the only one where the non-coverage rate of the total is larger than 5%. In the second domain, the non-coverage is relatively large for some other models.

In a similar way the coverage rates of linear combinations of the monthly unemployment rates can be calculated. For example the mean of the 12 months for each calendar year to evaluate whether the model estimates are biased in these periods. It is also interesting to compute the coverage rates for the mean for each month over the different years (all Januaries, Februaries etc) to check whether the seasonal patterns are modelled adequately. The rates are calculated for the separate domains. In case there are two or more model estimates outside the confidence interval, this is considered as suspicious, since there are only nine years and twelve months available.
For all domains and all models, the yearly coverage for the smoothed estimates is 100%. Table 3 shows the seasonal non-coverage rate. As expected, the seasonal effects are not described well in the first model. Furthermore, the confidence interval of October does not cover the model estimate of model 7 and 8 for domain 3. Figure 5 shows that the difference between domain 3 and the other domains is quite large for this month.

Table 3: Seasonal non-coverage rate (in %) for smoothed estimates

	Model						
Domain	*1*	*2*	*4*	*5*	*6*	*7*	*8*
1	17	0	0	0	0	0	0
2	8	0	0	0	0	0	0
3	42	0	0	0	0	8	8
4	17	0	0	0	0	0	0
5	8	0	0	0	0	0	0
6	17	0	0	0	0	0	0

5.2 Bias

Another way to look at possible bias is to plot the direct estimates against the smoothed model estimates $\hat{\theta}_{t,d}$. If the model estimates are unbiased, the regression line should be close to the line $Y_{t,d} = \hat{\theta}_{t,d}$ (see Brown e.a.). Therefore, β_0 and β_1 for the regression $y_{t,d} = \beta_0 + \beta_1\hat{\theta}_{t,d}$ are estimated together with their standard errors. For all models and all domains, the OLS estimates for β_0 and β_1 are not significantly different from 0 and 1 respectively. Therefore, there is no reason to reject one of the models based on this evaluation test. Because the results of paragraph 5.1, in particular Table 3, show that model 1 is biased, it seems that this test is not very distinctive for bias in seasonal patterns.

5.3 Tests for normally distributed prediction errors

The one-step forecast errors or prediction errors e_t are defined as the difference between the one-step forecasts and the GREG estimates. These errors should be normally distributed. To check this, it is tested whether the skewness S and kurtosis K fit the normality assumption, see e.g. Durbin and Koopman (2001). The skewness is not significantly different from 0 for all domains and all models, except for domain 6 under model 8, where it is 0.504.

The kurtosis is not significantly different from 3 for all domains and all models except domain 4 under model 1, where it is 3.95, and domain 6 under model 4, 5, and 8, where it is 4.11, 4.03 and 4.65, respectively.

5.4 Heteroscedasticity

Another model assumption is that the variance of the prediction errors is constant in time. A test statistic for heteroscedasticity is given by

$$H = \sum_{t=79}^{132} e_t^2 \Big/ \sum_{t=25}^{78} e_t^2 \ .$$

Under the null hypothesis of homoscedasticity, H is $F_{54,54}$-distributed (Durbin and Koopman, 2001). The test detects heteroscedasticity in the prediction errors for domain 1 and domain 5 under all models, domain 3 under model 7 and 8, and for domain 6 under model 8.

5.5 Test for independent prediction errors

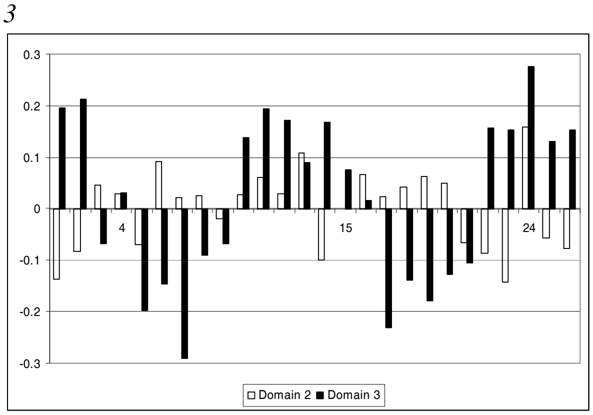
To check whether the prediction errors are serially independent, the sample autocorrelation functions are computed for time lags h=1,2,…,26. Since the sample autocorrelation functions of white noise are normally distributed with expectation 0 and variance $1/n$, no more than 5% of the sample autocorrelations of the prediction errors should be outside the bounds ± 0.189 . The number of sample autocorrelation functions outside the bounds is given in table 4.

