Discussion Paper

Estimation of level and change for unemployment using multilevel and structural time-series models

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

2016 | 10

Harm Jan Boonstra
Jan van den Brakel

Monthly estimates of provincial unemployment based on the Dutch Labour Force Survey (LFS) are obtained using time-series models. The LFS uses a rotating panel design so that in each month the total sample consists of five independent waves. The time-series model accounts for rotation group bias and serial correlation due to the rotating panel design. We consider two approaches to time-series modelling: structural time-series modelling and multilevel modelling. The structural time-series models are fit using a Kalman filter and smoother whereas the multilevel models are fit using a Gibbs sampler approach. Monthly unemployment estimates and standard errors based on these models are compared for the twelve provinces of the Netherlands. Pros and cons of the multilevel approach and structural time-series approach are discussed. Time-series models are appropriate to borrow strength over time and space. Modelling the full correlation matrix between time-series components rapidly increases the numbers of hyperparameters to be estimated. Modelling common factors is one possibility to obtain more parsimonious models that still account for cross-sectional correlation. In this paper an even more parsimonious approach is proposed, where domains share one overall trend and a parameter for the domain specific deviation from this overall trend. By borrowing strength over time as well as over areas, the time-series models more than halve the standard errors for the monthly provincial unemployment estimates, as compared to the design-based variances of the direct estimates. The reduction of the standard errors with equivalent models under a state space approach is larger, since the uncertainty of using maximum likelihood estimates for the hyperparameters is ignored. The time-series modelling approach is particularly appropriate to estimate month-to-month and year-to-year change of unemployment.

This work was funded by the European Union under grant no 07131.2015.001-2015.257.

1 Introduction

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

Until 2015 municipal estimates were produced annually by means of direct generalized regression estimation (GREG, see e.g. Särndal et al. (1992)), but only for municipalities with at least 30 thousand inhabitants. For municipalities with 10 to 30 thousand inhabitants, three-year moving averages of GREG estimates were used in order to reduce the variance. No estimates were published for municipalities with fewer than 10 thousand inhabitants. To improve the municipal estimates, a model-based small area estimation (SAE) strategy has been adopted starting 2015 (Boonstra et al., 2011; Boonstra and Michiels, 2013). The model used for this purpose is the Battese-Harter-Fuller basic unit-level model (Battese et al., 1988; Rao, 2003), a linear multilevel model with random municipality effects. Based on this model, estimates of labour status are produced annually for all municipalities.

The continuous nature of the LFS allows to borrow strength not only from other areas, but also over time. A structural time-series model (STM) is already being used to estimate national monthly labour status for 6 gender by age classes (van den Brakel and Krieg, 2009, 2015). In this paper we aim to combine the borrowing of strength over time of the time-series model with borrowing of strength over space of cross-sectional small area models to produce reliable monthly estimates of provincial unemployment.

Until now, provincial estimates are produced quarterly using the GREG. In order to produce figures on a monthly basis, a model-based estimation strategy is necessary to overcome the problem of too small monthly provincial sample sizes. The models considered are applied at an aggregate level. First, initial estimates are computed for each province in each month and for each panel wave using the survey regression estimator that uses auxiliary information to reduce non-response bias. The initial estimates and their estimated variances are subsequently modeled in a multilevel or structural time-series model. As a consequence of the LFS panel design, the initial estimates are autocorrelated and estimates based on follow-up waves are biased relative to the first wave estimates. Both features need to be accounted for in the model.

Previous accounts of regional small area estimation of unemployment, where strength is borrowed over both time and space, include Rao and Yu (1994); Datta et al. (1999); You et al. (2003); You (2008); Pfeffermann and Burck (1990); Pfeffermann and Tiller (2006), see also Rao (2003) for an overview. In Boonstra (2014) several multilevel time-series models have been applied to the estimation of annual unemployment levels for Dutch municipalities.

The time-series multilevel model is fit using a Gibbs sampler. Models with different combinations of fixed and random effects are compared based on the Deviance Information Criterion (DIC). For the STM a frequentist approach is followed, where hyperprarameters are estimated with maximum likelihood. The estimates based on multilevel and STM models and their standard errors are compared graphically and contrasted with the initial survey regression estimates. This paper further elaborates on the time-series modelling approach to obtain stable and precise estimates for level, month-to-month change and year-to-year change for

unemployment. Connections between structural time-series models and multilevel models have been explored before from several points of view in Knorr-Held and Rue (2002), Chan and Jeliazkov (2009), McCausland et al. (2011), Ruiz-Cárdenas et al. (2012), Piepho and Ogutu (2014). A comparison between multilevel time-series models and state-space models applied to time series of the Dutch National Travel Survey is carried out in Bollineni-Balabay et al. (2016).

This report is structured as follows. In Section 2 the LFS data used in this study are described. Section 3 describes how the initial estimates are computed. Section 4 discusses the multilevel and STM models used to model the initial estimates. In Section 5 the results based on several multilevel and STM models are compared, including estimates for month-to-month change and year-to-year change for monthly data. Section 6 contains a discussion of the results as well as some ideas on further work.

2 Data from the Dutch Labour Force Survey

The Dutch LFS is a household survey conducted according to a rotating panel design in which the respondents are interviewed five times at quarterly intervals. In the years considered in this study, the first wave of the panel consists of data collected by means of computer assisted personal interviewing (CAPI), whereas the four follow-up waves contain data collected by means of computer assisted telephone interviewing (CATI). For a more detailed description of the sampling design, we refer to Boonstra et al. (2008). In more recent years the LFS has undergone several changes, the most significant being the introduction of internet as an additional mode of observation in the first wave. These changes have led to discontinuities that can be accounted for by adding intervention effects to the time-series model (van den Brakel and Krieg, 2015). The treatment of these discontinuities in the case of monthly provincial figures is postponed to a future paper.

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

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

In the present study we use 72 months of LFS data from 2003 to 2008. Data from all five waves of the rotating panel are used. The Netherlands is divided into twelve provinces which serve as the domains for which monthly unemployment figures are to be estimated.

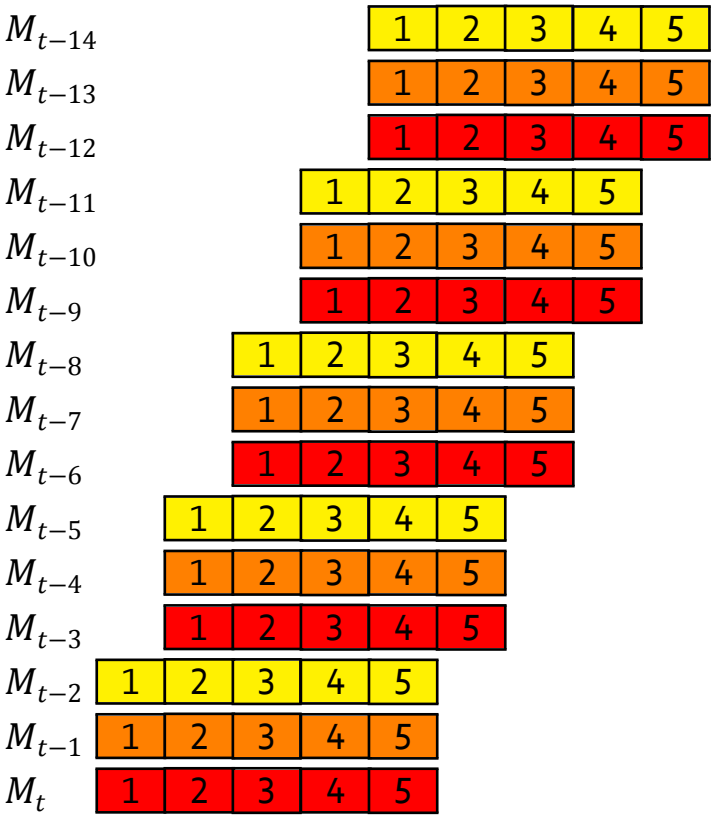

Figure 1 LFS rotating panel design for months $t-14$,...,t. Waves are coded with their corresponding wave number. The first wave is conducted through CAPI whereas the second to fifth waves are conducted through CATI. Vertically aligned squares with similar colours represent observations from the same households (barring panel attrition) in different months. Horizontally aligned squares combine into a monthly dataset.

Monthly national sample sizes vary between 5 and 7 thousand persons in the first wave and between 3 and 5 thousand in the fifth wave. Provincial sample sizes are diverse, ranging from 31 to 1949 persons for single wave monthly samples.

LFS data are available at the level of units, i.e. persons. A wealth of auxiliary data from several registrations is also available at the unit level. Among these auxiliary variables is registered unemployment, a strong predictor for the unemployment variable of interest. These predictors are used to compute initial estimates, which are input to the time-series models.

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

3 Initial estimates

Let $\hat{\bar{Y}}_{itp}$ denote the initial estimate for area i and period t based on data from wave p. The initial estimates used as input for the time-series small area models are survey regression estimates (Woodruff, 1966; Battese et al., 1988; Särndal et al., 1992)

$$\hat{\bar{Y}}_{itp} = \bar{y}_{itp} + \hat{\beta}'_{tp}(\bar{X}_{it} - \bar{x}_{itp}) , \tag{1}$$

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

Even though the regression coefficients used in (1) are not area-specific, the survey regression estimate for a particular area is a direct estimate in the sense that it is primarily based on the data obtained in that area, i.e. on $\bar{y}_{itp}$. The second term in (1) is a correction term that reduces bias due to selection effects, i.e. differences between the observed data and population with regard to the distribution of auxiliary variables x used in the model. The available auxiliary variables are listed in Table 1. The model selected to compute the survey regression estimates is

$$ru \times (sex + age3 + ethn) + ru5 + sex \times age5 + hhtype . \tag{2}$$

Note the prominent role played by registered unemployment. Despite being based on a very different concept of unemployment, it is still a strong predictor for the unemployment variable of interest.

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

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

Table 1 Available covariates

supported by the fact that panel attrition is limited, with response fractions close to 90% in waves two to five, given initial response in the first wave. Fig. 3 displays the survey regression estimates averaged only over areas, for wave 1 and the average of waves 2 to 5. The figure suggests that rotation group bias varies over time. Finally, Fig. 4 shows the rotation group bias per province, averaged over time. The two provinces with highest unemployment, Groningen and Flevoland, also show a larger rotation group bias.

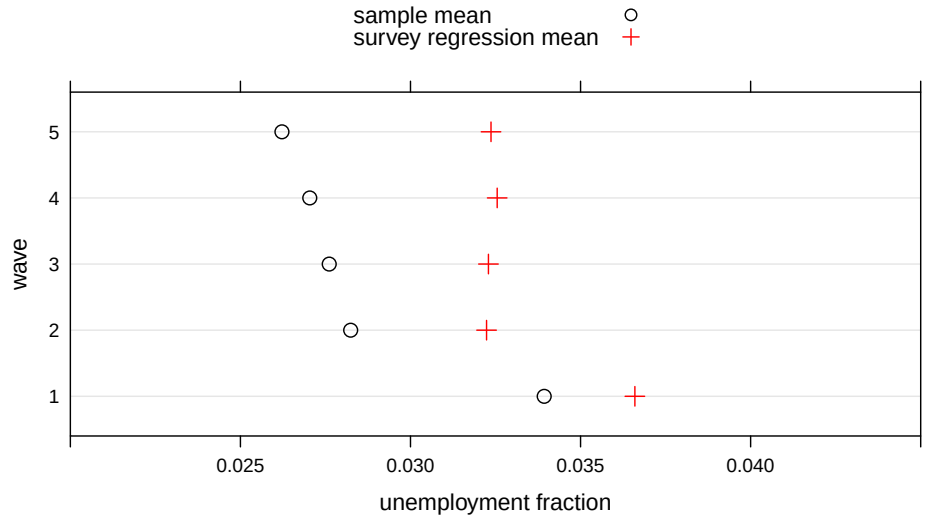

Figure 2 Systematic differences between estimates based on different waves (rotation group biases): sample means vs. survey regression means aggregated over periods and areas.

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

$$v(\hat{\bar{Y}}_{itp}) = \widehat{\sigma^2_{tp}}/n_{itp}\,, \tag{3}$$

based on the estimated within-area variances $\widehat{\sigma^2_{tp}}$ of the regression residuals pooled over areas, and the area-specific sample sizes n_{itp}.

The panel design induces several non-zero correlations among initial estimates for the same province and different time periods and waves. These positive correlations are due to partial overlap of the sets of sample units on which the estimates are based. Such correlations exist between estimates for the same province in months t_1, t_2 and based on waves p_1, p_2 whenever $t_2 - t_1 = 3(p_2 - p_1) \leq 12$, see Fig. 1. The covariances between $\hat{\bar{Y}}_{it_1p_1}$ and $\hat{\bar{Y}}_{it_2p_2}$ are estimated

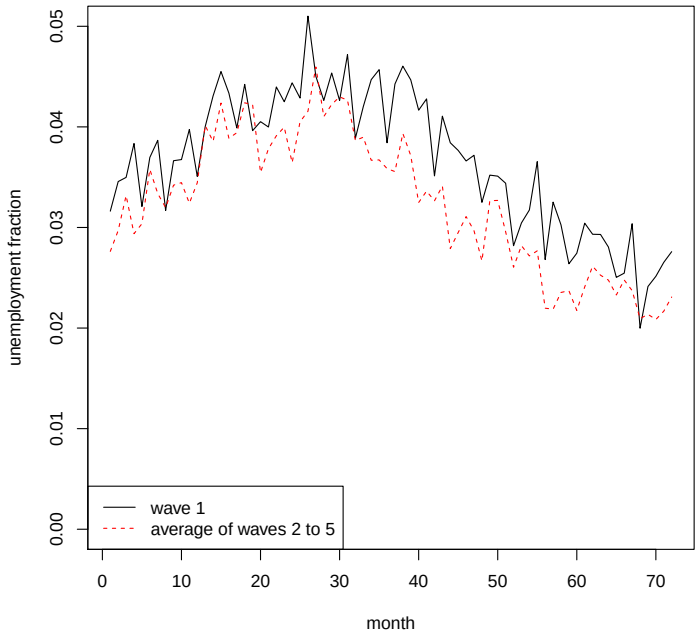

Figure 3 Systematic differences between survey regression estimates based on different waves over time. The estimates are aggregated over areas.

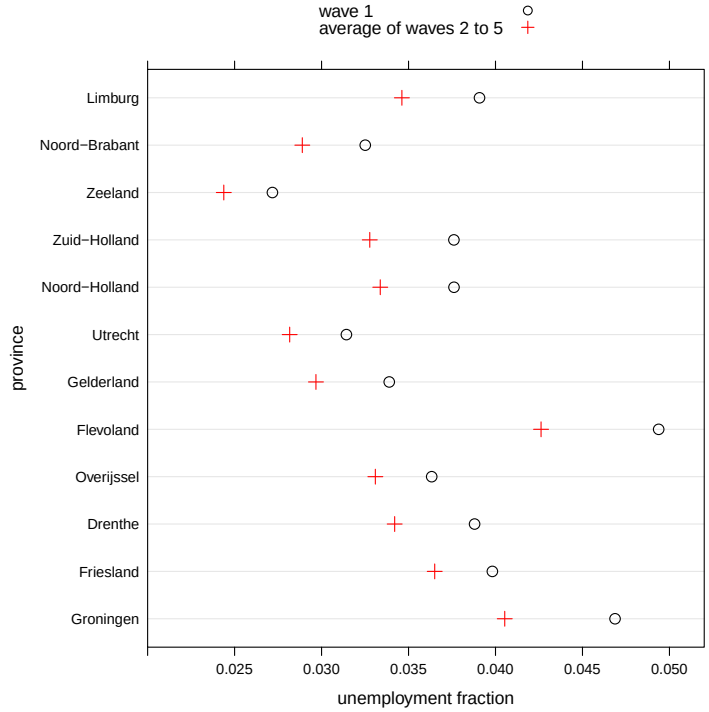

Figure 4 Systematic differences between survey regression estimates based on different waves by province. The estimates are aggregated over time.

as (see e.g. Kish (1965))

$$v(\hat{\bar{Y}}_{it_1p_1}, \hat{\bar{Y}}_{it_2p_2}) = \frac{n_{it_1p_1t_2p_2}}{\sqrt{n_{it_1p_1}n_{it_2p_2}}}\hat{\rho}_{t_1p_1t_2p_2}\sqrt{v(\hat{\bar{Y}}_{it_1p_1})v(\hat{\bar{Y}}_{it_2p_2})}, \tag{4}$$

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

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

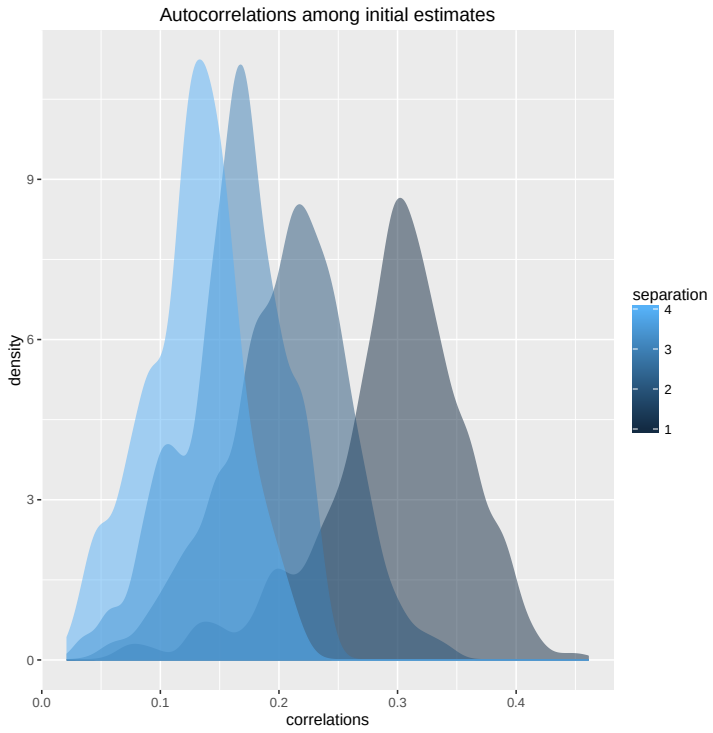

Figure 5 Distribution of non-zero autocorrelations between initial estimates, grouped by the separation in time (quarters).

Time-series model estimates for monthly provincial unemployment figures will be compared with direct estimates. The procedure for calculating monthly direct estimates is based on the approach that was used before 2010 to calculate official rolling quarterly figures for the labour force. Monthly direct estimates for provinces are calculated as the weighted mean over the five panel survey regression estimates (1):

$$\hat{\bar{Y}}_{it.} = \sum_{p=1}^{5} \alpha_{itp}\hat{\bar{Y}}_{itp}, \tag{5}$$

where α_{itp} are weights based on the variance estimates (3) and are obtained by:

$$\alpha_{itp} = \frac{v(\hat{\bar{Y}}_{itp})^{-1}}{\sum_{p=1}^{5} v(\hat{\bar{Y}}_{itp})^{-1}}. \tag{6}$$

To correct for RGB, the direct estimates (5) are multiplied by a ratio, say f_{it}, where the numerator is the mean of the survey regression estimates (1) for the first wave over the last three years and the denominator is the mean of monthly direct estimates (5) also over the last three years, i.e.

$$f_{it} = \frac{\sum_{j=0}^{35} \hat{\bar{Y}}_{i(t-j)1}}{\sum_{j=0}^{35} \hat{\bar{Y}}_{i(t-j)\cdot}}, \tag{7}$$

With this correction the direct estimates (5) are benchmarked to the level of the estimates in the first wave. The observed series start in January 2003. During the first three years of the series the correction factors f_{it} with t=January 2006 are used. Finally the corrected monthly survey regression estimate for month t is equal to

$$\tilde{\bar{Y}}_{it\cdot} = f_{it}\hat{\bar{Y}}_{it\cdot}, \tag{8}$$

with variance

$$v(\tilde{\bar{Y}}_{it\cdot}) = f_{it}^2 \sum_{p=1}^{5} \alpha_{itp}^2 v(\hat{\bar{Y}}_{itp}). \tag{9}$$

4 Time-series small area estimation

To fit a model to all unit-level data of all periods at once would be very challenging, not only because of the size of the combined dataset, but also because such a model would become quite complex since it must account for many effects at once. In addition to area and time effects the model would also need to include person or household effects to account for repeated measurements of the same persons/households. We therefore specify the time-series models at the area, i.e. provincial, level for the separate waves. The data for the time-series model then consist of a set of initial estimates at the area level, computed separately for each time period and for each wave, as described in Section 3.

Modeling and estimation are then divided into two stages. In the first stage initial estimates are computed using auxiliary information at the unit level to reduce non-response bias, as described in Section 3. The initial estimates are accompanied by variance estimates as well as estimates of covariances between estimates induced by the rotating panel design. In the second stage a structural or multilevel time-series model is applied to smooth the initial estimates, reduce standard errors and correct for rotation group bias. The estimated models are used to make predictions for provincial unemployment fractions, provinical unemployment trends, and month-to-month changes in the trends.

4.1 Structural time-series model

The first approach considered in this paper is a multivariate structural time-series model where the series of direct estimates for the separate waves and provinces are the input series. This approach is used by Statistics Netherlands to produce monthly figures about the Dutch labour force at national level and a break down in six domains that is based on the cross classification of gender and age (van den Brakel and Krieg, 2009, 2015). In this section we develop a structural time-series model for the monthly data at provincial level. Subsequently we model twelve provinces simultaneously to take advantage of temporal and cross-sectional sample information.

Let $\hat{\bar{Y}}_{it} = (\hat{\bar{Y}}_{it1}, \ldots, \hat{\bar{Y}}_{it5})^t$ denote the five-dimensional vector containing the survey regression estimates $\hat{\bar{Y}}_{itp}$ defined by (1) in period t and domain i. This vector can be modeled with the folowing structural time-series model (Pfeffermann, 1991; van den Brakel and Krieg, 2009):

$$\hat{\bar{Y}}_{it} = \iota_5 \theta_{it} + \lambda_{it} + e_{it}, \tag{10}$$

where ι_5 denotes a five-dimensional vector with each element equal to 1, θ_{it} the true population parameter for period t in domain i, λ_{it} a five-dimensional vector that models the rotation group bias and e_{it} a five-dimensional vector with sampling errors. The population parameter θ_{it} in (10) is modeled as

$$\theta_{it} = L_{it} + S_{it} + \epsilon_{it}, \tag{11}$$

where L_{it} denotes a stochastic trend model to capture low frequency variation (trend plus business cycle), S_{it} a stochastic seasonal component to model monthly fluctuations and ϵ_{it} a white noise for the unexplained variation in θ_{it}. For the stochastic trend component, the so-called smooth trend model is used, which is defined by the following set of equations:

$$\begin{aligned} L_{it} &= L_{it-1} + R_{it-1}, \\ R_{it} &= R_{it-1} + \eta_{R,it}, \\ \eta_{R,it} &\overset{ind}{\sim} N(0, \sigma^2_{Ri}). \end{aligned} \tag{12}$$

For the stochastic seasonal component the trigonometric form is used, which has the following form for a monthly pattern:

$$S_{it} = \sum_{l=1}^{6} S_{itl}, \tag{13}$$

where

$$\begin{aligned} S_{itl} &= \cos(\frac{\pi l}{6}) S_{it-1;l} + \sin(\frac{\pi l}{6}) S^*_{it-1;l} + \omega_{itl}, \qquad \omega_{itl} \overset{ind}{\sim} N(0, \sigma^2_{\omega_i}), \\ S^*_{itl} &= -\sin(\frac{\pi l}{6}) S_{it-1;l} + \cos(\frac{\pi l}{6}) S^*_{it-1;l} + \omega^*_{itl}, \quad \omega^*_{itl} \overset{ind}{\sim} N(0, \sigma^2_{\omega_i}), \quad l = 1, \ldots, 5, \\ S_{it6} &= -S_{it-1;6} + \omega_{it6}, \qquad\qquad \omega_{it6} \overset{ind}{\sim} N(0, \sigma^2_{\omega_i}). \end{aligned}$$

The white noise in (11) is defined as $\epsilon_{it} \overset{ind}{\sim} N(0, \sigma^2_{\epsilon_i})$.

Systematic differences between the series of the survey regression estimates observed in the five waves, i.e. the RGB, are modeled in (10) with $\lambda_{it} = (\lambda_{it1}, \lambda_{it2}, \lambda_{it3}, \lambda_{it4}, \lambda_{it5})^t$. To identify the model, it is assumed that the $\lambda_{it1} = 0$. This implies that it is assumed that the survey regression estimates of the first wave are the most reliable approximations for θ_{it}, see van den Brakel and Krieg (2009) for a motivation. The remaining components model the systematic difference between wave p with respect to the first wave and are modeled as random walks. As a result,

$$\begin{aligned} \lambda_{it1} &= 0, \\ \lambda_{itp} &= \lambda_{it-1;p} + \eta_{\lambda,itp}, \ \eta_{\lambda,itp} \overset{ind}{\sim} N(0, \sigma^2_{\lambda_i}), \ p = 2,3,4,5. \end{aligned} \tag{14}$$

The use of random walks allows for time dependent patterns in the RGB.

Finally, a time-series model for the survey errors is developed. Let $e_{it} = (e_{it1}, e_{it2}, e_{it3}, e_{it4}, e_{it5})^t$ denote the five-dimensional vector containing the survey errors of the five waves. The direct estimates for the variance of the survey regression estimates are used as prior information in the time-series model to account for heteroscedasticity due to varying sample sizes over time using the following survey error model:

$$e_{itp} = \sqrt{v(\hat{\bar{Y}}_{itp})}\tilde{e}_{itp}, \tag{15}$$

and $v(\hat{\bar{Y}}_{itp})$ defined by (3). The autocorrelation between survey errors due to panel overlap is modeled with an AR(q) model, where $q = p - 1$ denotes the order of the AR model and p the number of the wave. This results in the following model for the survey errors:

$$\begin{aligned} \tilde{e}_{it1} &= \nu_{it1}, \ \nu_{it1} \overset{ind}{\sim} N(0, \sigma^2_{\nu_{i1}}), \\ \tilde{e}_{it2} &= \varrho_{it12}\tilde{e}_{it-3;1} + \nu_{it2}, \ \nu_{it2} \overset{ind}{\sim} N(0, \sigma^2_{\nu_{i2}}), \\ \tilde{e}_{it3} &= \varrho_{it23}\tilde{e}_{it-3;2} + \varrho_{it13}\tilde{e}_{it-6;1} + \nu_{it3}, \ \nu_{it3} \overset{ind}{\sim} N(0, \sigma^2_{\nu_{i3}}), \\ \tilde{e}_{it4} &= \varrho_{it34}\tilde{e}_{it-3;3} + \varrho_{it24}\tilde{e}_{it-6;2} + \varrho_{it14}\tilde{e}_{it-9;1} + \nu_{it4}, \ \nu_{it4} \overset{ind}{\sim} N(0, \sigma^2_{\nu_{i4}}), \\ \tilde{e}_{it5} &= \varrho_{it45}\tilde{e}_{it-3;4} + \varrho_{it35}\tilde{e}_{it-6;3} + \varrho_{it25}\tilde{e}_{it-9;2} + \varrho_{it15}\tilde{e}_{it-12;1} + \nu_{it5}, \ \nu_{it5} \overset{ind}{\sim} N(0, \sigma^2_{\nu_{i5}}), \end{aligned} \tag{16}$$

with $\varrho_{itpp'}$ the partial autocorrelation coefficient between waves p and p' of domain i in period t. These partial autocorrelation coefficients are derived from the correlation coefficients

$$\frac{n_{it_1p_1t_2p_2}}{\sqrt{n_{it_1p_1}n_{it_2p_2}}}\hat{\rho}_{t_1p_1t_2p_2} \tag{17}$$

obtained from the micro data as described in Section 3. The Yule-Walker equations are applied to the correlation coefficients (17) for each wave, domain and time period to obtain estimates for the partial correlation coefficients $\varrho_{itpp'}$, (Box et al., 2008). Finally an AR(1) model is chosen for wave 2 through 5 to model the autocorrelation in the survey errors. Strictly spoken this is an AR(3) model with the first partial autocorrelations for time periods $t - 1$ and $t - 2$ set equal to zero, because of the monthly frequency of the data and the quarterly correlation pattern of the observations. From (15) it follows for the first waves that $Var(\tilde{e}_{it1}) = \sigma^2_{\nu_{i1}}$ and $Var(e_{it1}) = v(\hat{\bar{Y}}_{it1})\sigma^2_{\nu_{i1}}$. For the preceding waves it holds that

$Var(\tilde{e}_{itp}) = \sigma^2_{\nu_{ip}}/(1-\varrho_{it(p-1)p})$ and $Var(e_{itp}) = v(\hat{\bar{Y}}_{itp})\sigma^2_{\nu_{ip}}/(1-\varrho_{it(p-1)p})$. The variances $\sigma^2_{\nu_{ip}}$ are scaling parameters with values close to one for the first wave and close to $1/T\sum_{t=1}^{T}(1-\varrho_{it(p-1)p})$ for the other waves, where T denotes the length of the observed series.

Model (10) uses sample information observed in preceding periods within each domain to improve the precision of the survey regression estimator and accounts for RGB and serial correlation induced by the rotating panel design. To take advantage of sample information across domains, model (10) for the separate domains can be combined in one multivariate model:

$$\begin{pmatrix} \hat{\bar{Y}}_{1t} \\ \vdots \\ \hat{\bar{Y}}_{m_A t} \end{pmatrix} = \begin{pmatrix} \iota_5\theta_{1t} \\ \vdots \\ \iota_5\theta_{m_A t} \end{pmatrix} + \begin{pmatrix} \lambda_{1t} \\ \vdots \\ \lambda_{m_A t} \end{pmatrix} + \begin{pmatrix} e_{1t} \\ \vdots \\ e_{m_A t} \end{pmatrix}, \tag{18}$$

where m_A denotes the number of domains, which is equal to twelve in this application. This multivariate setting, allows for several ways to use sample information across domains. In this paper models with cross-sectional correlation between the slope disturbance terms of the trend (12) are considered, i.e.

$$Cov(\eta_{R,it}, \eta_{R,i't'}) = \begin{cases} \sigma^2_{Ri} & if \quad i = i' \ and \ t = t' \\ \varsigma_{Rii'} & if \quad i \neq i' \ and \ t = t' \\ 0 & if \quad i \neq i' \ and \ t \neq t' \end{cases}. \tag{19}$$

Strong correlation between the slope disturbances across the domains can result in cointegrated trends. This implies that $q < m_A$ common trends are required to model the dynamics of the trends for the m_A domains and allows the specification of so-called common trend models, (Koopman et al., 1999; Krieg and van den Brakel, 2012).

Initial STM analyses showed that the seasonal and RGB component turned out to be time independent. It is therefore not sensible to model correlations between seasonal and RGB disturbance terms. Since the hyperparameters of the white noise population domain parameters tend to zero, it turned out to be better to remove this component completely from the model implying that modelling correlations between population noise is not considered. Correlation between survey errors for different domains is also not considered, since the domains are geographical regions from which samples are drawn independently.

As an alternative to a model with a full covariance matrix for the slope disturbances, a trend model is considered that has one common smooth trend model for all provinces plus $m_A - 1$ trend components that describe the deviation of each domain from this overall trend. In this case (11) is given by

$$\begin{aligned} \theta_{1t} &= L_t + S_{1t} + \epsilon_{1t}, \\ \theta_{it} &= L_t + L^*_{it} + S_{it} + \epsilon_{it}, \ \ i = 2, \ldots, m_A. \end{aligned} \tag{20}$$

In (20), L_t is the overal smooth trend component, defined by (12), and L_{it}^* the deviation from the overall trend for the separate domains, which are defined as smooth trend models defined by (12) or local levels, defined as

$$\begin{aligned} L_{it}^* &= L_{it-1}^* + \eta_{L,it}, \\ \eta_{L,it} &\overset{ind}{\sim} N(0, \sigma_{Li}^2). \end{aligned} \tag{21}$$

This trend model implicitly allows for correlations between the trends of the different domains.

Another more synthetic approach to use sample information across domains is to formulate seemingly unrelated structural time-series models, which implies that the disturbance terms of a factor share the same hyperparameter across the domains. This is equivalent to a multilevel model component with a single common variance parameter. The most synthetic approach to use sample infromation from other domains is to define the state variables equal over domains.

The parameters to be estimated with the time-series modelling approach are the trend and the signal. The latter is defined as the trend plus the seasonal component. The time-series approach is particularly suitable for estimating month-to-month changes. Seasonal patterns and sampling errors hamper a straightforward interpretation of month-to-month changes of direct estimates and smoothed signals. Therefore month-to-month changes are calculated for the trends only. Due to the strong positive correlation between the levels of consecutive periods, the standard errors of month-to-month changes in the level of the trends are much smaller than those of e.g. month-to-month changes of the direct estimates. The month-to-month change of the trend is defined as $\Delta_{it}(1) = L_{it} - L_{it-1}$ for models with separate trends for the domains or $\Delta_{it}(1) = L_t - L_{t-1} + L_{it}^* - L_{it-1}^*$ for models with an overall trend and $m_A - 1$ trends for the deviation from the overall trend for the separate domains. This modelling approach is also usefull to estimate year to year developments for trend and are defined as $\Delta_{it}(12) = L_{it} - L_{it-12}$ or $\Delta_{it}(12) = L_t - L_{t-12} + L_{it}^* - L_{it-12}^*$. Year to year differences are also sensible for signals, since the main part of the seasonal component cancels out. These developments are defined equivalently to the year to year developments of the trend.

The aforementioned structural time-series models are analyzed by putting them in the so-called state-space form. This implies that the STM is expressed in terms of a measurement equation and a transition equation. The measurement equation states how the observed times series depend on the underlying state variables (trend, seasonal, RGB and survey errors). The transition equation describes how the state variables gradually evolve over time. Under the assumption of normally distributed disturbance terms, the Kalman filter is applied to obtain optimal estimates for the state variables, including the month-to-month changes for the trends. For a general introduction of state-space models and the Kalman filter, see Harvey (1989) or Durbin and Koopman (2001). The Kalman filter assumes that the hyperparameters are known in advance. In this application the unknown hyperparameters are substituted with their maximum likelihood (ML) estimates. The uncertainty in the small area predictions of using ML estimates for the unknown hyperparameters is ignored. The analysis is conducted with software developed in OxMetrics in combination with the subroutines of SsfPack 3.0, (Doornik, 2009; Koopman et al., 1999, 2008). All state variables are non-stationary with the exception of the survey errors. The non-stationary variables are initialized with a diffuse prior, i.e. the expectations of the initial states are equal to zero and the initial covariance matrix of

the states is diagonal with elements diverging to infinity. The survey errors are stationary and therefore initialized with a proper prior. The initial values for the survey errors are equal to zero and the covariance matrix is available from the aforementioned model for the survey errors. The exact initial solution for the Kalman filter with diffuse initial conditions, proposed by Koopman (1997) is used. ML estimates for the hyperparameters are obtained using the numerical optimization procedure maxBFGS in OxMetrics.

If correlations between the slope disturbances of the trends are non-zero, then the correlation matrix of the transition equation is non-diagonal. Let V_R denote the $m_A \times m_A$ block of the covariance matrix of the transition equation that models the variances and covariances of the slope disturbances of the trend. In the case of strong correlations between the domains, the rank of V_R can become smaller than the number of domains m_A. Therefore the covariance matrix is implemented as a Cholesky decomposition, i.e. $V_R = A_R D_R A_R'$, with D_R a diagonal matrix and A_R a lower triangular matrix with ones on the diagonal. If the slope disturbances for the m_A domains of a component, are strongly correlated, then this component might be driven by $q < m_A$ common factors. This implies that V_R is not of full rank. In the context of structural time-series models this is often referred to as cointegration. If $m_A - q$ ML estimates of the diagonal elements of D_t tend to zero, then the variances of the slope disturbances are obtained as a linear combination of q common trend factors. Also the state variables of the trends for the m_A domains can be expressed as a linear combination of q common factors, up to a time invariant correction term. See Harvey (1989), Section 8.5 or Koopman et al. (2007), Section 9.1 for more details concerning cointegration and common factor state-space models.

4.2 Time-series multilevel model

For the description of the multilevel time-series model the initial estimates $\hat{\bar{Y}}_{itp}$ are combined into a vector $\hat{\bar{Y}} = (\hat{\bar{Y}}_{111}, \hat{\bar{Y}}_{112}, \dots, \hat{\bar{Y}}_{115}, \hat{\bar{Y}}_{121}, \dots)'$, i.e., wave index runs faster than time index which runs faster than area index. The numbers of areas, periods and waves are denoted by m_A, m_T and m_P, respectively. The total length of $\hat{\bar{Y}}$ is therefore $m = m_A m_T m_P = 12(areas) * 72(months) * 5(waves) = 4320$. Similarly, the variance estimates $v(\hat{\bar{Y}}_{itp})$ are put in the same order along the diagonal of a $m \times m$ covariance matrix Φ.

The covariance matrix Φ is not diagonal because of the correlations induced by the panel design. It is a sparse band matrix, and the ordering of the vector $\hat{\bar{Y}}$ is such that it achieves minimum possible bandwidth, which is advantageous from a numerical computational point of view. To illustrate, the pattern of non-vanishing elements of the top-left 500×500 block of Φ is displayed in Fig. 6. This part of the covariance matrix refers to all $72 * 5$ estimates for the first province, and some estimates for the second province. The figure clearly shows the correlation structure induced by the rotating panel design, as well as independence between estimates for different provinces, resulting in a block diagonal matrix with a relatively small number of nonzero bands around the diagonal.

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

$$\hat{\bar{Y}} = X\beta + \sum_{\alpha} Z^{(\alpha)} v^{(\alpha)} + e\,, \tag{22}$$

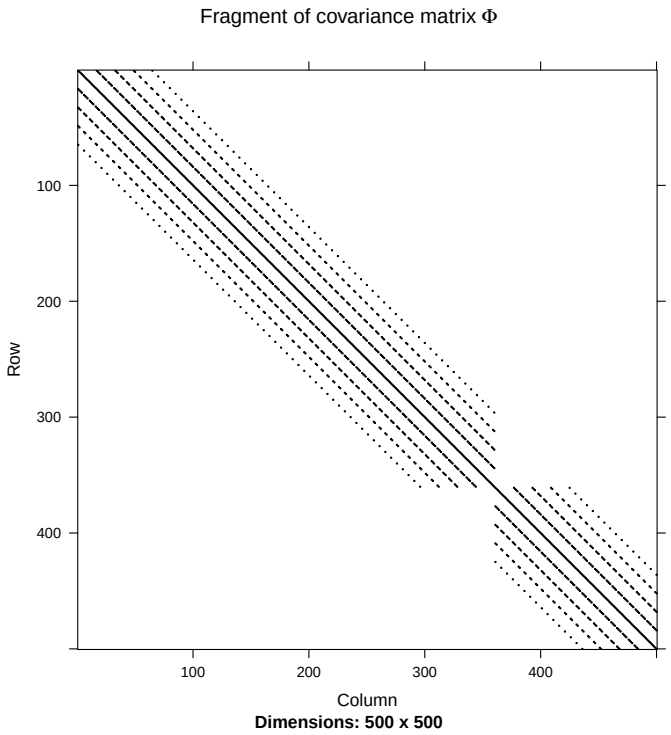

Figure 6 The top-left 500×500 block of Φ.

where X is a $m \times p$ design matrix for the fixed effects β, and the $Z^{(\alpha)}$ are $m \times q^{(\alpha)}$ design matrices for random effect vectors $v^{(\alpha)}$. The sampling errors $e = (e_{111}, e_{112}, \ldots, e_{115}, e_{121}, \ldots)'$ are taken to be normally distributed as

$$e \sim N(0, \Sigma) \tag{23}$$

where $\Sigma = \bigoplus_{i=1}^{m_A} \lambda_i \Phi_i$ with Φ_i the covariance matrix for the initial estimates for province i, and λ_i a province-specific variance scale parameter to be estimated. As described in Section 3 the design variances in $\Phi = \bigoplus_i \Phi_i$ are pooled over provinces and because of the discrete nature of the unemployment data they thereby lose some of their dependence on the unemployment level. It was found that incorporating the variance scale factors λ_i allows the model to rescale the estimated design variances to a level that better fits the data.

To describe the general model for each vector $v^{(\alpha)}$ of random effects, we suppress the superscript α. Each vector v has $q = dl$ components corresponding to d effects allowed to vary over l levels of a factor variable. In particular,

$$v \sim N(0, A \otimes V)\,, \tag{24}$$

where V and A are $d \times d$ and $l \times l$ covariance matrices, respectively. The covariance matrix V is allowed to be parameterised in three different ways. Most generally, it is an unstructured, i.e. fully parameterised covariance matrix. More parsimonious forms are $V = \mathrm{diag}(\sigma_{v;1}^2, \ldots, \sigma_{v;d}^2)$ or $V = \sigma_v^2 I_d$. If $d = 1$ the three parameterisations are equivalent. The covariance matrix A describes the covariance structure between the levels of the factor variable, and is assumed to be known. It is typically more convenient to use the precision matrix $Q_A = A^{-1}$ as it is sparse for many common temporal and spatial correlation structures (Rue and Held, 2005).

A single smooth trend can be represented as a random intercept ($d = 1$) varying over time ($l = m_T$), with temporal correlation determined by the $m_T \times m_T$ band sparse precision matrix

(see e.g. Rue and Held (2005)),

$$Q_A = \begin{pmatrix} 1 & -2 & 1 & & & & & \\ -2 & 5 & -4 & 1 & & & & \\ 1 & -4 & 6 & -4 & 1 & & & \\ & 1 & -4 & 6 & -4 & 1 & & \\ & & \ddots & \ddots & \ddots & & & \\ & & & 1 & -4 & 6 & -4 & 1 \\ & & & & -2 & 5 & -4 & 1 \\ & & & & & 1 & -2 & 1 \end{pmatrix}. \tag{25}$$

In this case $V = \sigma_v^2$ and the design matrix Z is the $m \times m_T$ indicator matrix for month, i.e. the matrix with a single 1 in each row for the corresponding month and 0s elsewhere. The sparsity of both Q_A and Z can be exploited in computations.

The precision matrix (25) has two singular vectors, $\iota_{m_T} = (1, 1, \ldots, 1)$ and $(1, 2, \ldots, m_T)'$. This means that the corresponding specification (24) is completely uninformative about the overall level and linear trend. In order to prevent unidentifiability among various terms in the model, the overall level and trend can be removed from v by imposing the constraints $Rv = 0$, where R is the $2 \times m_T$ matrix with the two singular vectors as its rows. The overall level and trend are then included in the vector β of fixed effects.

A smooth trend for each province is obtained with $d = m_A$, $l = m_T$, and V a $m_A \times m_A$ covariance matrix, either diagonal with a single variance parameter, diagonal with m_A variance parameters, or unstructured, i.e. fully parametrised in terms of m_A variance parameters and $m_A(m_A - 1)/2$ correlation parameters. The design matrix is $I_{m_A} \otimes I_{m_T} \otimes \iota_{m_P}$ in this case.

An alternative trend model consists of a single global smooth trend (also known as a second order random walk) supplemented by a local level trend, i.e. an ordinary (first-order) random walk, for each province. The latter can be modeled as discussed in the previous paragraph, but with Q_A in (25) replaced by the precision matrix for a first order random walk, given by equation (34) in the appendix. In contrast to the STM approach, it is not necessary to remove one of the provincial random walk trends from the model for identifiability. The reason for that is that in the multilevel approach constraints are imposed to ensure that the smooth overall trend as well as all provincial random walk trends sum to zero over time. The constrained components correspond to a global and provincial intercepts, which are separately included in the model as fixed effects with one provincial fixed effect excluded.

Seasonal effects can be expressed in terms of correlated random effects (24) as well. The trigonometric seasonal (13) is equivalent to the balanced dummy variable seasonal model (Proietti, 2000; Harvey, 2006), corresponding to first-order random walks over time for each month, subject to a sum-to-zero constraint over the months. In this case $d = 12$ (seasons), $V = \sigma_v^2 I_{12}$, and $l = m_T$ with Q_A the first-order random walk precision matrix (34). The sum-to-zero constraints over seasons at each time, together with the sum-to-zero constraints over time of each random walk can be imposed as $Rv = 0$ with R the $(m_T + 12) \times 12m_T$ matrix

$$R = \begin{pmatrix} \iota_{12}' \otimes I_{m_T} \\ I_{12} \otimes \iota_{m_T}' \end{pmatrix}. \tag{26}$$

Together with fixed effects for each season (again with a sum-to-zero constraint imposed) this random effect term is equivalent to the trigonometric seasonal. It can be extended to a seasonal for each province, with a separate variance parameter for each province.

To account for the RGB, the multilevel model includes fixed effects for waves 2 to 5. These effects can optionally be modeled dynamically by adding random walks over time for each wave. Another choice to be made is whether the fixed and random effects are crossed with province.

Further fixed effects can be included in the model, for example those associated with the auxiliary variables used in the survey regression estimates. Some fixed effect interactions, for example season × province or wave × province might alternatively be modeled as random effects to reduce the risk of overfitting.

Finally, a white noise term can be added to the model, to account for unexplained variation by area and time in the signal.

Model (22) can be regarded as a generalization of the Fay-Herriot area-level model. The Fay-Herriot model only includes a single vector of uncorrelated random effects over the levels of a single factor variable (typically areas). The models used in this paper contain various combinations of uncorrelated and correlated random effects over areas and months. Earlier accounts of multilevel time-series models extending the Fay-Herriot model are Rao and Yu (1994); Datta et al. (1999); You (2008). Datta et al. (1999) and You (2008) use time-series models with independent area effects and first-order random walks over time for each area. In Rao and Yu (1994) a model is used with independent random area effects and a stationary autoregressive AR(1) instead of a random walk model over time. In You et al. (2003) the random walk model was found to fit the Canadian unemployment data slightly better than AR(1) models with autocorrelation parameter fixed at 0.5 or 0.75. Compared to the aforementioned references a novel feature of our model is that smooth trends are considered instead of or in addition to first-order random walks or autoregressive components. We also include independent area-by-time random effects as a white noise term accounting for unexplained variation at the aggregation level of interest.