Table 4: Number of sample autocorrelation functions outside the 95% confidence interval

	Model						
Domain	1	2	4	5	6	7	8
1	6	2	1	1	2	3	3
2	0	0	0	0	0	1	1
3	7	0	0	0	0	0	0
4	1	3	2	2	2	0	2
5	0	1	0	0	1	1	0
6	1	0	0	0	0	0	0

Since model 1 does not contain a seasonal component, the typical pattern for seasonal autocorrelation should be found in the correlograms. Surprisingly, the number of autocorrelations that exceed the bounds of the 95% confidence interval are smaller than 5% for four out of the six domains under this model. The correlogram shows a seasonal pattern only for domain 1, 3 and 6. The correlograms for domain 2 and 3 are shown in Figure 12.

Figure 12: Correlogram for model 1 and domain 2 and 3



For some of the other models, the autocorrelation is suspicious for domain 1. Figure 13 shows the correlograms for domain 1 under model 5 and 8. It can be seen that there is a cyclical pattern and a large autocorrelation for h=1. This pattern is similar for all models, except model 1. The cyclical pattern is more pronounced in model 8, here more autocorrelations are outside the confidence interval.

Figure 14 shows the correlograms for domain 4 under model 7 and 8. The pattern is similar for all models except model 1.

Figure 13: Correlogram for model 5 and model 8 and domain 1

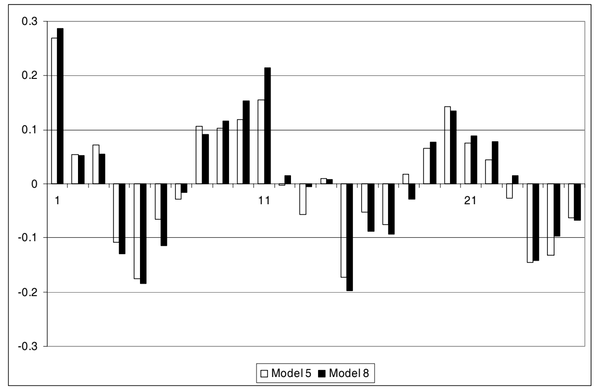
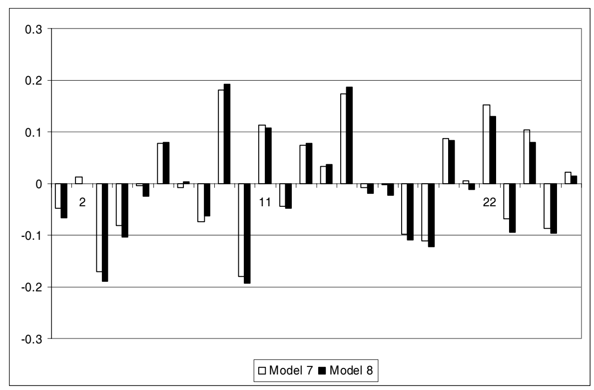


Figure 14: Correlogram for model 7 and model 8 and domain 4



5.6 Mean absolute prediction error

The mean absolute prediction error is computed as

$$APE = \frac{1}{T-24}\sum_{t=25}^{T}\left|e_t\right| .$$

A small value for APE implies that the model predicts the true population value well. Table 5 shows the APE for all models and all domains.