4.2.1 Estimating the time-series multilevel model

A Bayesian approach is used to fit model (22)-(24). This means we need prior distributions for all (hyper)parameters in the model. The following priors are used:

- The data-level variance parameters λ_i for $i = 1, \ldots, m_A$ are assigned inverse chi-squared priors with both degree of freedom parameter and scale parameter equal to 1.
- The fixed effects are assigned a normal prior with zero mean and fixed diagonal variance matrix with very large values (1e10).
- For a fully parameterized covariance matrix V in (24) we use the scaled-inverse Wishart prior as proposed in O'Malley and Zaslavsky (2008) and recommended by Gelman and Hill (2007). Conditionally on a d-dimensional vector parameter ξ,

$$V|\xi \sim \text{Inv} - \text{Wishart}\left(V|\nu, \text{diag}(\xi)\Psi\text{diag}(\xi)\right) \tag{27}$$

  where $\nu = d + 1$ is chosen, and $\Psi = I_d$. The vector ξ is assigned a normal distribution $N(0, I_d)$.
- All other variance parameters appearing in a diagonal matrix V in (24) are assigned, conditionally on an auxiliary parameter ξ, inverse chi-squared priors with 1 degree of freedom and scale parameter ξ^2. Each parameter ξ is assigned a $N(0, 1)$ prior. Marginally, the standard deviation parameters have half-Cauchy priors. Gelman (2006) demonstrates that these priors are better default priors than the more common inverse chi-squared priors.

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

For each model considered, the Gibbs sampler is run in three independent chains with randomly generated starting values. Each chain is run for 2500 iterations. The first 500 draws are discarded as a "burn-in sample". From the remaining 2000 draws from each chain, we keep every fifth draw to save memory while reducing the effect of autocorrelation between successive draws. This leaves $3 * 400 = 1200$ draws to compute estimates and standard errors. The convergence of the MCMC simulation is assessed using trace and autocorrelation plots as well as the Gelman-Rubin potential scale reduction factor (Gelman and Rubin, 1992), which diagnoses the mixing of the chains. The diagnostics suggest that all chains converge well within 500 draws. Also, the estimated Monte Carlo simulation errors (accounting for the remaining autocorrelation in the chains) are small compared to the posterior standard errors for all parameters, so that the number of retained draws is sufficient for our purposes.

The estimands of interest can be expressed as functions of the parameters, and applying these functions to the MCMC output for the parameters results in draws from the posteriors for these estimands. In this paper we summarize those draws in terms of their mean and standard deviation, serving as estimates and standard errors, respectively. All estimands considered can be expressed as linear predictors, i.e. as linear combinations of the model parameters. Estimates and standard errors for the following estimands are computed:

- Signal: the vector θ_{it} including all fixed and random effects, except those associated with waves 2 to 5. These correspond to the fitted values $X\beta + \sum_{\alpha} Z^{(\alpha)} v^{(\alpha)}$ associated with each fifth row $1, 6, 11, \ldots$ of $\bar{Y}$ and the design matrices.
- Trend: prediction of the long-term trend. This is computed by only incorporating the trend components of each model in the linear predictor. For most models considered the trend corresponds to seasonally adjusted figures, i.e. predictions of the signal with all seasonal effects removed.
- Growth of trend: the differences between trends at two consecutive months.

5 Results

In this section, the results obtained with the structural time-series models and multilevel time-series models are described in subsections (5.1) and (5.2) respectively. Before results are discussed, two discrepancy measures are defined to evaluate and compare the different models. The first measure is the Mean Relative Bias (MRB) and summarizes the differences

between model estimates and direct estimates averaged over time, as percentage of the latter. For a given model M, the MRB_i is defined as

$$MRB_i = \frac{\sum_t (\hat{\theta}_{it}^{\mathrm{M}} - \hat{\theta}_{it}^{\mathrm{direct}})}{\sum_t \hat{\theta}_{it}^{\mathrm{direct}}} \times 100\% . \tag{28}$$

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

$$RRSE_i = 100\% \times \frac{1}{m_T} \sum_t (se(\hat{\theta}_{it}^{\mathrm{direct}}) - se(\hat{\theta}_{it}^{\mathrm{M}}))/se(\hat{\theta}_{it}^{\mathrm{direct}}) , \tag{29}$$

for a given model M.

5.1 Results STM

Ten different structural time-series models are compared. Four different trend models are distinguished. The first trend component is a smooth trend model without correlations between the domains (12), abbreviated as T1. The second trend model is a smooth trend model (12) with a full correlation matrix for the slope disturbances (19), abbreviated as T2. The third trend component is a common smooth trend model for all provinces with eleven local level trend models for the deviation of the domains from this overall trend ((20) in combination with (21)), abbreviated as T3. The fourth trend model is a common smooth trend model for all provinces with eleven smooth trend models for the deviation of the domains from this overlall trend ((20) in combination with (12)), abbreviated as T4. In T3 and T4 the province Groningen is taken equal to the overall trend. The component for the RGB (14) can be domain specific (indicated with an R) or chosen equal for all domains (no indication for RGB). An alternative simplicfication is to assume that RGB for waves 2, 3, 4 and 5 are equal but domain specific (indicated with R2). In a similar way the seasonal component (13) can be chosen domain specific (indicated with an S) or taken equal for all domains (no indication for seasonal). All models share the same component for the survey error, i.e. an AR(1) model with time varying autocorrelation coefficients for wave 2, through 5 to model the autocorrelation in the survey errors. The following structural time-series models are compared:

- T1SR: Smooth trend model and no correlation between slope disturbances, seasonal and RGB domain specific.
- T2SR: Smooth trend model with a full correlation matrix for the slope disturbances, seasonal and RGB domain specific.
- T2S: Smooth trend model with a full correlation matrix for the slope disturbances, seasonal domain specific, RGB equal over all domains.
- T2R: Smooth trend model with a full correlation matrix for the slope disturbances, seasonal equal over all domains, RGB domain specific.
- T3SR: One common smooth trend model for all domains plus eleven local levels for deviations from the overall trend, seasonal and RGB domain specific.
- T3R: One common smooth trend model for all domains plus eleven local levels for deviations from the overall trend, seasonal equal over all domains, RGB domain specific.

- T3R2: One common smooth trend model for all domains plus eleven local levels for deviations from the overall trend, seasonal equal over all domains, RGB is domain specific but assumed to be equal for the four follow-up waves.
- T3: One common smooth trend model for all domains plus eleven local levels for deviations from the overall trend, seasonal and RGB equal over all domains.
- T4SR: One common smooth trend model for all domains plus eleven smooth trend models for deviations from the overall trend, seasonal and RGB domain specific.
- T4R: One common smooth trend model for all domains plus eleven smooth trend models for deviations from the overall trend, seasonal equal over all domains, RGB domain specific.

For all models, the ML estimates for the hyperparameters of the RGB and the seasonals tend to zero, which implies that these components are time invariant. Also the ML estimates for the variance components of the white noise of the population domain parameters tend to zero. This component is therefore removed from model (11). The ML estimates for the variance components of the survey errors in the first wave vary between 0.93 and 1.90. For the follow-up waves, the ML estimates vary between 0.86 and 1.80. The variances of the direct estimates are pooled over the domains (3), which might introduce some bias, e.g. underestimation of the variance in domains with high unemployment rates. Scaling the variances of the survey errors with the ML estimates for $\sigma^2_{v_{ip}}$ is neccessary to correct for this bias.

The ML estimates for the hyperparameters for the trend components are summarized in Table 2 for model T1SR, T2SR, T3SR and T4SR. Hyperparameters for similar trend models where the RGB or seasonal component is equal over the domains have almost similar hyperparameter estimates for the trend. For T2SR only the eigenvalues of the diagonal elements of D_R from the Cholesky decomposition are presented. In Table 3 the corresponding correlation matrix for the slope disturbances is given.

For model T1SR the hyperparameters for slope disturbances for the 12 domains are of the same order. For model T2SR it follows that the dynamics of these 12 domains can be modelled with only 2 underlying common trends. The hyperparameters of the overall trends in T3SR and T4SR are of the same order as the domain specific trend hyperparameters of T1SR. The hyperparameters of the domain specific deviations from the overall trend are larger under the local level model (T3SR) compared to the smooth trend model (T4SR).

Models are compared using the loglikelihoods. To account for differences in model complexity, Akaike Information Criteria (AIC) and Bayes Information Criteria (BIC) are used, Durbin and Koopman (2001), Section 7.4. Results are summarized in Table 4. Parsimonious models where the seasonals or RGB are equal over the domains are preferred by the AIC or BIC criteria. Note, however, that the likelihoods are not completely comparable between models. To obtain comparable likelihoods, the first 24 months of the series are ignored in the computation of the likelihood for all models. Some of the likelihoods are nevertheless odd. For example the likelihood of T2SR is smaller than the likelihood of T2S, although T2SR contains more model parameters. This is probably the result of large and complex time-series models in combination with relatively short time series, which gives rise to flat likelihood functions. Also from this point of view, sparse models that avoid over-fitting are still favorable, which is in line with the results of the AIC and BIC values in Table 4.

Modelling correlations between slope disturbances of the trend results in a significant model improvement. Model T1SR, e.g. is nested within T2SR and a likelihood ratio test clearly favours

Trend component	T1SR	T2SR	T3SR	T4SR
Overall			1.48e-04	1.51E-04
Grn	2.26e-04	1.94e-04		
Frsl	1.61e-04	5.73e-10	3.19e-04	3.04e-10
Drn	2.19e-04	0	2.32e-04	1.64e-09
Ovr	1.37e-04	0	1.50e-04	4.50e-10
Flv	2.98e-04	0	1.35e-03	1.41e-04
Gld	1.56e-04	0	1.97e-08	1.12e-09
Utr	1.33e-04	0	3.79e-04	4.73e-07
N-H	1.60e-04	0	3.07e-04	1.63e-10
Z-H	1.59e-04	0	2.60e-04	1.65e-07
Zln	1.30e-04	0	3.93e-04	4.23e-10
N-B	1.39e-04	0	3.11e-04	3.03e-07
Lmb	1.31e-04	0	1.06e-09	1.01e-09

Table 2 ML hyperparameter estimates trend components.

dom.	Grn	Frsl	Drn	Ovr	Flv	Gld	Utr	N-H	Z-H	Zln	N-B
Frsl	1.000										
Drn	0.977	0.977									
Ovr	0.999	0.999	0.968								
Flv	1.000	1.000	0.979	0.999							
Gld	0.991	0.991	0.997	0.985	0.992						
Utr	1.000	1.000	0.983	0.997	1.000	0.995					
N-H	1.000	1.000	0.983	0.998	1.000	0.994	1.000				
Z-H	0.999	0.999	0.964	1.000	0.998	0.983	0.997	0.997			
Zln	1.000	1.000	0.981	0.998	1.000	0.994	1.000	1.000	0.997		
N-B	0.997	0.997	0.991	0.992	0.997	0.999	0.999	0.999	0.991	0.998	
Lmb	0.986	0.986	0.928	0.992	0.984	0.955	0.980	0.981	0.993	0.982	0.969

Table 3 Correlation matrix slope disturbances (diagonal elements are omitted)

Model	log likelihood	states	hyperparameters	AIC	BIC
T1SR	9813.82	204	24	-399.41	-390.52
T2SR	9862.86	204	35	-400.99	-391.68
T2S	9879.03	160	35	-403.50	-395.90
T2R	9859.97	83	35	-405.92	-401.32
T3SR	9855.35	193	24	-401.60	-393.14
T3R	9851.62	72	24	-406.48	-402.74
T3R2	9871.65	36	24	-408.82	-406.48
T3	9881.16	28	24	--409.55	-407.52
T4SR	9857.47	204	24	-401.23	-392.34
T4R	9853.65	83	24	-406.11	-401.94

Table 4 AIC and BIC for the structural time-series models.

the latter. The correlations between the slope disturbances of the domains are very strong as follows from Table 3. The rank of this 12×12 covariance matrix equals two, which implies that the covariance matrix is actually driven by two underlying common factors. This also follows from Table 2 where only 2 out of the 12 eigenvalues are unequal to zero. As a result the full covariance matrix for the slope disturbances of the 12 domains is actually modelled with 23 instead of 78 hyperparameters.

In Table 5 the MRB, defined by (28), for the ten different models are shown. Models that assume that the RGB is equal over the domains, i.e. T2S and T3, have large relative biases for some of the domains. Large biases occur in the domains where unemployment is large (e.g. Groningen) or small (e.g. Utrecht) compared to the national average. A possible compromise between parsimony and bias is to assume that the RGB is equal for the four follow-up waves but still domain specific (T3R2). For this model the bias is small, with the exception of Gelderland.

	Grn	Frs	Drn	Ovr	Flv	Gld	Utr	N-H	Z-H	Zln	N-B	Lmb
T1SR	1.1	0.5	2.0	-0.2	0.1	3.4	0.1	0.6	1.7	-2.1	0.5	2.1
T2SR	1.2	0.7	2.2	-0.1	0.2	3.5	0.2	0.6	1.7	-2.1	0.5	2.1
T2S	-3.1	3.1	0.7	0.9	-4.4	2.8	2.4	0.8	0.5	1.7	1.8	1.5
T2R	0.9	0.8	1.8	-0.2	-0.4	3.4	0.1	0.6	1.7	-1.6	0.6	2.2
T3SR	0.8	0.6	2.0	-0.2	-0.3	3.5	0.3	0.5	1.7	-2.0	0.6	2.0
T3R2	-0.1	1.3	2.1	-0.6	-0.8	3.6	0.9	0.6	1.5	-1.1	1.0	1.2
T3R	0.5	0.7	1.8	-0.2	-0.8	3.5	0.3	0.5	1.6	-1.5	0.7	2.1
T3	-4.0	2.5	0.1	0.9	-5.0	2.8	2.3	0.7	0.6	2.5	2.0	1.3
T4SR	0.8	0.7	2.1	-0.2	-0.0	3.5	0.2	0.6	1.7	-1.9	0.5	2.1
T4R	0.6	0.7	1.8	-0.2	-0.6	3.4	0.1	0.6	1.7	-1.3	0.7	2.1

Table 5 Mean Relative Bias averaged (28) over time (%), per province for structural time-series models.

In Figure 7 and 8 the smoothed trends and standard errors of models T1SR, T2SR and T2S are compared. The month-to-month development of the trend and the standard errors for these three models are compared in Figures 9 and 10. The smoothed trends obtained with the common trend model are slightly more flexible compared to a model without correlation between the slope disturbances. This is clearly visible in the month-to-month change of the trends. Modeling the correlation between slope disturbances clearly reduces the standard error of the trend and the month-to-month change of the trend. Assuming that the RGB is equal for all domains (model T2S) clearly affects the level of the trend and further reduces the standard error, mainly since the number of state variables are reduced. The difference between the trend under T2SR and T2S is a level shift. This clearly follows from the month-to-month changes of the trend under model T2SR and T2S, which are exactly equal. According to AIC and BIC the reduction of the number of state variables by assuming equal RGB for all domains is an improvement of the model. In this application, however, interest is focused on the model fit for the separate domains. Assuming that the RGB is equal over all domains is on average efficient for overall goodness of fit measures, like AIC and BIC, but not necessarily for all separate domains. The bias introduced in the trends of some of the domains by taking the RGB equal over the domains is undesirable.

In Figure 11 and 12 the smoothed trends and standard errors of models T2SR, T3SR and T4SR are compared. The month-to-month development of the trend and the standard errors for these three models are compared in Figures 13 and 14. The trends obtained with one overall

smooth trend plus eleven trends for the domain deviations of the overall trend resemble trends obtained with the common trend model. In this application, where only two common trends are required to model the dynamics of the trends observed in the twelve domains is reasonably well approximated with the two alternative trends. This is an empirical finding that cannot be generalized to other situations, particularly when more common factors are required. The common trend model, however, has the smallest standard errors for the trend. Furthermore, the trends under the model with a local level for the domain deviations from the overall trend are in some domains more volatile compared to the other two models. This is most obvious in the month-to-month changes of the trend. It is a general feature for trend models with random levels to have more volatile trends, Durbin and Koopman (2001), Chapter 3. The more flexible trend model of T3 also results in a higher standard error of the month-to-month changes.

Assuming that the seasonals are equal for all domains is another way of reducing the number of state variables and avoid over-fitting of the data. This assumption does not affect the level of the trend since the MRB is small (see Table 5) and results in a significant improvement of the model according to AIC and BIC. Particularly if interest is focused on trend estimates, some bias in the seasonal patterns are acceptable and a model with a trend based on T2, or T4, with the seasonal component assumed equal over the domains, might be a good compromise between a model that accounts sufficiently for differences between domains and model parsimony to avoid over-fitting of the data.

In Figure 15 and 16 the smoothed trends and standard errors of models T2R, T3 and T4R are compared. The month-to-month development of the trend and the standard errors for these three models are compared in Figures 17 and 18. Model T3 is the most parsimonious model that is the best model according to AIC and BIC. Particularly the assumption of equal RGB results in biased trend estimates in some of the domains (see Table 5). The local level component used to model the deviations of the domain-specific trend from the overall trend makes the trends more volatile, which is especially visible in the month-to-month changes. There are some differences between the trends and the month-to-month changes of the trends for models T2R and T4R. It might be expected that the common trend model better fits the dynamics of the trends since it is based on the full covariance matrix of the slope disturbances, while T4 and also T3 implicitly attempt to capture this covariance structure.

To further illustrate the advantages of using time-series models for estimating change, Figure 19 compares the year-to-year change of the trends under models T2R and T3R2 and the direct estimates. Since the seasonal effects under the STM approach are time invariant, the year-to-year changes for the signals are exactly equal to the year-to-year changes of the trends. Note that for year-to-year changes, it makes sense to use direct estimates since seasonal effects cancel out. Figure 20 compares the standard errors of these year-to-year changes. Estimates for year-to-year changes are, compared to the direct estimates, very stable and precise and greatly improve the direct estimates for year-to-year change.

Finally in Figure 21 and 22 the smoothed signals and standard errors for models T2SR and T2R are compared. A main part of the volatility in the series of the direct estimates are sampling errors that are removed from the smoothed signals under T2SR and T2R. Assuming that the seasonals are equal over the domains, results in a less pronounced seasonal pattern. The smoothed signal under T2R has indeed a less pronounced seasonal component than T2SR. The reduction in standard errors under both models compared to the direct estimates under T2SR varies between 40 and 50 percent. Under model T2R the reduction is more than 60 percent.

Table 6 shows the RRSE, defined by (29), for the ten structural time-series models. Table 7 contains the averages of standard errors for signal, trend, and growth (time-differences of trend). The average is taken over all months and provinces. Modelling the correlation between the trends explicitly (Model T2) or implicitly (Model T3 or T4) reduces the standard errors for the trend and signal significantly. The time-series modelling approach is particularly appropriate to estimate month-to-month changes through the trend component. The precision of the month-to-month changes, however, strongly depends on the choice of the trend model. A local level trend model (T3) results in more volatile trends and has a clearly larger standard error for the month-to-month change. Parsimonious models where RGB or the seasonal components are assumed equal over the domains result in further strong standard error reductions at the cost of introducing bias in the trend or the seasonal patterns.

	Grn	Frs	Drn	Ovr	Flv	Gld	Utr	N-H	Z-H	Zln	N-B	Lmb
T1SR	36	36	38	42	43	44	47	47	45	50	47	43
T2SR	43	42	43	48	49	49	53	53	50	54	53	48
T2S	49	48	51	53	55	54	58	56	54	58	56	54
T2R	64	63	62	65	66	63	68	68	63	73	67	64
T3SR	45	41	45	48	42	51	49	50	48	53	49	50
T3R	67	62	63	66	56	61	62	64	65	70	60	66
T3R2	68	63	64	67	57	62	62	65	65	70	60	67
T3	79	74	76	75	65	69	69	69	69	76	63	76
T4SR	43	41	45	48	45	50	49	53	50	54	51	49
T4R	65	63	64	65	62	62	63	68	63	73	63	65

Table 6 Relative reductions in standard error (29) compared to those of the direct estimates (%), per province for the structural time-series models.

	se(signal)	se(trend)	se(growth)
direct	100		
T1SR	57	41	6
T2SR	51	33	4
T2S	46	23	4
T2R	34	33	4
T3SR	53	35	9
T3R	36	35	9
T3R2	36	34	9
T3	28	26	9
T4SR	52	34	4
T4R	35	34	4

Table 7 Means of standard errors over all months and provinces relative to the mean of the direct estimator's standard errors (%) for the structural time-series models.

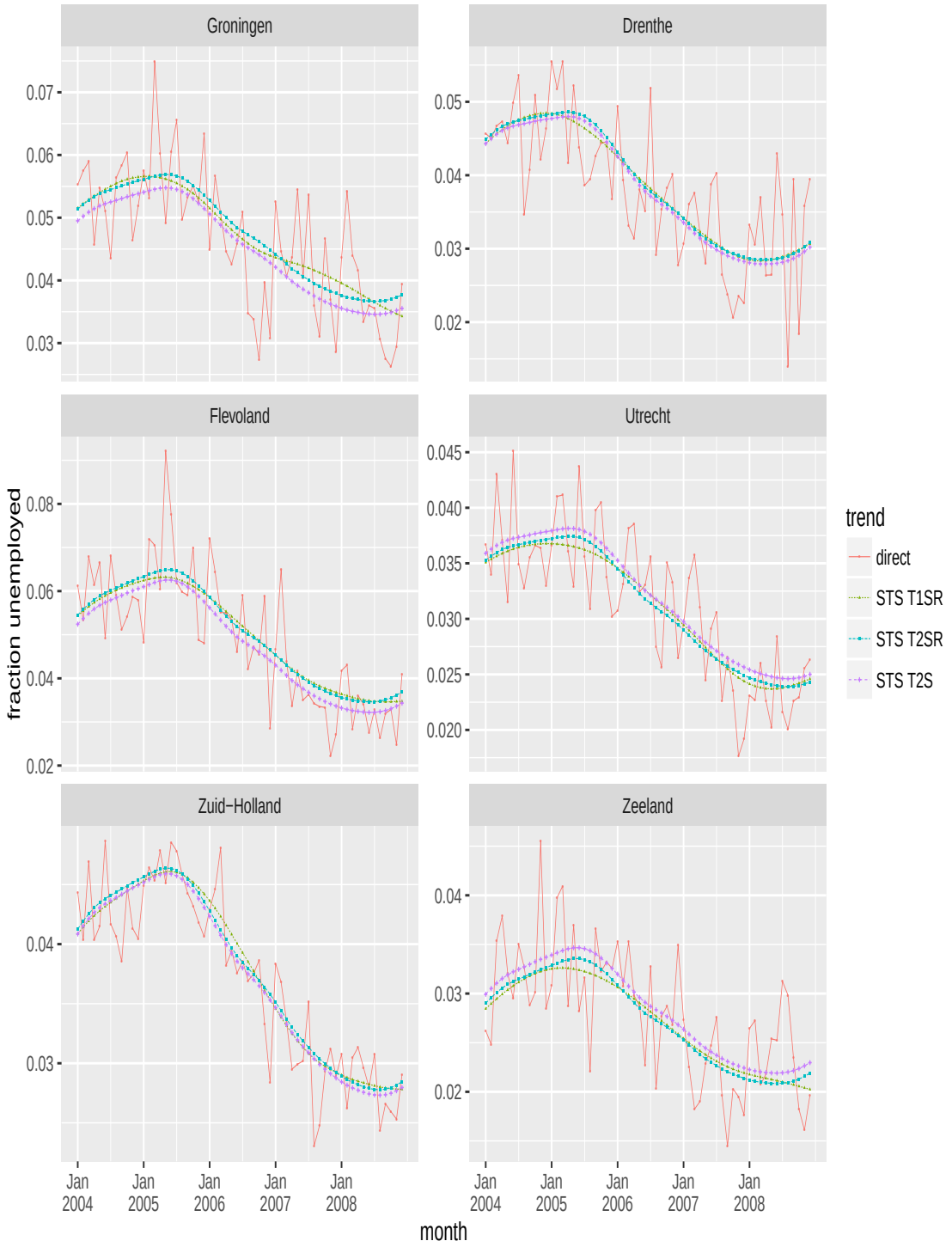

Figure 7 Comparison of estimates.

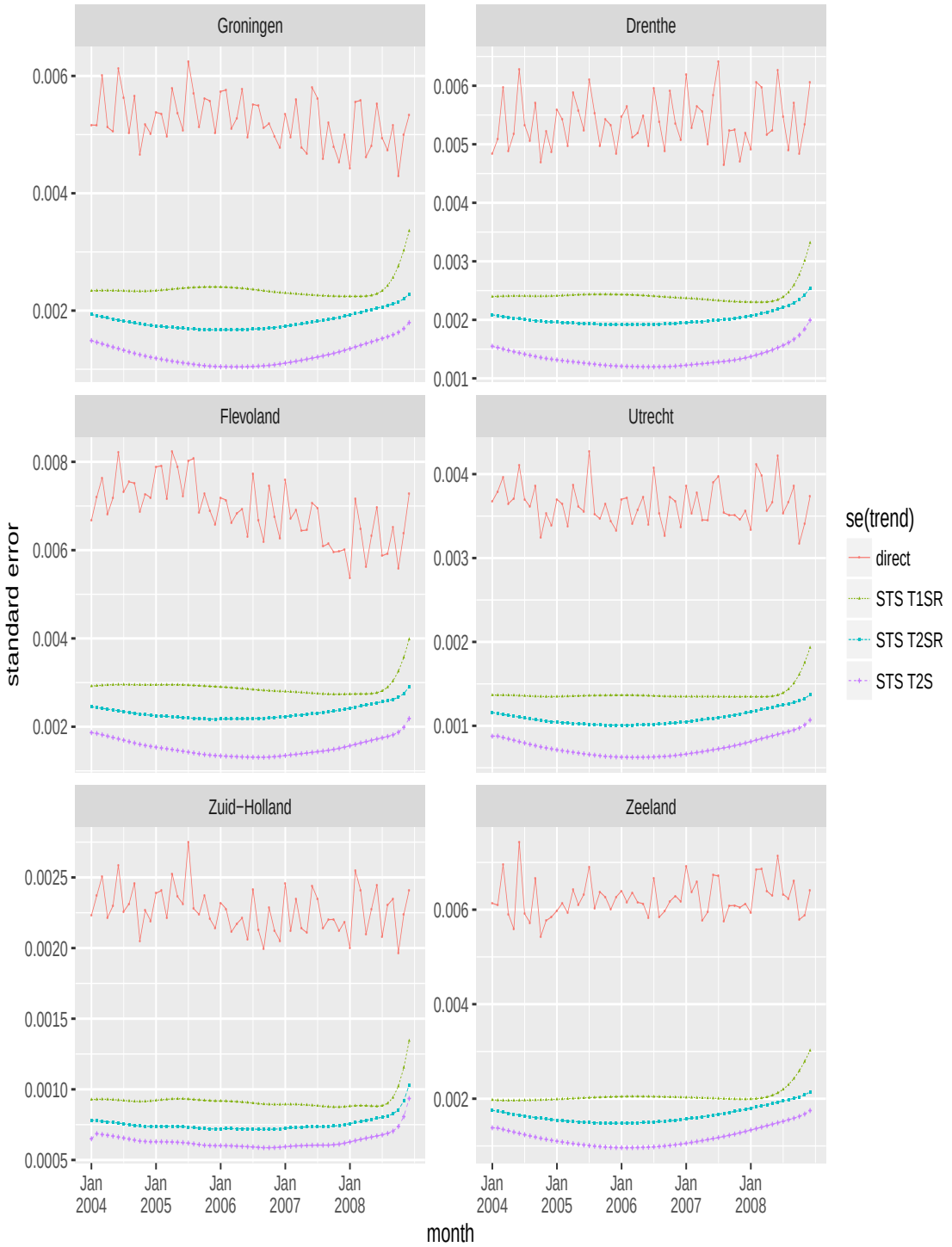

Figure 8 Comparison of standard errors.

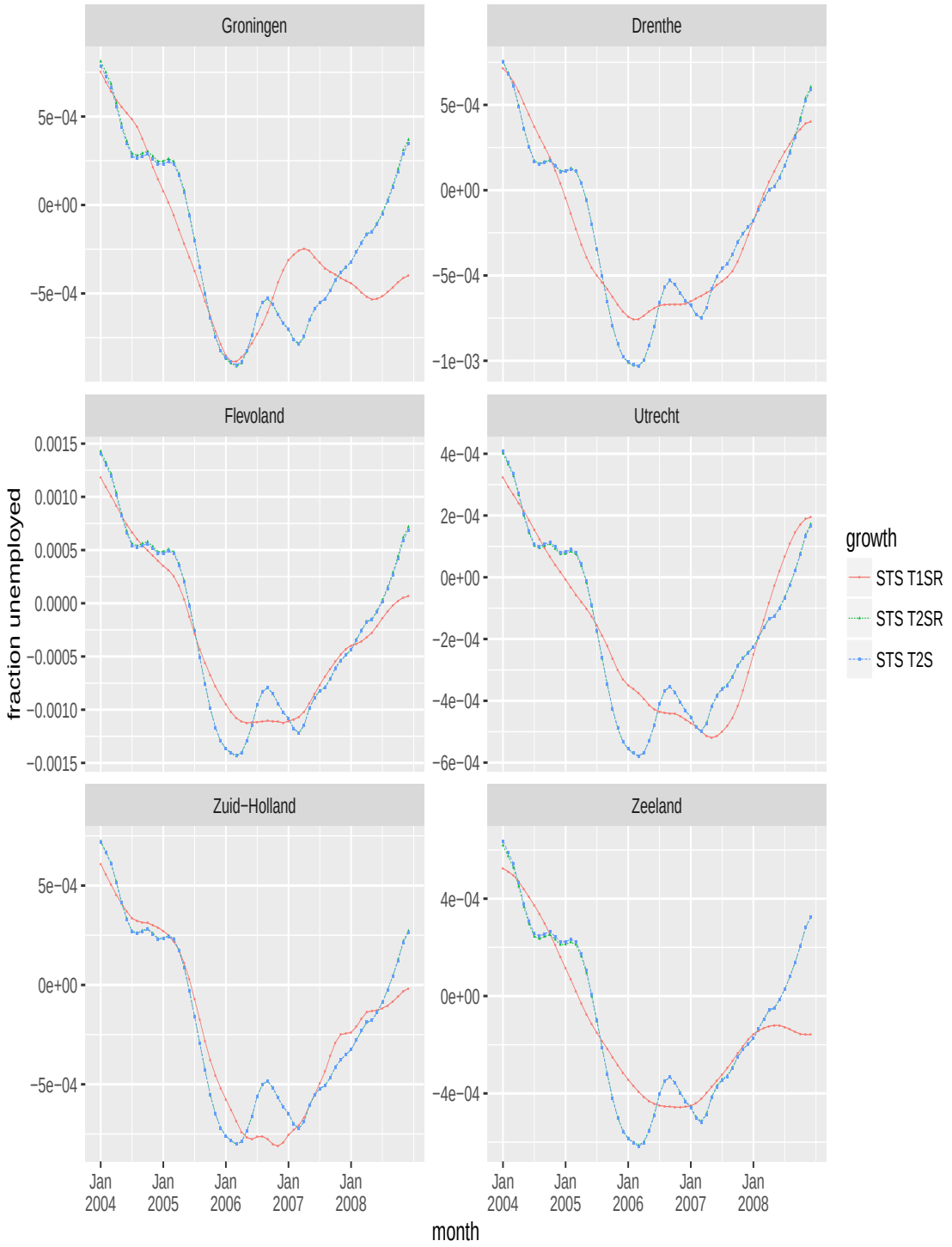

Figure 9 Comparison of estimates.

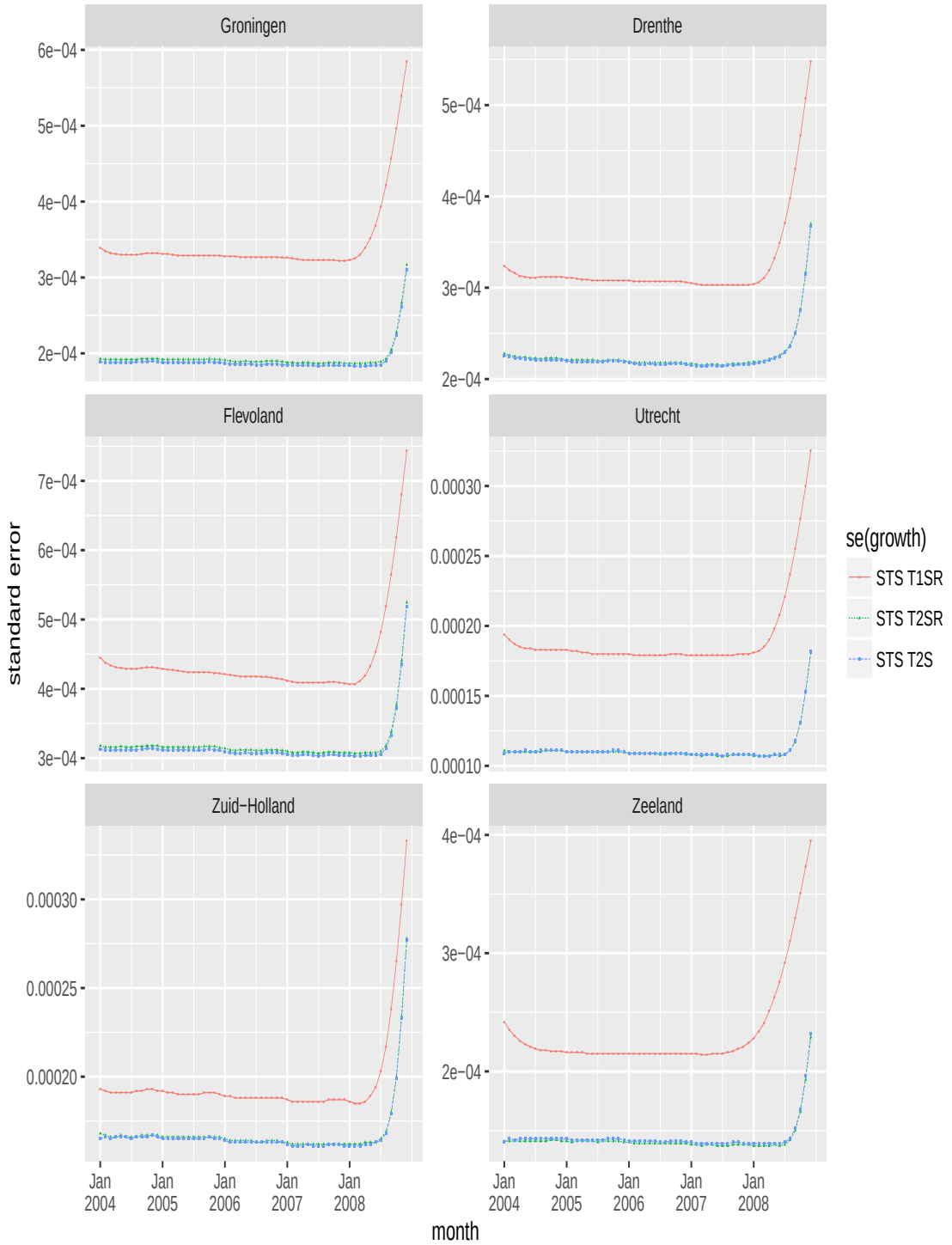

Figure 10 Comparison of standard errors.

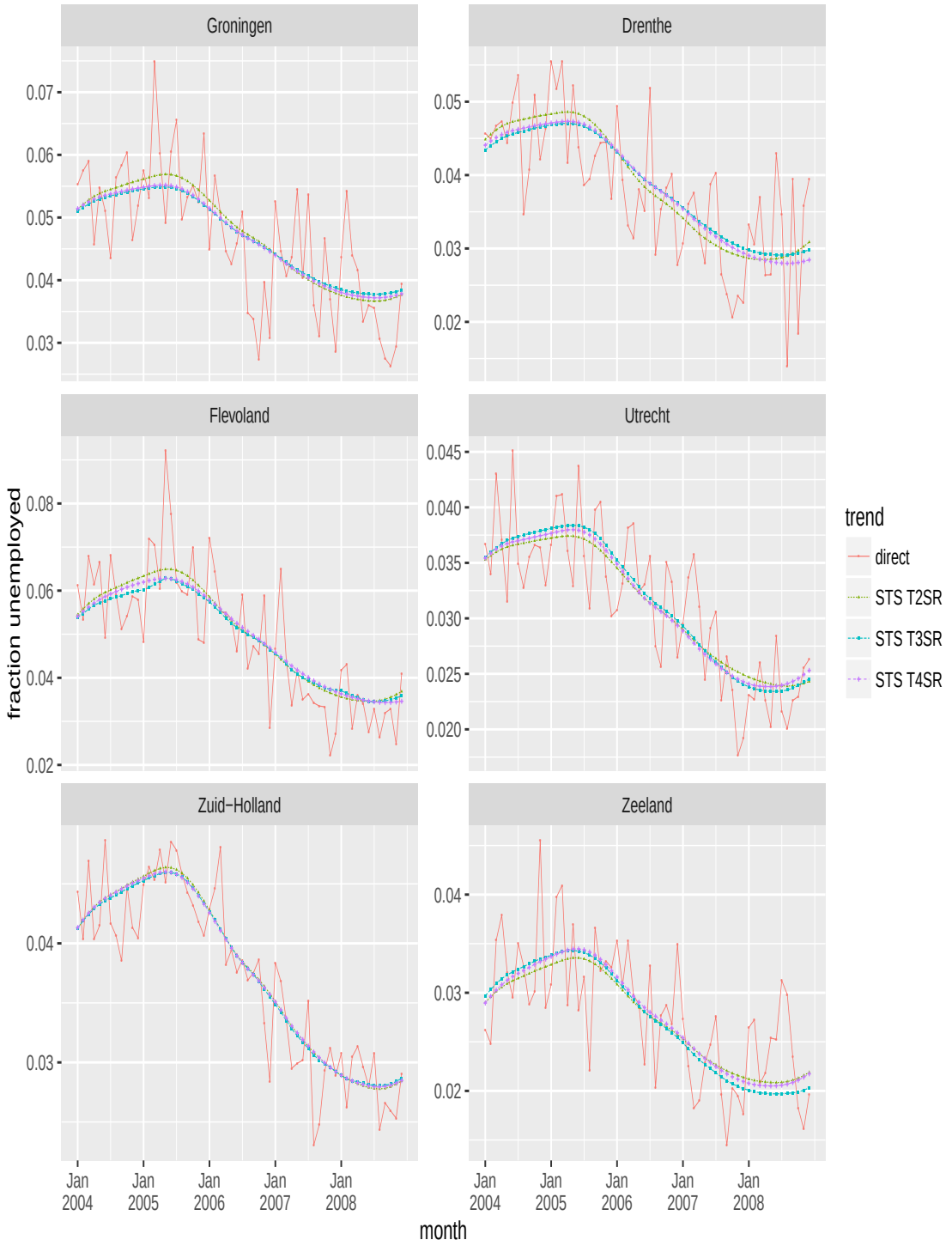

Figure 11 Comparison of estimates.

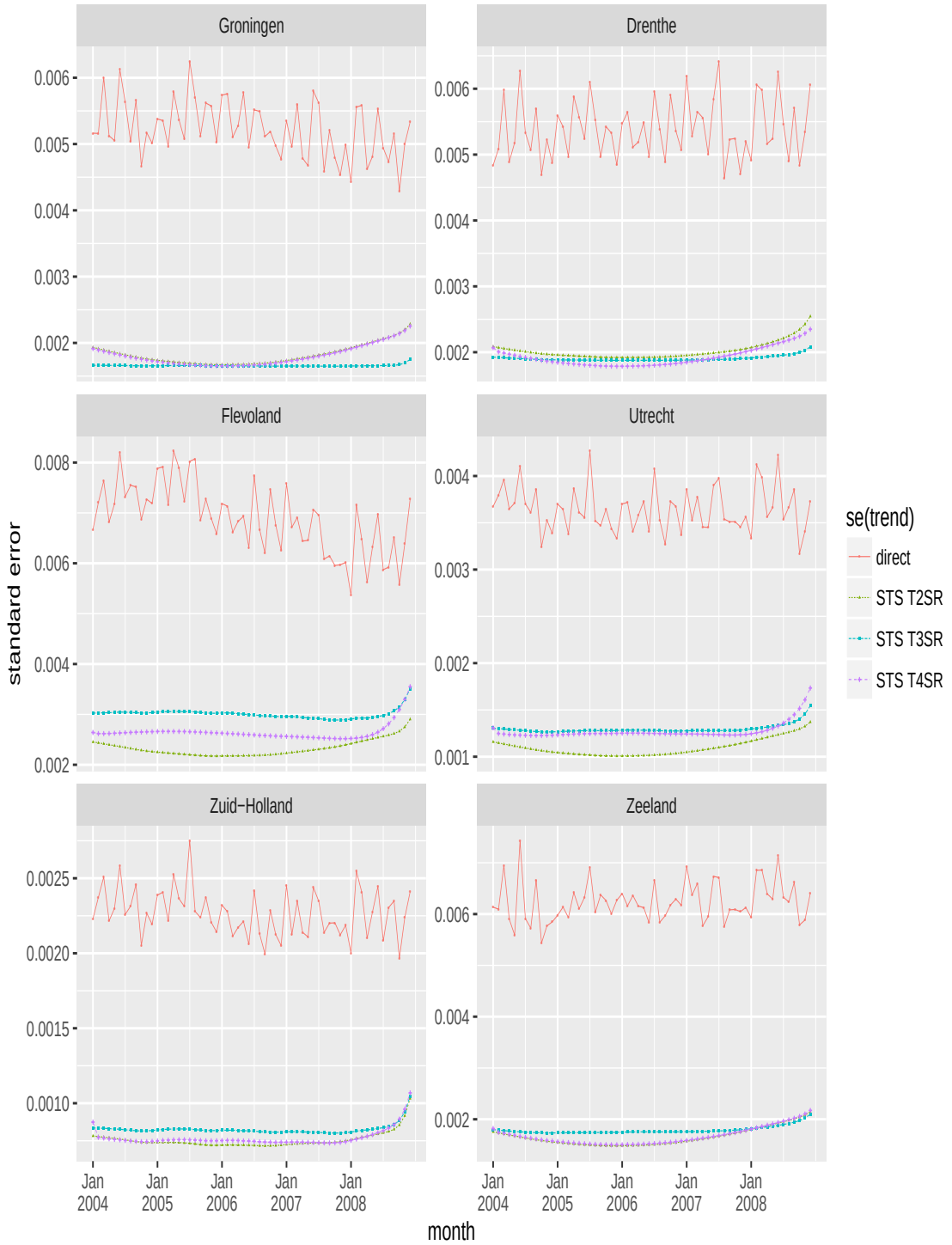

Figure 12 Comparison of standard errors.

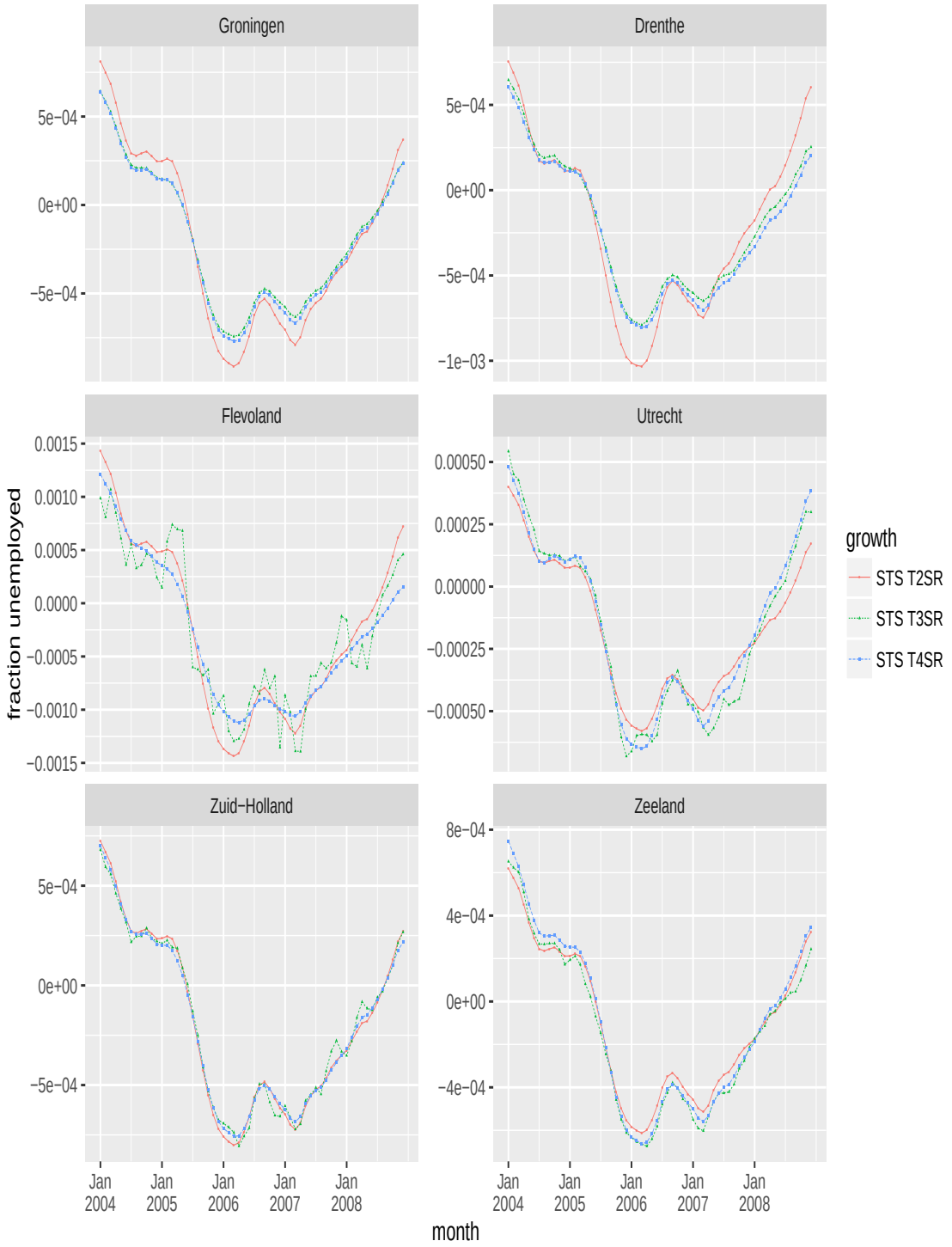

Figure 13 Comparison of estimates.