Table 5: The APE (x 100) for all models

	Model						
Domain	*1*	*2*	*4*	*5*	*6*	*7*	*8*
1	1.72	1.72	1.69	1.67	1.74	1.68	1.58
2	1.96	2.04	2.04	2.03	2.03	2.00	1.97
3	0.59	0.52	0.51	0.51	0.52	0.53	0.51
4	0.69	0.73	0.70	0.69	0.70	0.65	0.66
5	0.56	0.57	0.54	0.54	0.57	0.55	0.53
6	0.90	0.97	0.92	0.93	0.95	0.85	0.84

Surprisingly, the APE under model 1 is smaller than under model 2 for all domains except domain 1 and 3. These are the domains where the sample autocorrelation functions are suspicious (see Table 4).

Borrowing strength from other domains generally reduces the APE, compare e.g. model 4 and 5 with model 2. For model 6 however, where strength is only borrowed by bivariate models, the APE are similar as for model 2. When some seasonal effects are chosen equal, the APE are reduced for most domains (compare model 6 with model 7 and model 5 with model 8).

6. Discussion and conclusions

In this paper different multivariate structural time series models are applied to the monthly unemployment rate of the Dutch LFS. With these models, the precision of the estimates of six domains can be improved substantially compared with the GREG estimates. The most important improvement is achieved by borrowing information from other time periods. Further, but substantially smaller, improvements are possible by borrowing information from other domains by modelling the correlation between the parameters of the time series models for the domains or by assuming that these parameters are equal for different domains.

Seven different models are evaluated using different diagnostic tests. The smallest values of the prediction error and the standard error are obtained with model 8. There are, however, numerical problems with this model which are avoided by restricting the correlations of the slope parameters in an artificial way. The model evaluation does not indicate that the estimates under this model are biased, but detects that the prediction errors are heteroscedastic, deviate from normality, and are serially correlated. Model 7 might be an alternative, since this model results in slightly larger standard errors and better evaluation results.

For all considered models, however, heteroscedasticity and autocorrelation for a part of the domains is detected. The models could be improved, for example by using trigonometric functions for the seasonal components (see Durbin and Koopman, 2001) and modelling of outliers. Also auxiliary information can be used to improve the models, e.g. using information about the registered unemployment. Finally a larger set of diagnostic tests than considered in this paper is required for adequate model evaluation. In the continuation of this project we will investigate which diagnostic tests are most relevant for model evaluation and selection.

Acknowledgement

The authors wish to thank professor Pfeffermann for his advice during this research. The views expressed in this paper are those of the authors and do not necessarily reflect the policies of Statistics Netherlands.

References

Brakel, J.A. van den and S. Krieg (2007). Modelling Rotation Group Bias and Survey Errors in the Dutch Labour Force Survey. *Proceedings of the section on Survey Research Methods*, American Statistical Association.

Brown, G., R. Chambers, P. Heady and D. Heasman (2001). Evaluation of Small Area Estimation Methods – an Application to Unemployment estimates from the UK LFS. Proceedings of Statistics Canada Symposium 2001.

Doornik, J.A. (1998). *Object-Oriented Matrix Programming using Ox 2.0*. London: Timberlake Consultants Press.

Durbin, J. and S.J. Koopman (2001). *Time series analysis by state space methods*. Oxford: Oxford University Press.

Koopman, S.J., N. Shephard and J.A. Doornik (1999). Statistical Algorithms for Models in State Space using SsfPack 2.2. *Econometrics Journal*, 2, pp. 113-166.

Pfeffermann, D. and S.R. Bleuer (1993). Robust Joint Modelling of Labour Force Series of Small Areas. *Survey Methodology*, vol. 19, pp. 149-163.

Pfeffermann, D. and L. Burck (1990). Robust Small Area Estimation Combining Time Series and Cross-Sectional Data. *Survey Methodology*, vol. 16, pp. 217-237.

Rao, J.N.K. and M. Yu (1994), Small Area Estimation by Combining Time Series and Cross-Sectional Data, *Canadian Journal of Statistics*, vol. 22, pp. 511-528.

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

Scott, A.J. and T.M.F. Smith (1974). Analysis of Repeated Surveys using Time Series Methods. *Journal of the American Statistical Association*, 69, pp. 674-678.