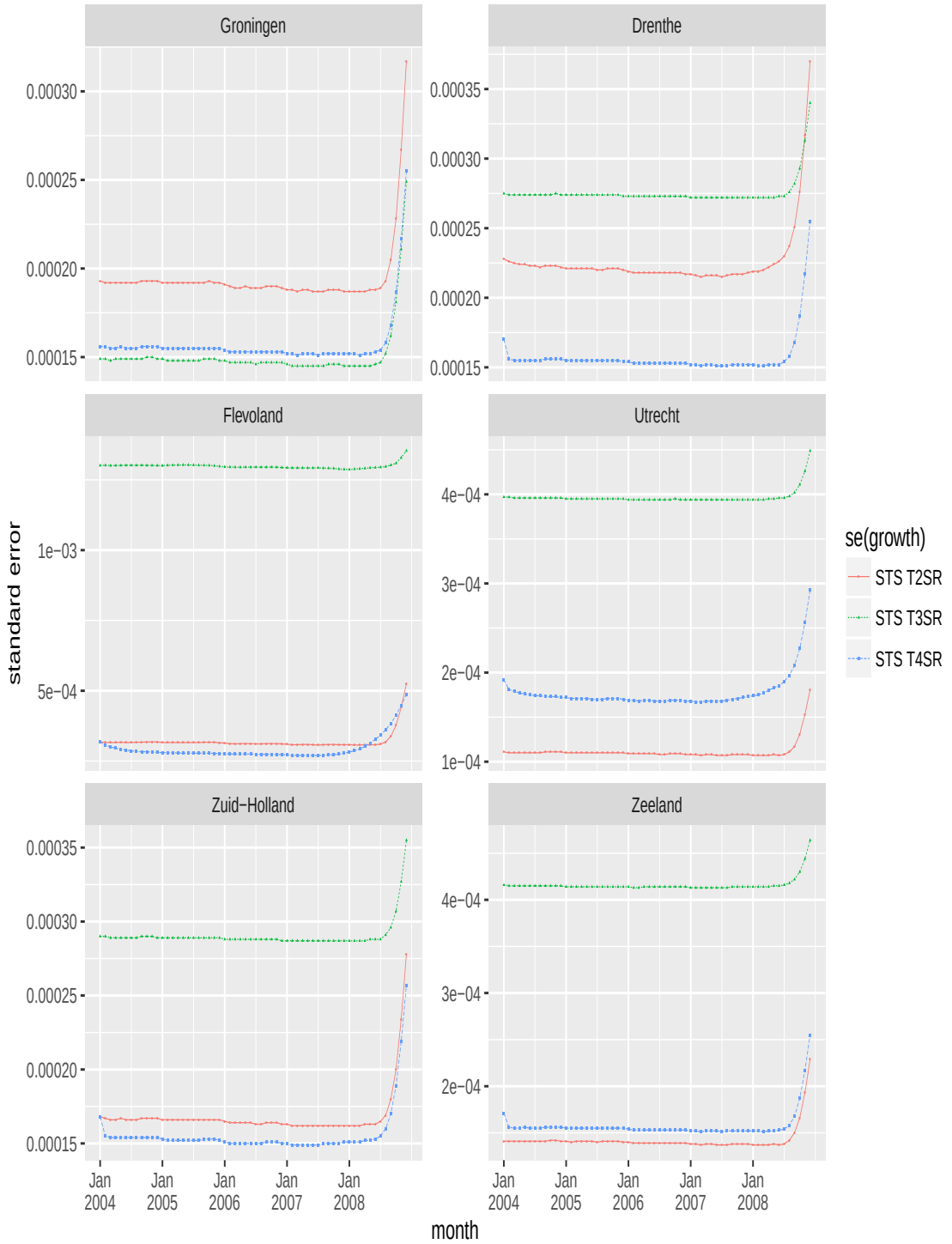

Figure 14 Comparison of standard errors.

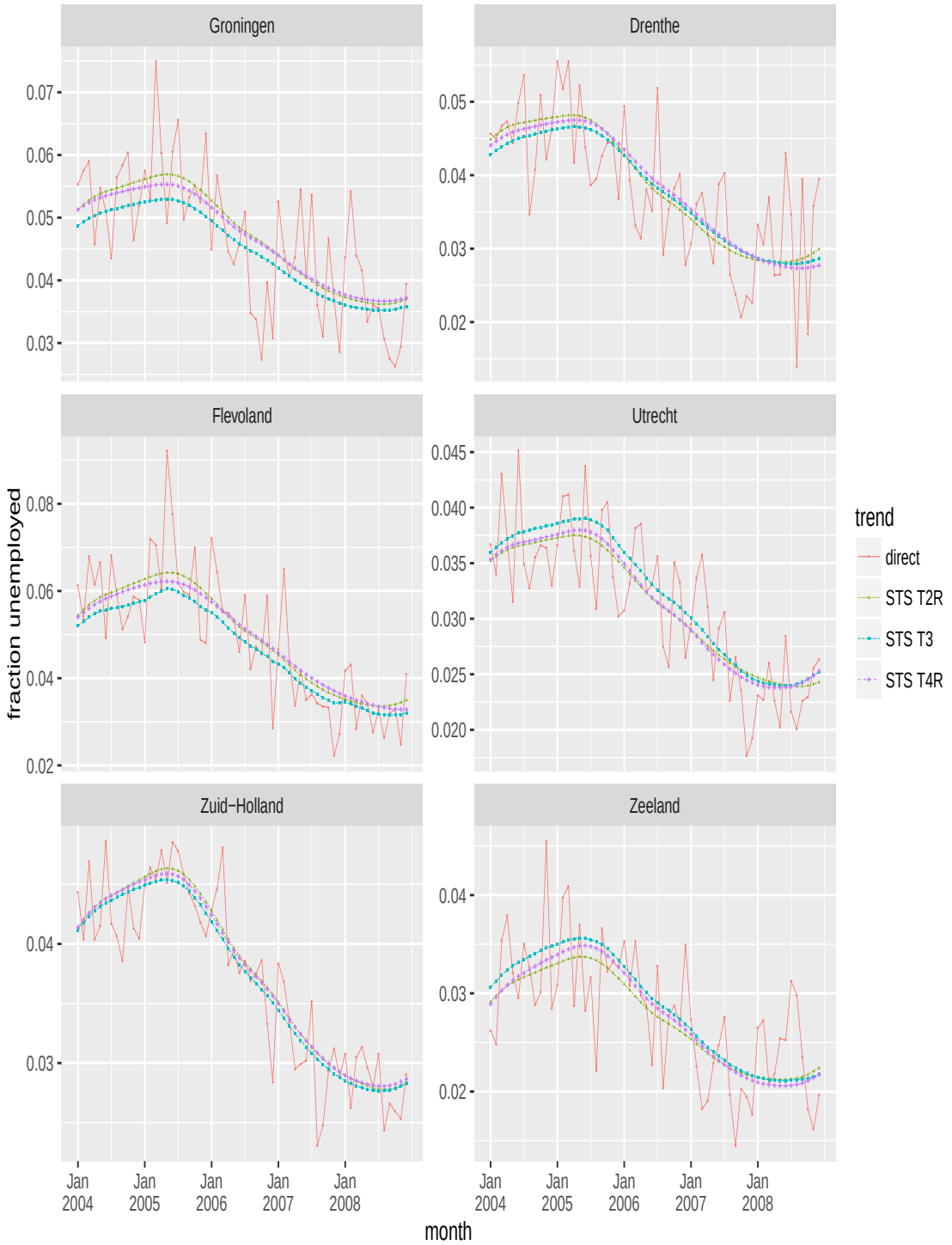

Figure 15 Comparison of estimates.

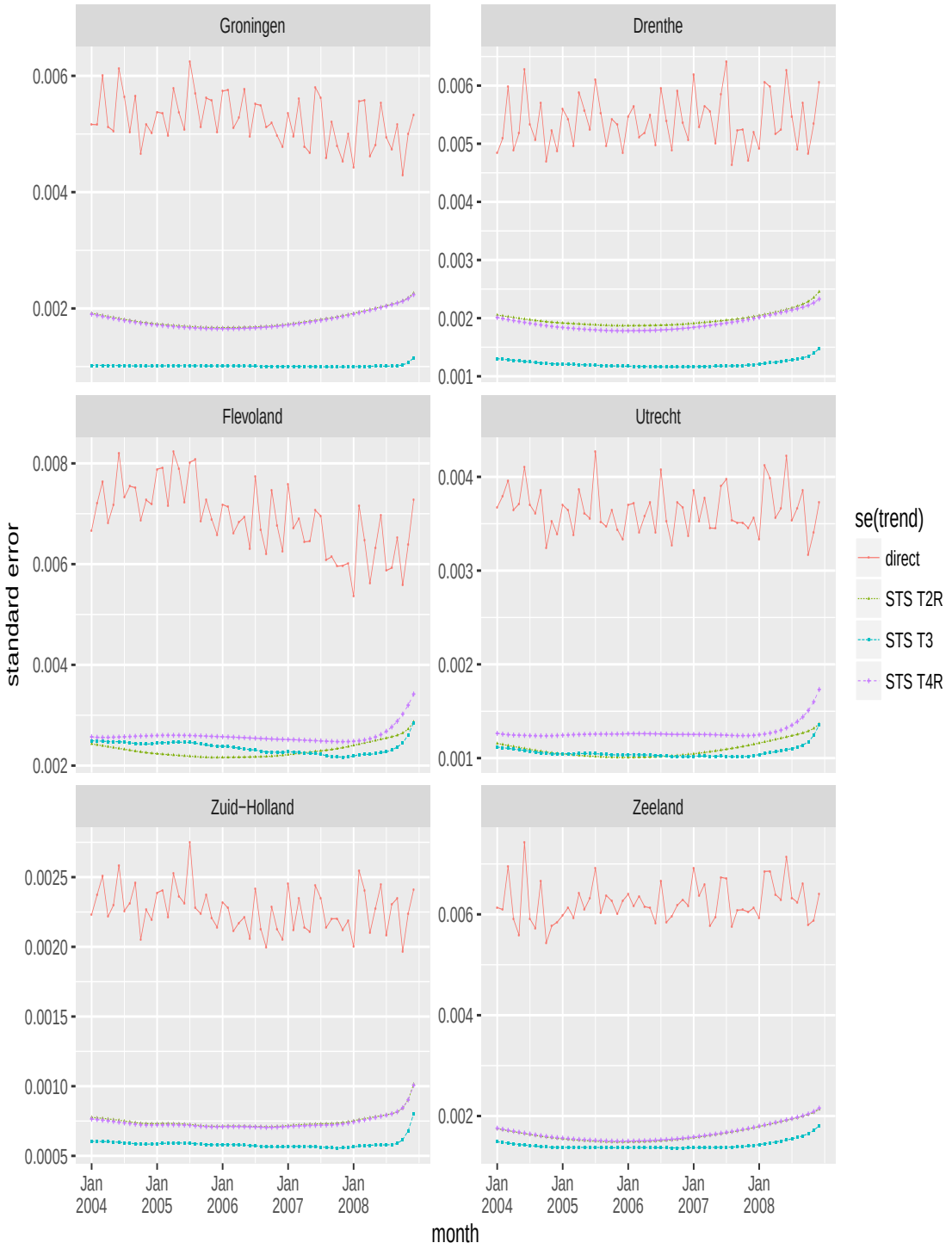

Figure 16 Comparison of standard errors.

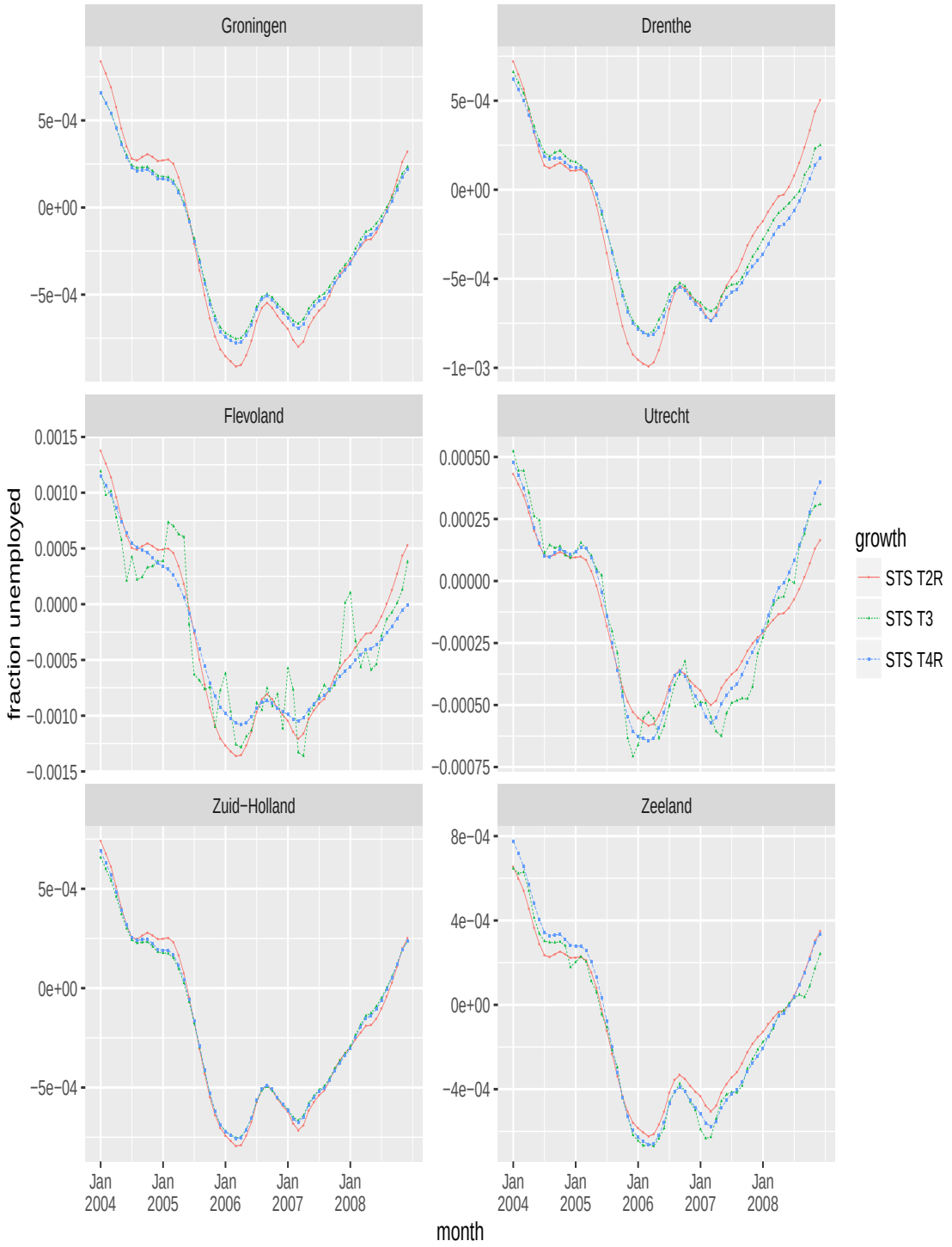

Figure 17 Comparison of estimates.

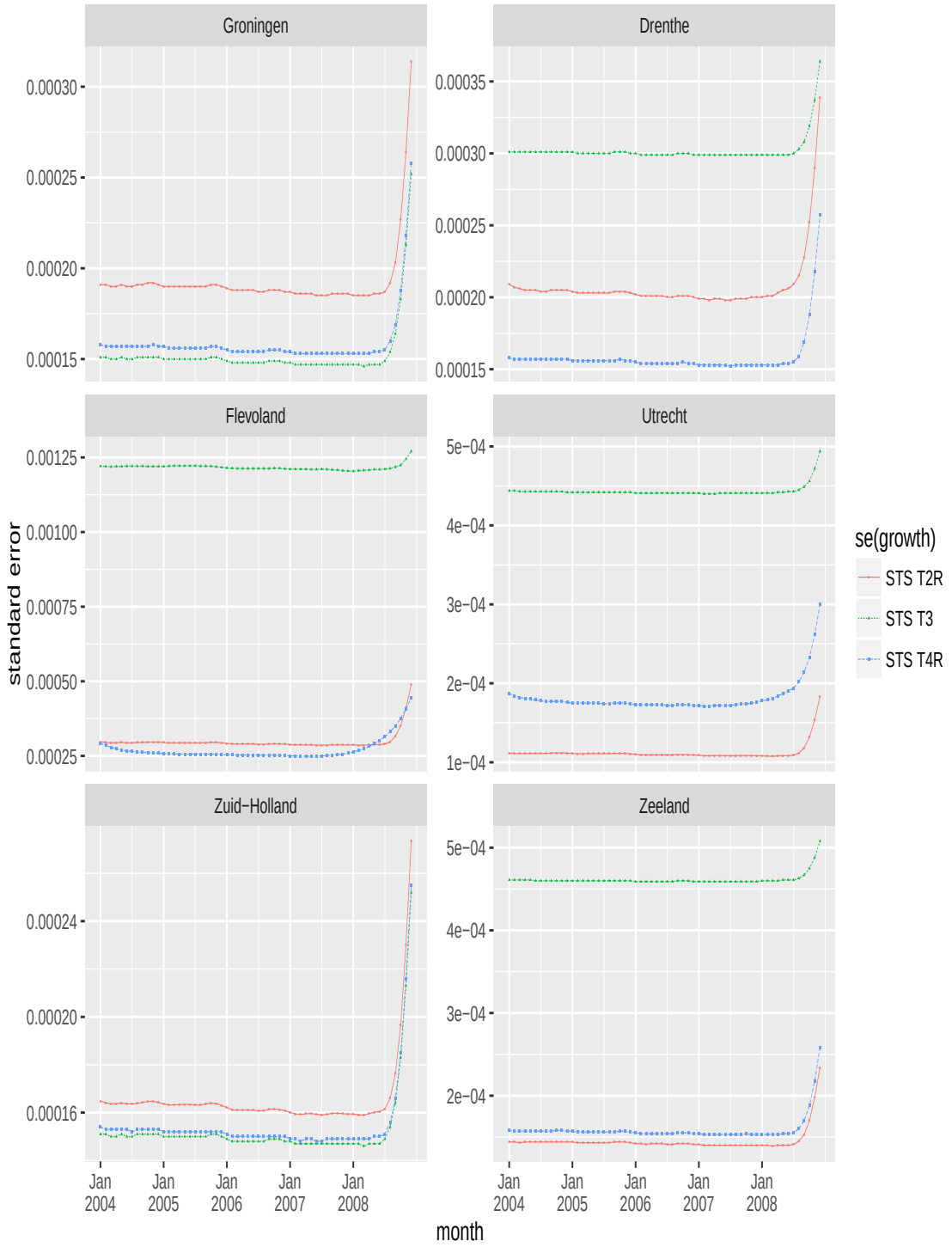

Figure 18 Comparison of standard errors.

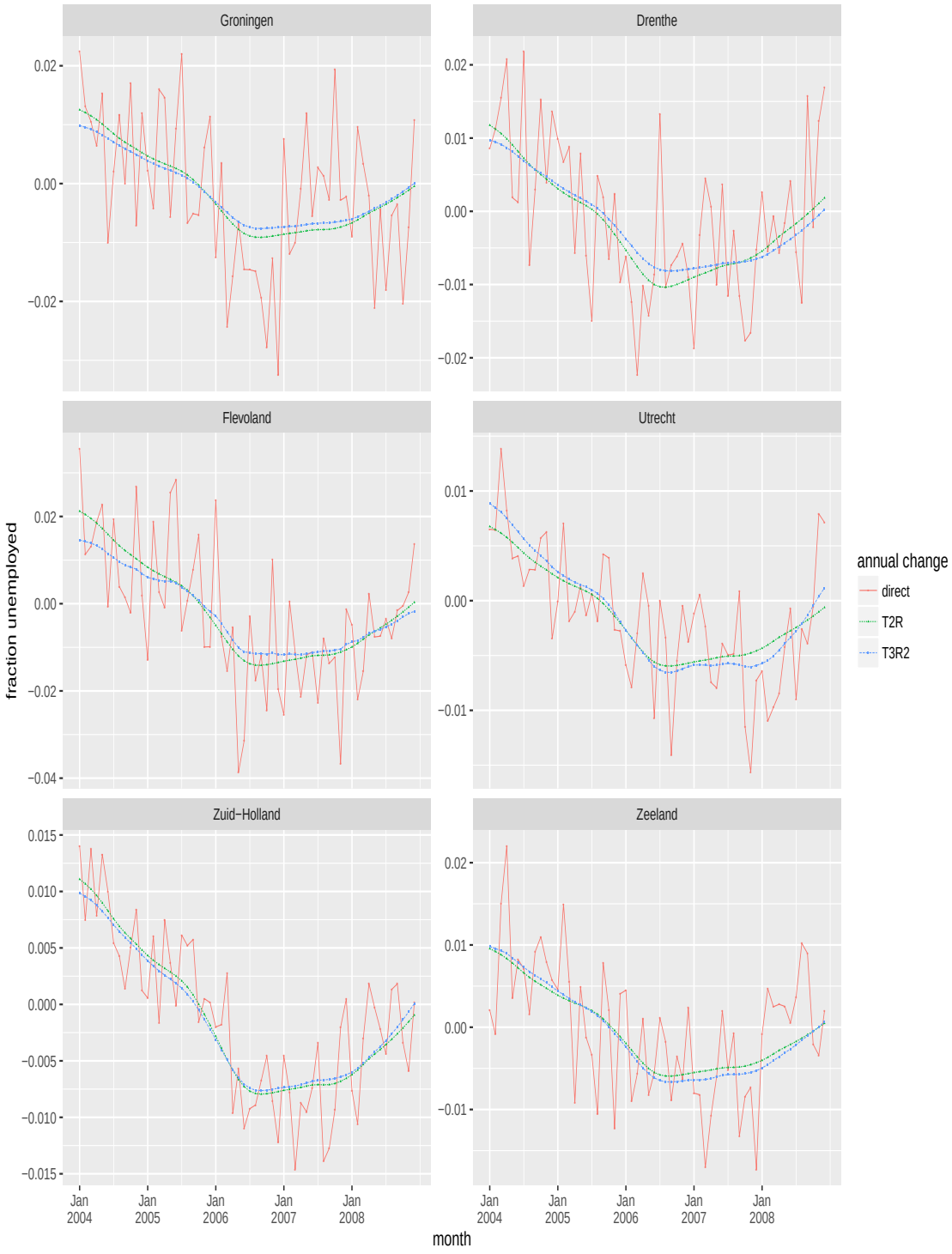

Figure 19 Comparison of estimates for annual change.

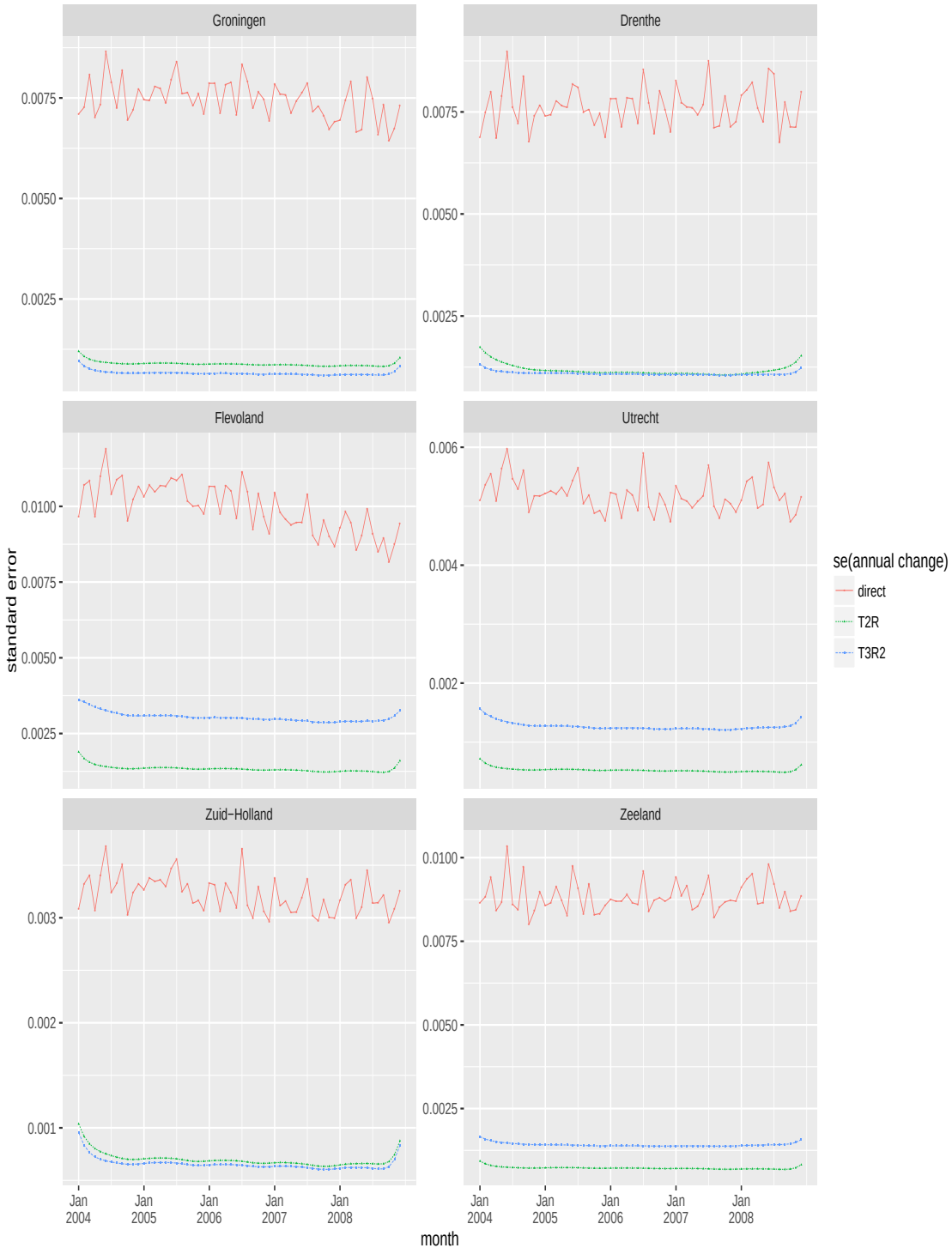

Figure 20 Comparison of standard errors for annual change.

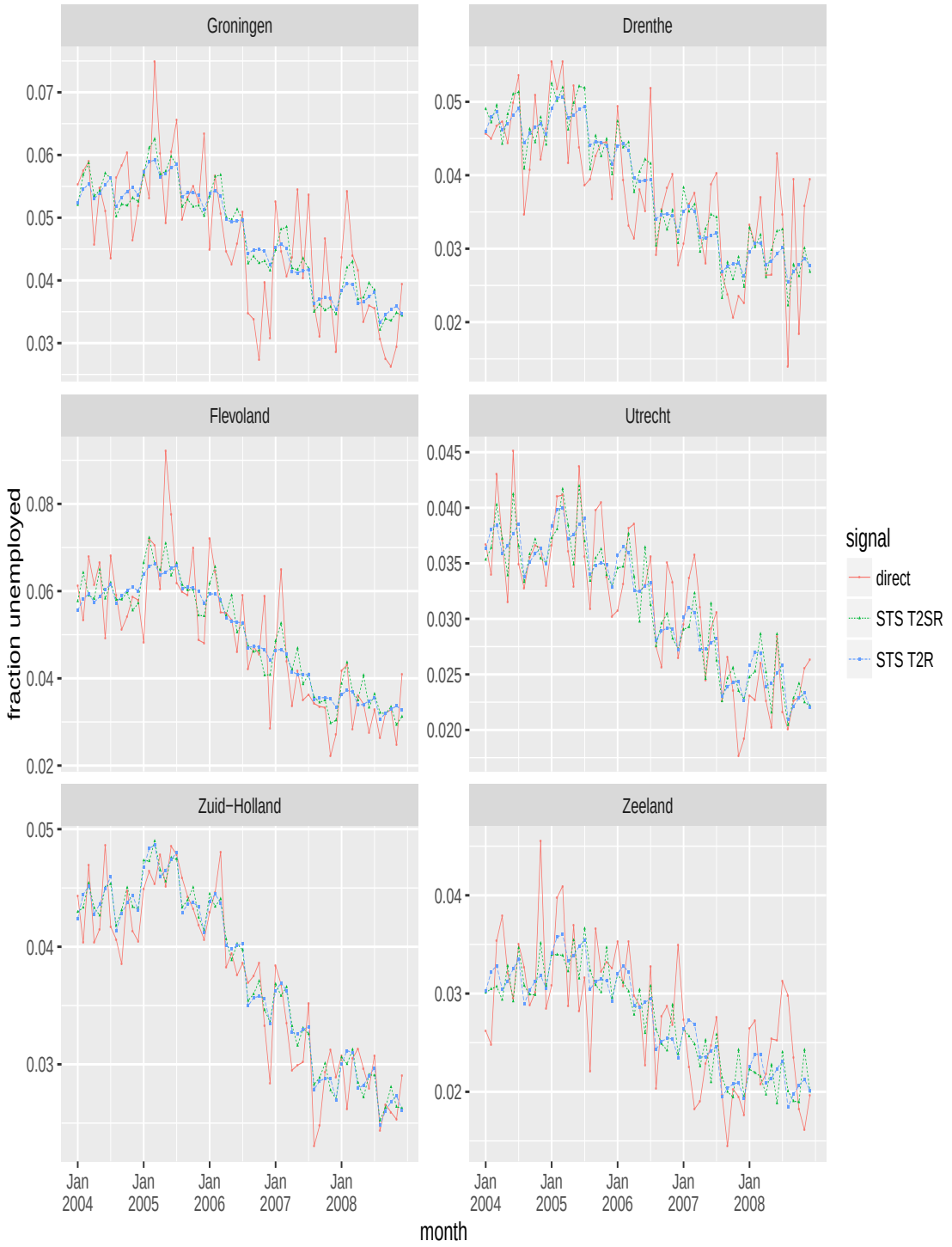

Figure 21 Comparison of estimates.

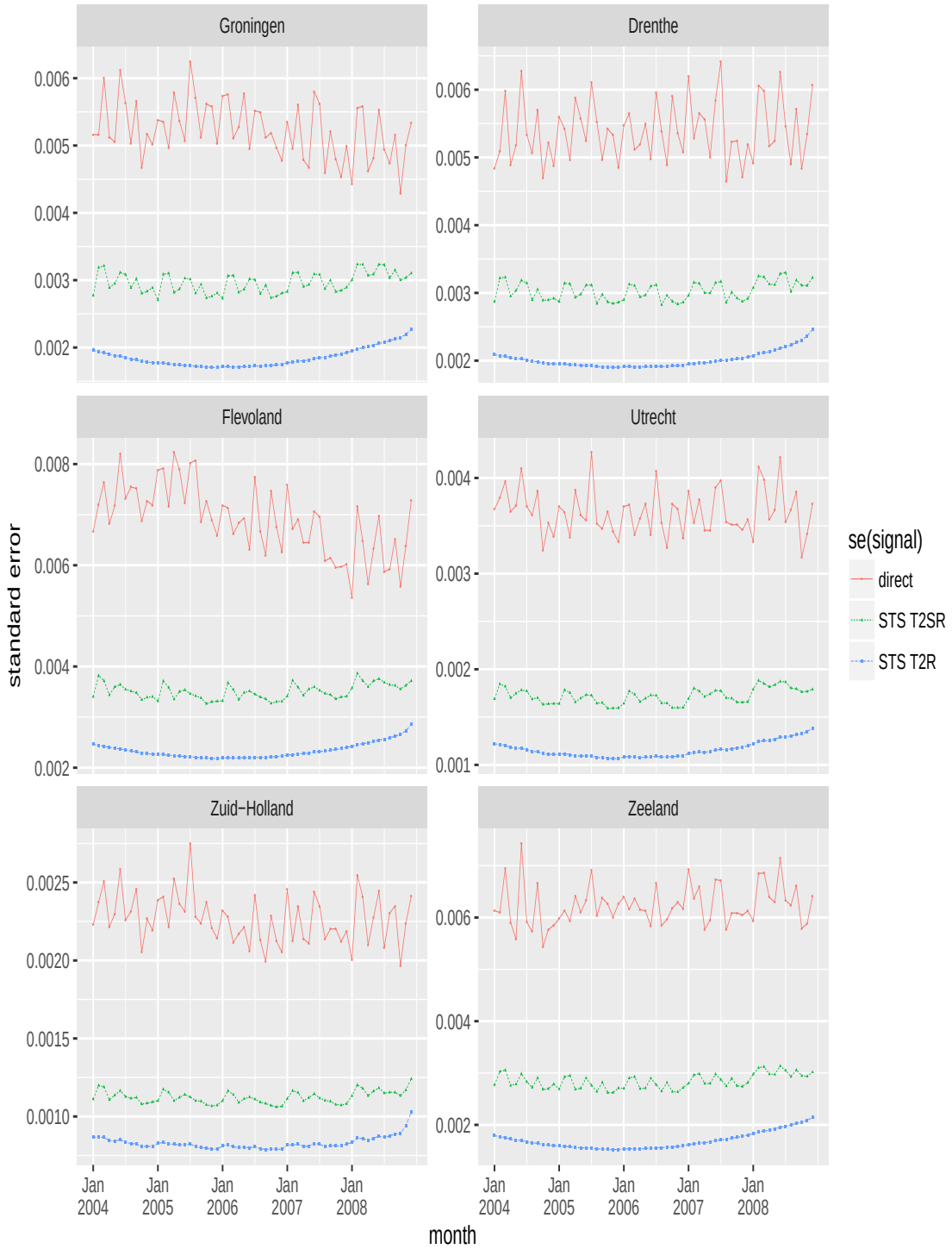

Figure 22 Comparison of standard errors.

5.2 Results multilevel models

The ten models T1SR to T4R on page 20 fitted using STM have also been fitted using the Bayesian multilevel approach. In II the multilevel models are specificied in terms of the fixed and random effect components included. One difference between the two approaches is that the Bayesian approach accounts for uncertainty in the hyperparameters by considering their posterior distributions, implying that variance parameters do not actually become zero, as frequently happens for the maximum likelihood estimates in the STM approach. For comparison purposes, however, effects absent from the STM model due to zero maximum likelihood estimates have also been suppressed in the corresponding multilevel models. In addition to these ten models we consider one more model with extra terms including a dynamic RGB component as well as a white noise term.

Differences between STM and multilevel estimates based on the ten models considered can arise because of

- the different estimation methods, maximum likelihood versus MCMC
- the different modeling of survey errors. In the multilevel models the survey errors' covariance matrix is taken to be $\Sigma = \bigoplus_{i=1}^{m_A} \lambda_i \Phi_i$ with Φ_i the covariance matrix of estimated design variances for the initial estimates for province i, and λ_i scaling factors, one for each province. In the STM models the survey errors are allowed to depend on more parameters, as shown in equation (16), though eventually an AR(1) model is used to approximate these dependencies.
- the slightly different parameterizations of the trend components. For the trend in model T3, for example, the province of Groningen is singled out by the STM model used, because no local level component is added for that province.

The estimates and, to a lesser extent, the standard errors based on the multilevel models are quite similar to the STM results for the corresponding models, as can be seen from figures 23-34, showing the estimates and standard errors for the signals, trends, and month-to-month differences. In order to save space these results are only presented for two models, namely T2R and T3R2, but the qualitative differences between STM and multilevel results are quite consistent over all models.

Particularly the estimates of the signal are very similar for STM and multilevel models (Figures 23 and 29). The small differences are due to slightly more flexible trends in the estimated multilevel models. Larger differences can be seen in the standard errors of the signal (Figures 24 and 30): the multilevel models yield almost always larger standard errors for provinces with high unemployment levels (left column of the figures), whereas for provinces with smaller unemployment levels (right column of the figures) the differences are less pronounced and for some models the multilevel standard errors are smaller than the STM ones (for example for model T1SR, results not shown).

Concerning the trends, the differences between STM and multilevel estimates are again quite small (Figures 25 and 31), although it can be seen that the multilevel estimates are slightly more flexible. The standard errors for the trends are mostly larger for the multilevel model (Figures 26 and 32), especially for provinces with high unemployment. Also, the multilevel standard errors show more variation over time.

The somewhat larger flexibility of the multilevel trends is more clearly visible in the figures comparing the month-to-month differences (Figures 27 and especially 33). The standard errors of these differences are for all models notably larger in the multilevel case, see Figures 28 and 34. In some cases the difference is larger than a factor two. The multilevel standard errors are more variable over time as well. We have checked that the larger variation over time of the multilevel standard errors is not an artifact of Monte Carlo simulation error. Model T3R2 has also been estimated using 9000 instead of 1200 draws, and the results remained approximately the same. Only the fluctuations of the standard errors for month-to-month changes for unemployment in Flevoland became visibly smaller, but not much (not shown).

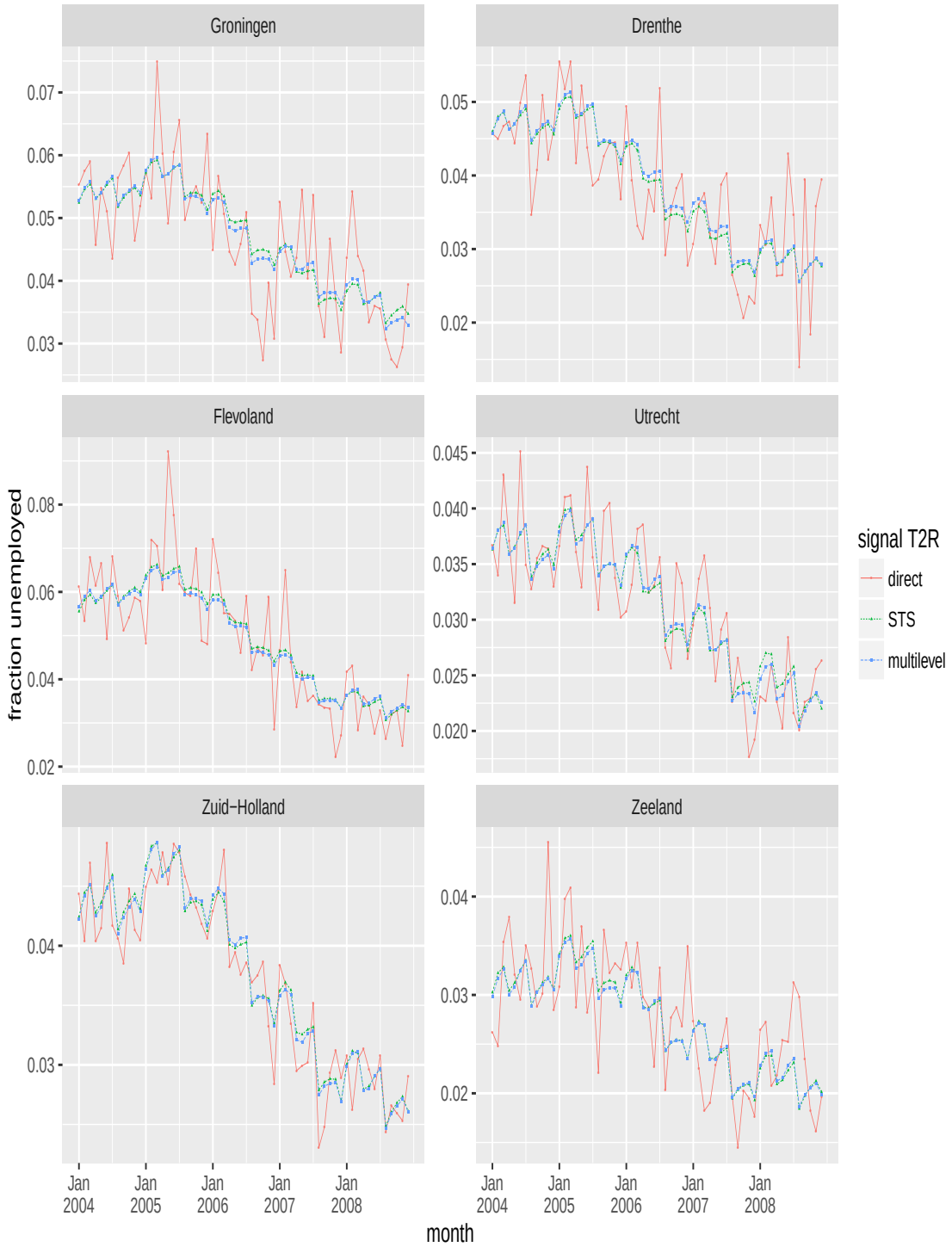

Figure 23 Comparison of estimates.

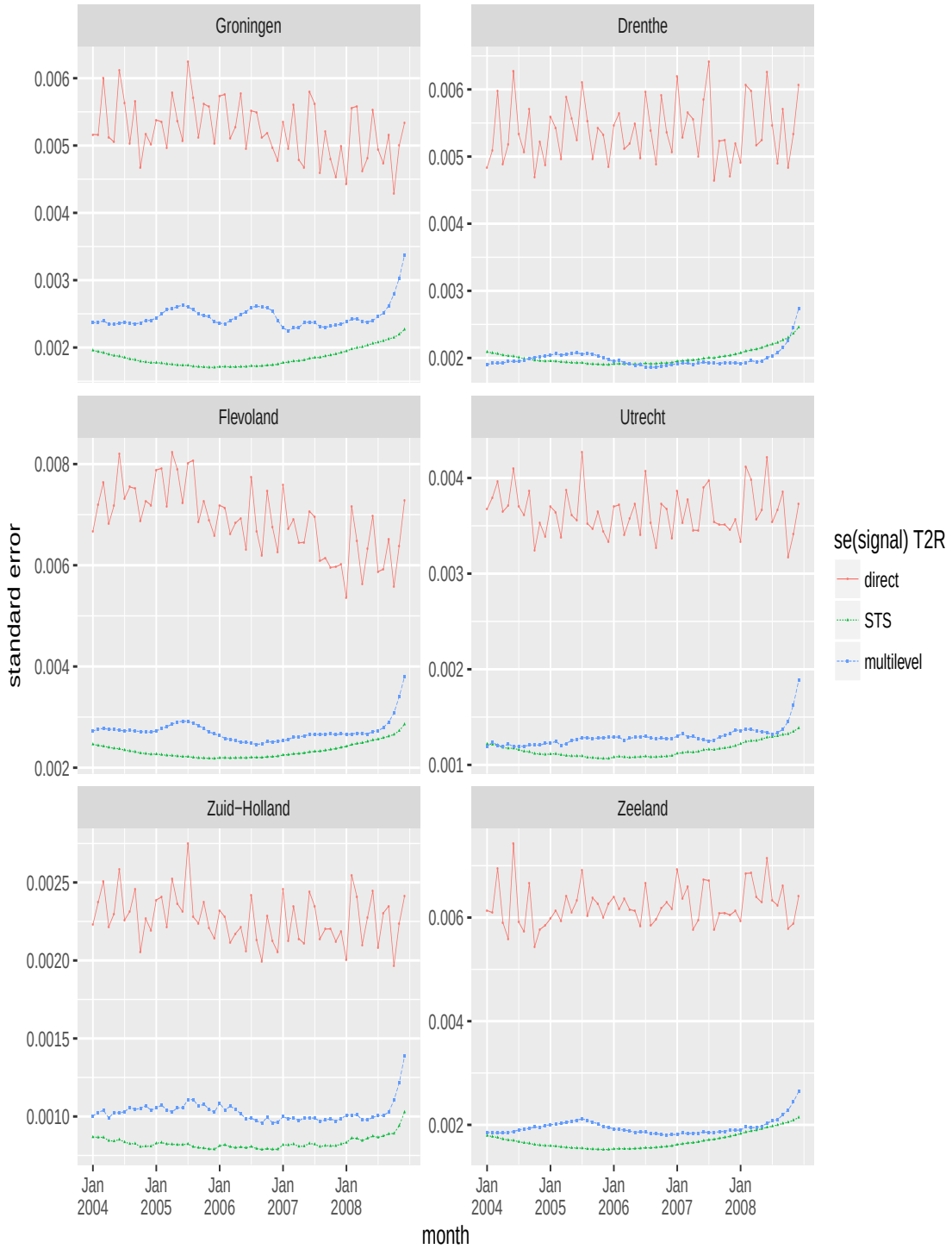

Figure 24 Comparison of standard errors.

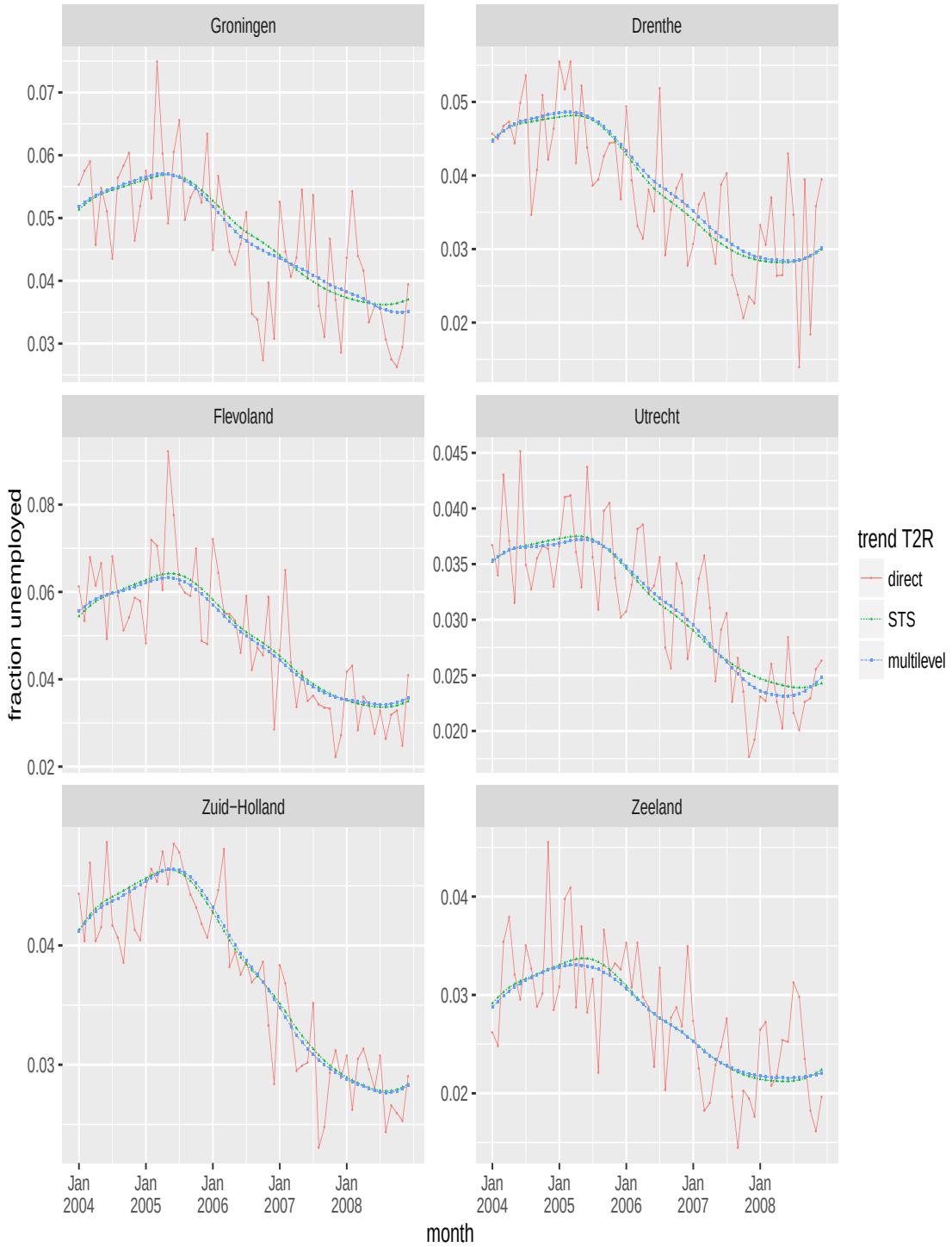

Figure 25 Comparison of estimates.

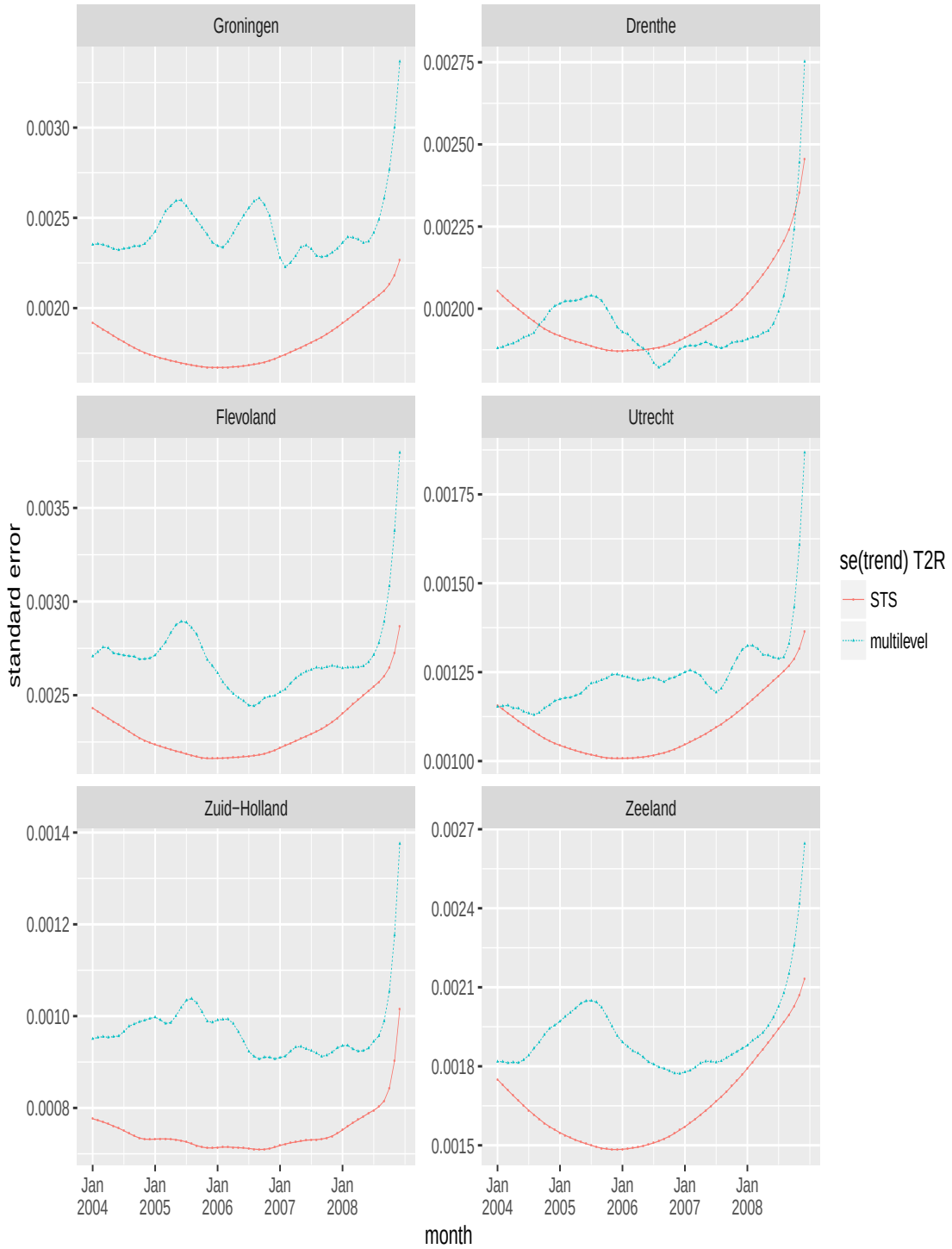

Figure 26 Comparison of standard errors.

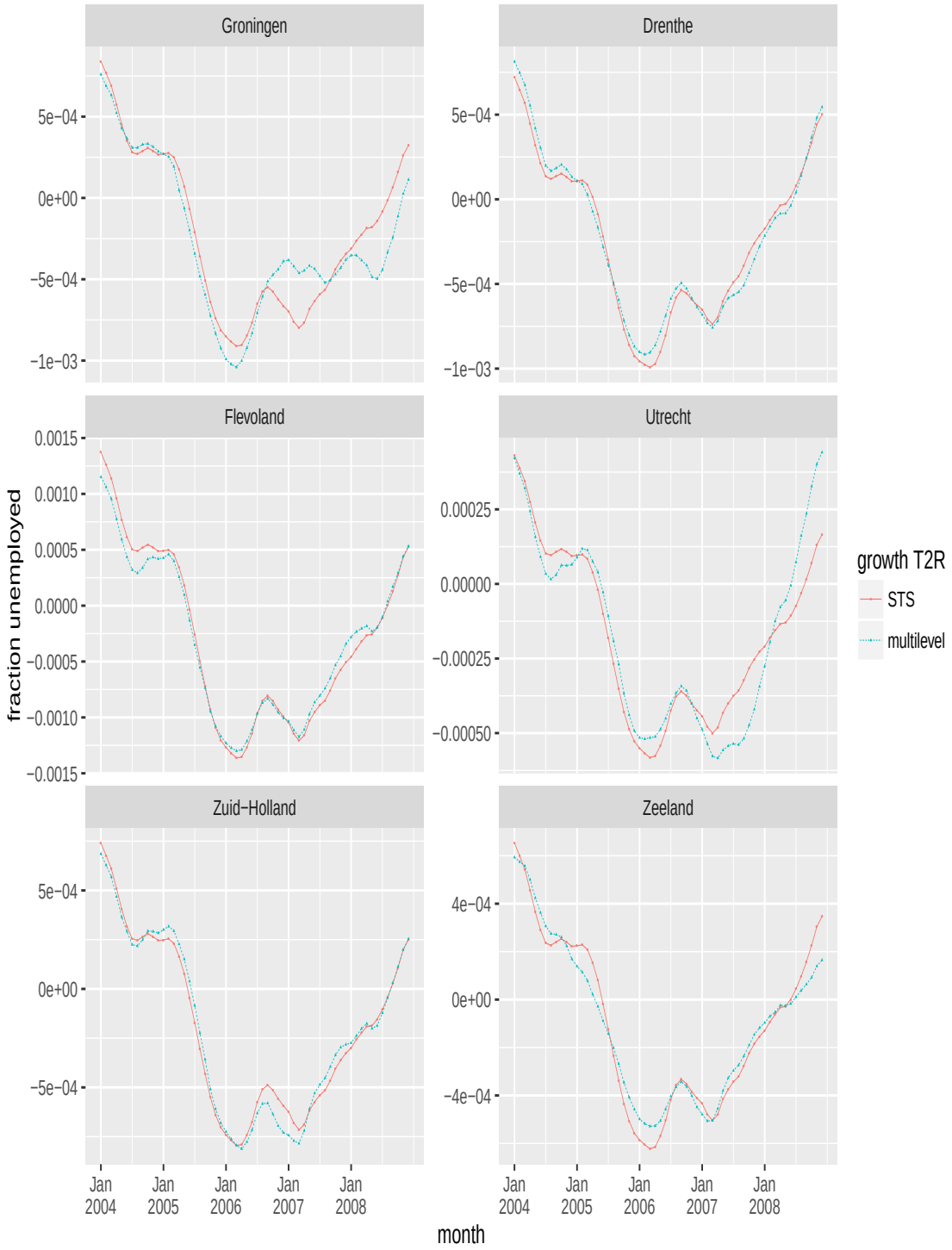

Figure 27 Comparison of estimates.

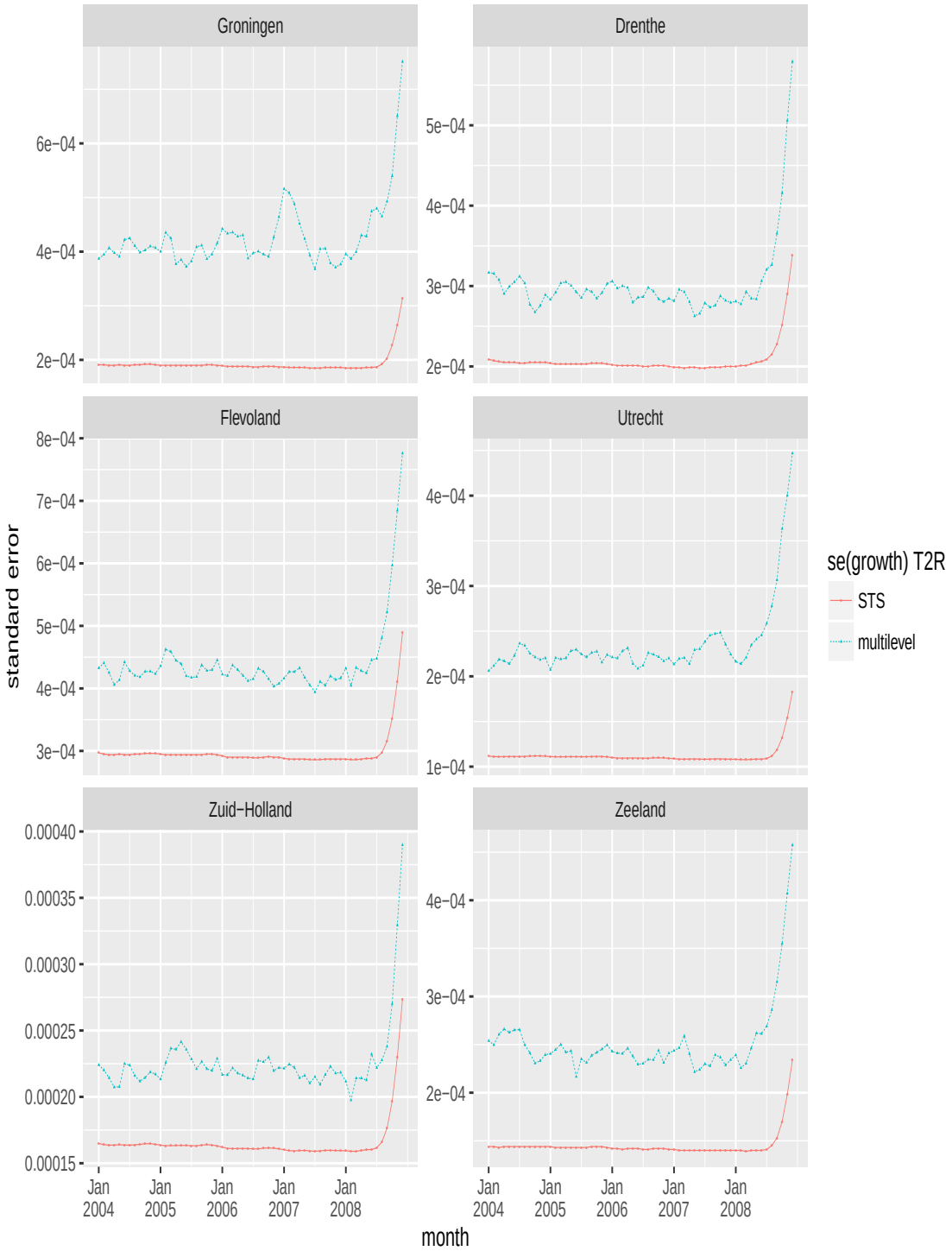

Figure 28 Comparison of standard errors.

Figure 29 Comparison of estimates.

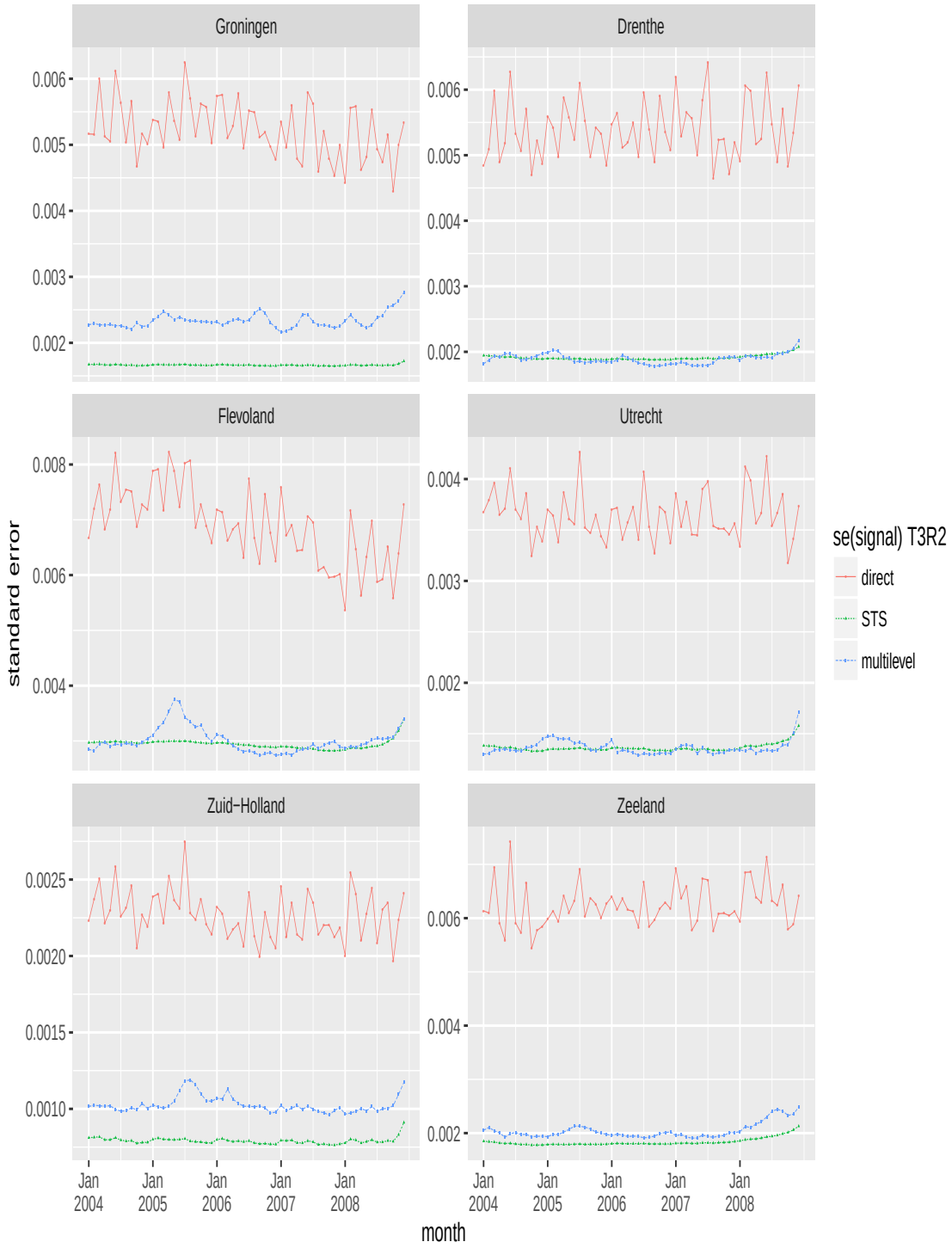

Figure 30 Comparison of standard errors.

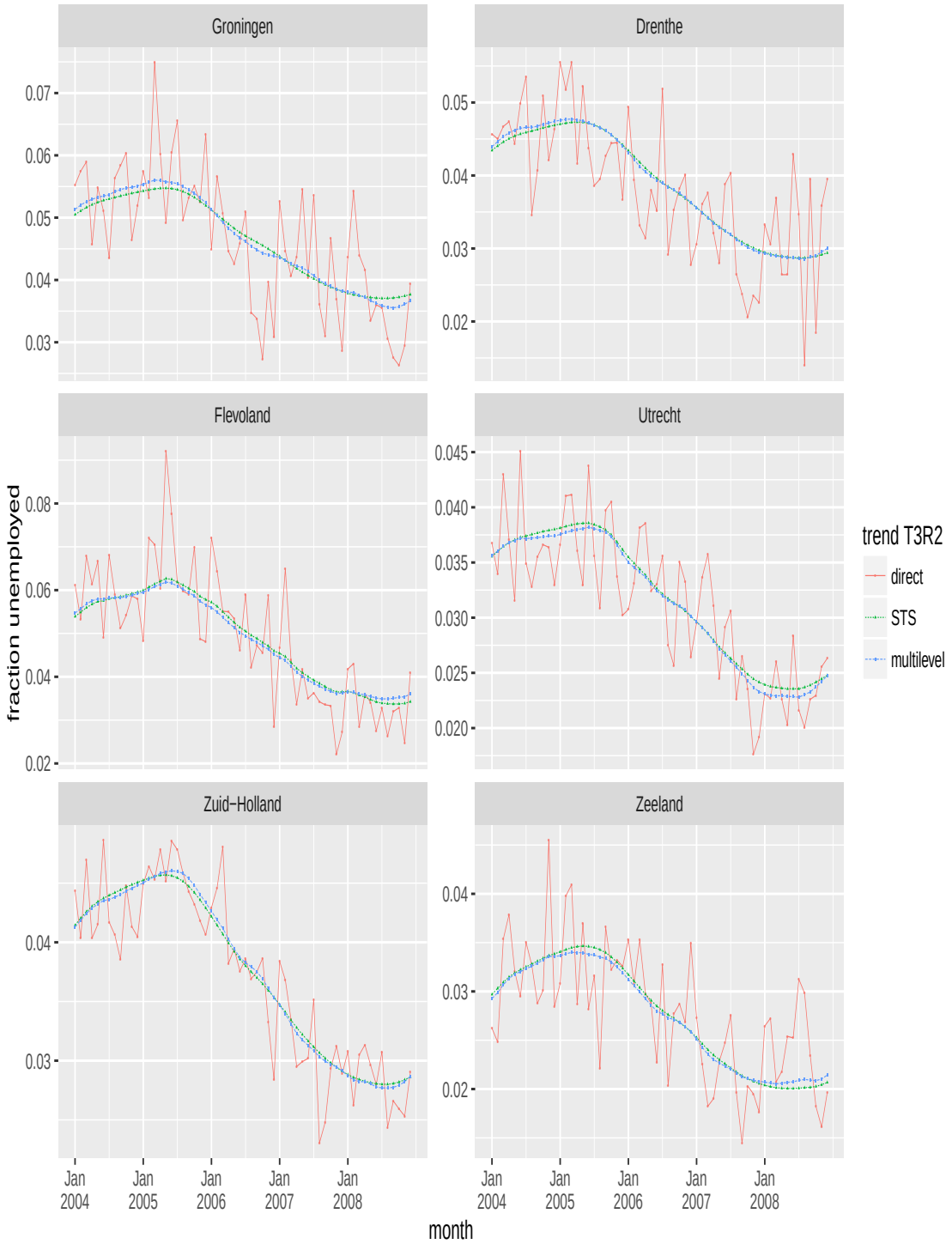

Figure 31 Comparison of estimates.

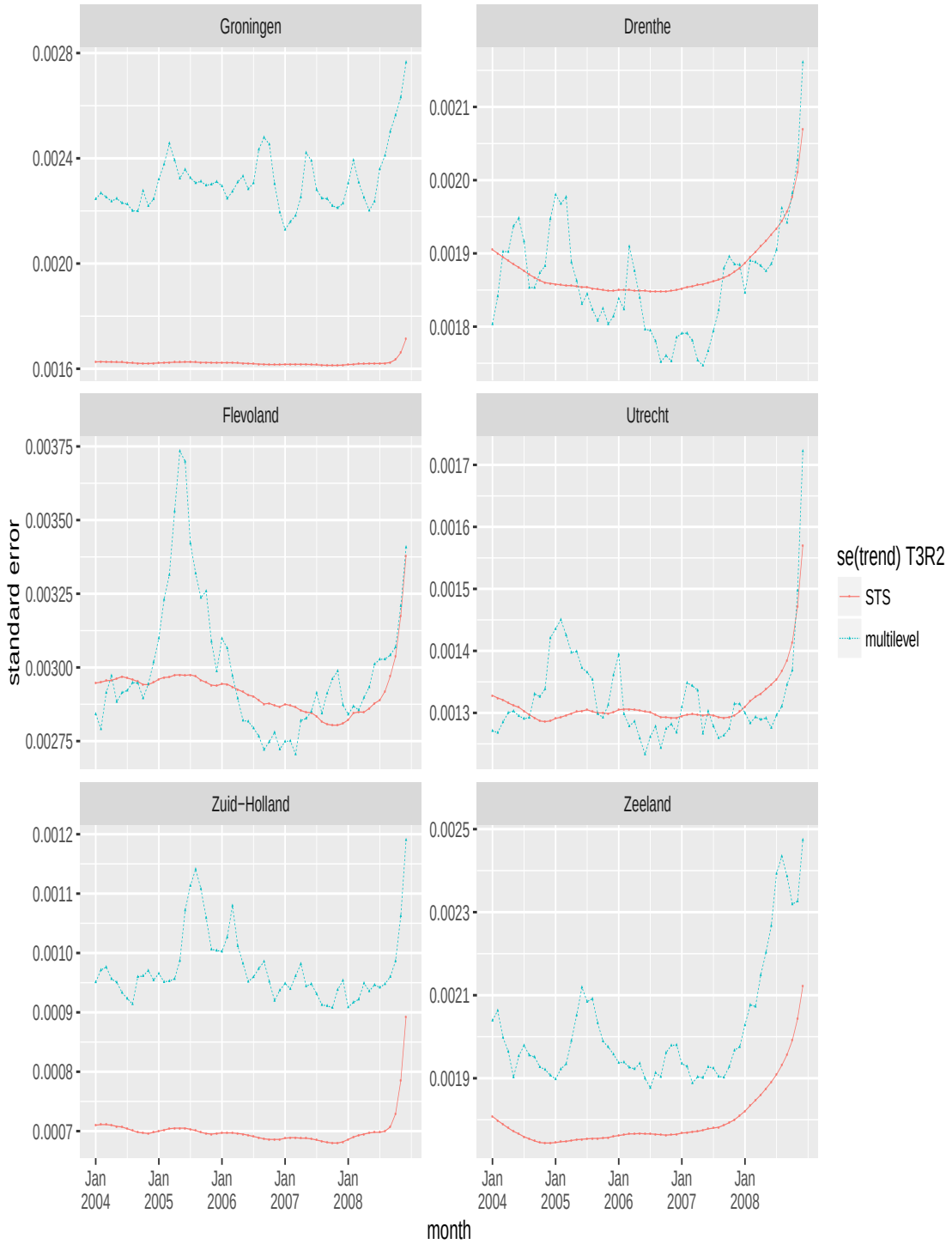

Figure 32 Comparison of standard errors.

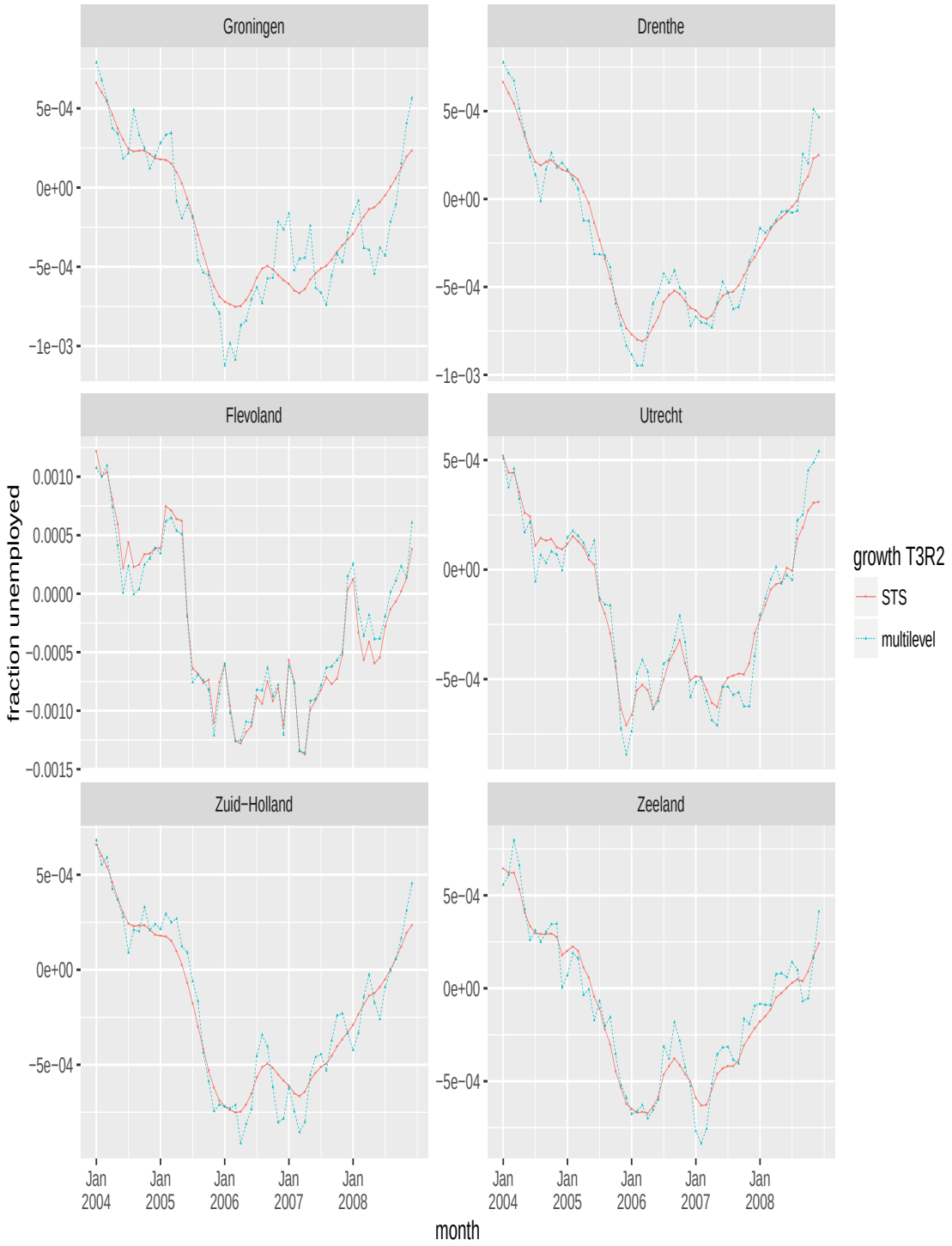

Figure 33 Comparison of estimates.

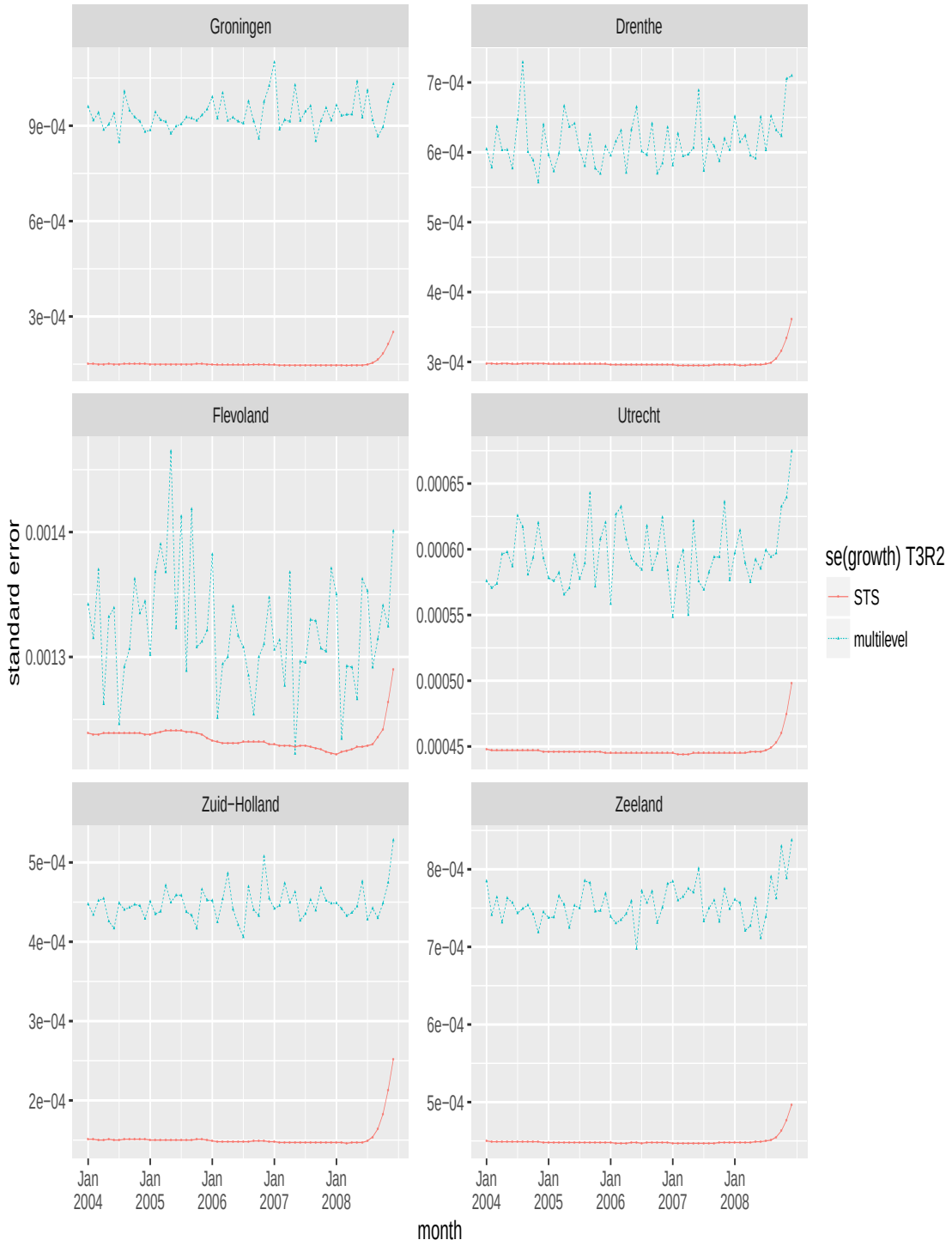

Figure 34 Comparison of standard errors.

Trend component	T1SR	T2SR	T3SR	T4SR
Overall			1.9e-4	2.0e-4
Grn	3.6e-4	2.9e-4	7.8e-4	1.4e-4
Frsl	2.7e-4	2.1e-4	5.7e-4	6.2e-5
Drn	3.1e-4	2.6e-4	5.1e-4	9.1e-5
Ovr	1.8e-4	1.6e-4	2.5e-4	4.6e-5
Flv	4.1e-4	3.4e-4	1.3e-3	2.2e-4
Gld	1.9e-4	1.7e-4	2.9e-4	3.9e-5
Utr	2.2e-4	1.6e-4	5.2e-4	1.1e-4
N-H	2.2e-4	1.9e-4	3.8e-4	3.6e-5
Z-H	2.0e-4	1.8e-4	3.8e-4	5.3e-5
Zln	2.0e-4	1.6e-4	6.3e-4	6.5e-5
N-B	1.8e-4	1.5e-4	3.9e-4	5.7e-5
Lmb	1.8e-4	1.6e-4	2.4e-4	4.2e-5

Table 8 Posterior means of standard errors for the multilevel models' trend components.

The larger flexibility of the multilevel model trends is most likely due to the relatively large uncertainty about the variance parameters for the trend, which is accounted for in the Bayesian multilevel approach but ignored in the maximum likelihood STM approach. The posterior distributions for the trend variance parameters are also somewhat right-skewed. The posterior means for the standard deviations listed in Table 8 are always larger than the maximum likelihood estimates for the corresponding STM hyperparameters (Table 2). For the models with trend type 2, i.e. with a fully parametrized covariance matrix over provinces, the multilevel models show positive correlations among the provinces, as do the STM maximum likelihood estimates, but the latter are much more concentrated near 1 (Table 3), whereas the posterior means for correlations in the corresponding multilevel model T2SR are all between 0.45 and 0.8.

The standard errors for the quantities of interest under the multilevel model are generally larger than those under the STM models. This also is most likely due to the fact that the hyperparameter uncertainty is taken into account by the Bayesian multilevel approach.

Table 9 contains values of the DIC model selection criterion (Spiegelhalter et al., 2002), the associated effective number of model parameters, and the posterior mean of the log-likelihood. The parsimonious model T3 is selected as the most favourable model by the DIC criterion. So in this case the DIC criterion selects the same model, as the AIC and BIC criteria for the STM models. An advantage of DIC is that it uses an effective number of model parameters depending on the size of random effects, instead of just the number of model parameters used in AIC/BIC.

Table 10 shows the MRB measures defined in equation (28) for the estimates based on the multilevel models. The results are quite similar to those based on the STM model, see Table 5. Again, the parsimonious model T3 selected as best by DIC comes with larger average bias over time for the provinces Groningen and Flevoland, which have the highest rates of unemployment. Model T3R2 has much smaller average biases for Groningen and Flevoland compared to model T3 (but slightly larger bias for Gelderland). Since its DIC value is not that much higher than for model T3, model T3R2 seems to be a good compromise between models

	DIC	p_DIC	mean llh
T1SR	-29054	255	14655
T2SR	-29076	235	14656
T2S	-29129	196	14662
T2R	-29164	118	14641
T3SR	-29081	242	14662
T3R2	-29217	94	14655
T3R	-29174	126	14650
T3	-29230	82	14656
T4SR	-29084	228	14656
T4R	-29170	109	14640

Table 9 DIC, effective number of model parameters and posterior mean of log-likelihood.

T3 and T3R, being more parsimonious than T3R and respecting provincial differences better than model T3.

	Grn	Frs	Drn	Ovr	Flv	Gld	Utr	N-H	Z-H	Zln	N-B	Lmb
T1SR	1.0	0.7	3.0	-0.5	0.1	3.1	-0.1	0.5	1.5	-2.1	0.1	2.6
T2SR	0.8	0.8	2.9	-0.5	0.3	3.1	-0.2	0.5	1.5	-1.8	0.1	2.6
T2S	-2.9	2.4	0.8	1.2	-3.8	2.5	1.5	0.7	0.2	2.0	1.5	1.2
T2R	0.7	0.5	2.7	-0.6	-0.0	3.1	-0.2	0.4	1.5	-1.8	0.4	2.5
T3SR	0.8	0.5	2.7	-0.4	-0.3	3.1	-0.0	0.3	1.4	-2.1	0.4	2.5
T3R	0.5	0.6	2.5	-0.4	-0.4	3.0	0.0	0.4	1.5	-1.9	0.5	2.4
T3R2	0.3	0.8	2.4	-0.7	-0.6	3.1	0.1	0.5	1.3	-1.6	0.6	2.0
T3	-3.4	2.0	0.3	1.2	-4.7	2.4	1.8	0.5	0.3	2.3	1.8	1.2
T4SR	0.8	0.8	2.9	-0.4	-0.1	3.1	-0.1	0.5	1.4	-2.0	0.2	2.6
T4R	0.8	0.6	2.7	-0.5	-0.4	3.2	-0.2	0.4	1.6	-1.7	0.4	2.5

Table 10 Mean Relative Bias averaged (28) over time (%), per province for multilevel models

Table 11 shows the RRSE defined in (29) between the multilevel time-series model (signal) estimates and the direct estimates. Table 12 contains the averages of standard errors for signal, trend and month-to-month differences in the trend. The average is taken over all months and provinces. The results are again similar to the STM results, see Tables 6 and 7, although especially the standard errors of month-to-month changes are larger under the multilevel models.

5.2.1 A model with additional random effects

Many of the variance parameters of the general STM model discussed in Section 4.1 were estimated at zero by maximum likelihood. This means that the trigonometric seasonal reduces to fixed effect dummies for the seasons, that a white noise term is absent, and that RGB corrections are time invariant. In the multilevel models considered so far, we have removed these same effects as well, in order to get more comparable results. We now consider an extended multilevel model where some of these effects as well as other random effects are included. Unlike maximum likelihood estimation, the Bayesian approach accounts for the uncertainty about the variance parameters, so that once random effects terms are included their effect will never be zero, although it can still be small.

	Grn	Frs	Drn	Ovr	Flv	Gld	Utr	N-H	Z-H	Zln	N-B	Lmb
T1SR	33	45	45	46	41	47	51	45	41	55	48	45
T2SR	35	47	46	50	44	49	53	47	43	56	51	47
T2S	40	50	51	53	48	53	56	51	45	59	53	52
T2R	53	62	62	62	61	60	63	59	55	69	60	61
T3SR	36	47	49	50	41	50	51	47	42	55	48	50
T3R	55	62	65	65	57	57	62	60	55	68	58	64
T3R2	55	62	65	65	56	57	62	60	55	67	59	64
T3	62	68	71	71	61	61	66	64	59	73	62	71
T4SR	37	49	48	50	44	51	52	49	43	57	50	49
T4R	56	65	65	63	61	63	63	63	57	71	60	63

Table 11 Reductions in standard error compared to those of the direct estimates (%), per province, for the multilevel time-series models

	se(signal)	se(trend)	se(growth)
direct	100		
T1SR	55	41	8
T2SR	52	37	6
T2S	49	33	7
T2R	39	38	6
T3SR	53	38	15
T3R	39	38	15
T3R2	39	38	15
T3	34	32	15
T4SR	51	36	6
T4R	37	36	6

Table 12 Means of standard errors over all months and provinces relative to the mean of the direct estimator's standard errors (%) for the multilevel time-series models

To investigate, we fit an extended version of model T3R2 where the linear predictor is extended with

- a white noise term with a single variance parameter, $\epsilon_{it} \overset{\text{ind}}{\sim} N(0, \sigma_\epsilon^2)$
- the balanced dummy seasonal effect as described earlier. We only include an overall such effect, as it turned out that including one for each province tends to overfit the data.
- a dynamic RGB component as described in Section 4.1. Again, we only include an overall effect, not one for each province. The effects corresponding to the first wave are constrained to zero.
- season by province random effects, summing to zero over the seasons for each province

Note that this extended model specification includes all second order interactions among the variables month, season, wave and area, except wave by season. Model T3R2 includes wave by area fixed effects differentiating only between wave 1 and follow-up waves. It was found that including wave by area random effects instead of fixed effects had a much smaller bias reducing effect for the provinces Groningen and Flevoland.

The results from this extended model, abbreviated by T3R2e are compared to those of models T3R2 and T3SR, and shown in Figures 35-40.

From Figure 35 it is clear that the estimates of the signal based on the extended model are quite close to the estimates for model T3R2. So the additional estimated random effects are small. Figure 36 shows that for the signal standard errors the results based on the extended model T3R2e are between those of models T3R2 and T3SR, but much closer to the former for most provinces. The DIC value for model T3R2e is -29260 (with 161 effective model parameters), which is well below the DIC value for model T3R2 (Table 9). It turns out that this improvement in the DIC is almost entirely due to the dynamic RGB component. The other additional random components of model T3R2e hardly affect the DIC value. Apparently, modelling the RGB as a dynamic effect results in a markedly better fit, despite the zero maximum likelihood estimates that were found for the variance of the time-dependence of RGB effects in the STM model. Referring back to Figure 3, it is not really surprising that some time-dependence of RGB results in a better fit.

Figure 37 shows the trends for the three models. For model T3R2e we have defined the trend by the predictions based only on the trend components in the model, excluding not only all seasonal effects, but also the white noise term. The trend for model T3R2e deviates slightly from those of models T3R2 and T3SR, presumably because of the (indirect) effect of the dynamic RGB component. The standard errors of the trend for model T3R2e are qualitatively similar to those for model T3R2, although somewhat larger for province Zuid-Holland, see Figure 38. The month-to-month changes in unemployment are also similar between the three models (Figure 39). However, their standard errors are quite different, as shown in Figure 40. The extended model T3R2e has larger standard errors for Utrecht and Zeeland, but smaller for the other displayed provinces. For the provinces not shown the differences are smaller, though there is one province (Gelderland) where the standard errors based on T3SR are about 25% smaller than the standard errors based on the other two models.

Figure 35 Comparison of estimates.

Figure 36 Comparison of standard errors.

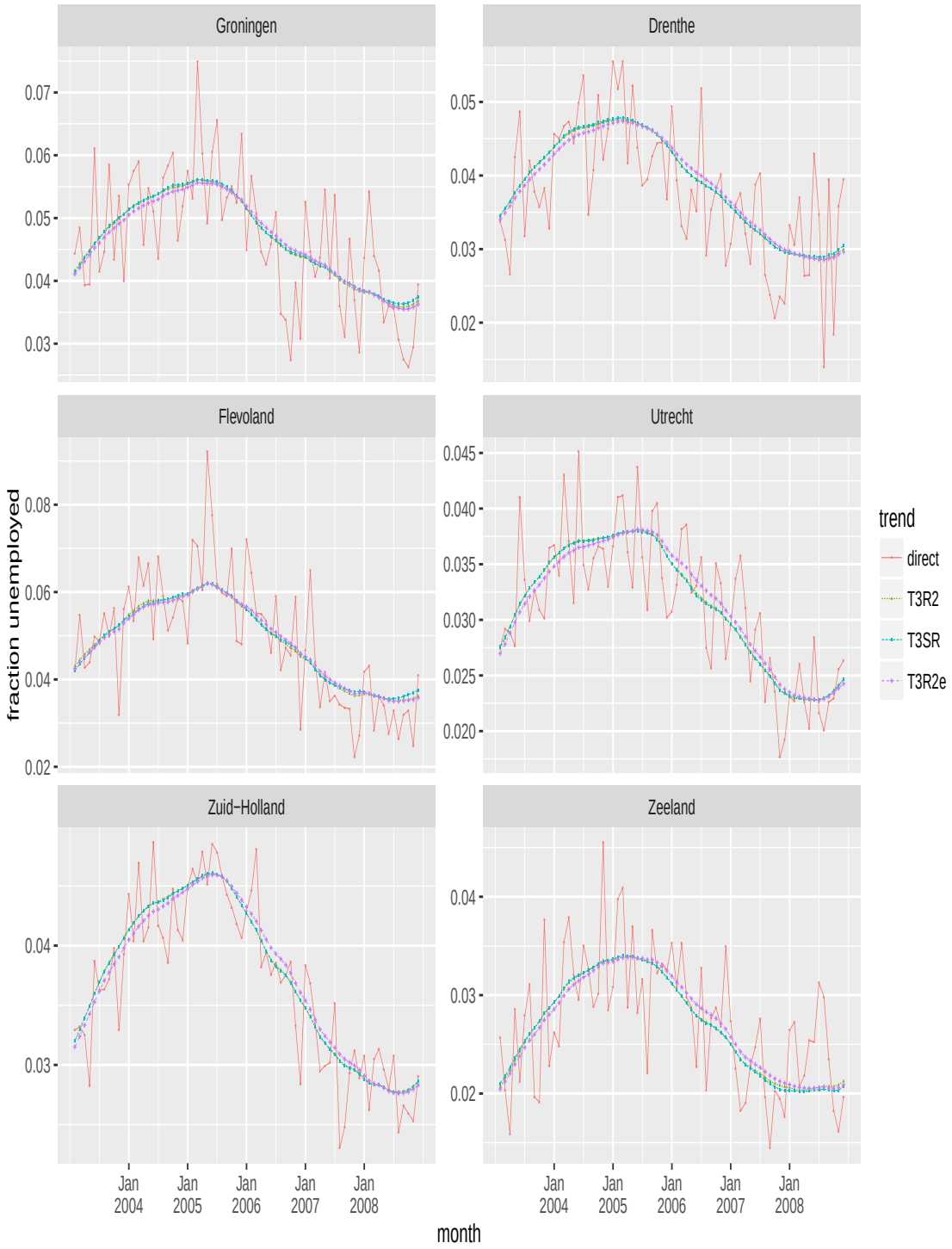

Figure 37 Comparison of estimates.

Figure 38 Comparison of standard errors.

Figure 39 Comparison of estimates.

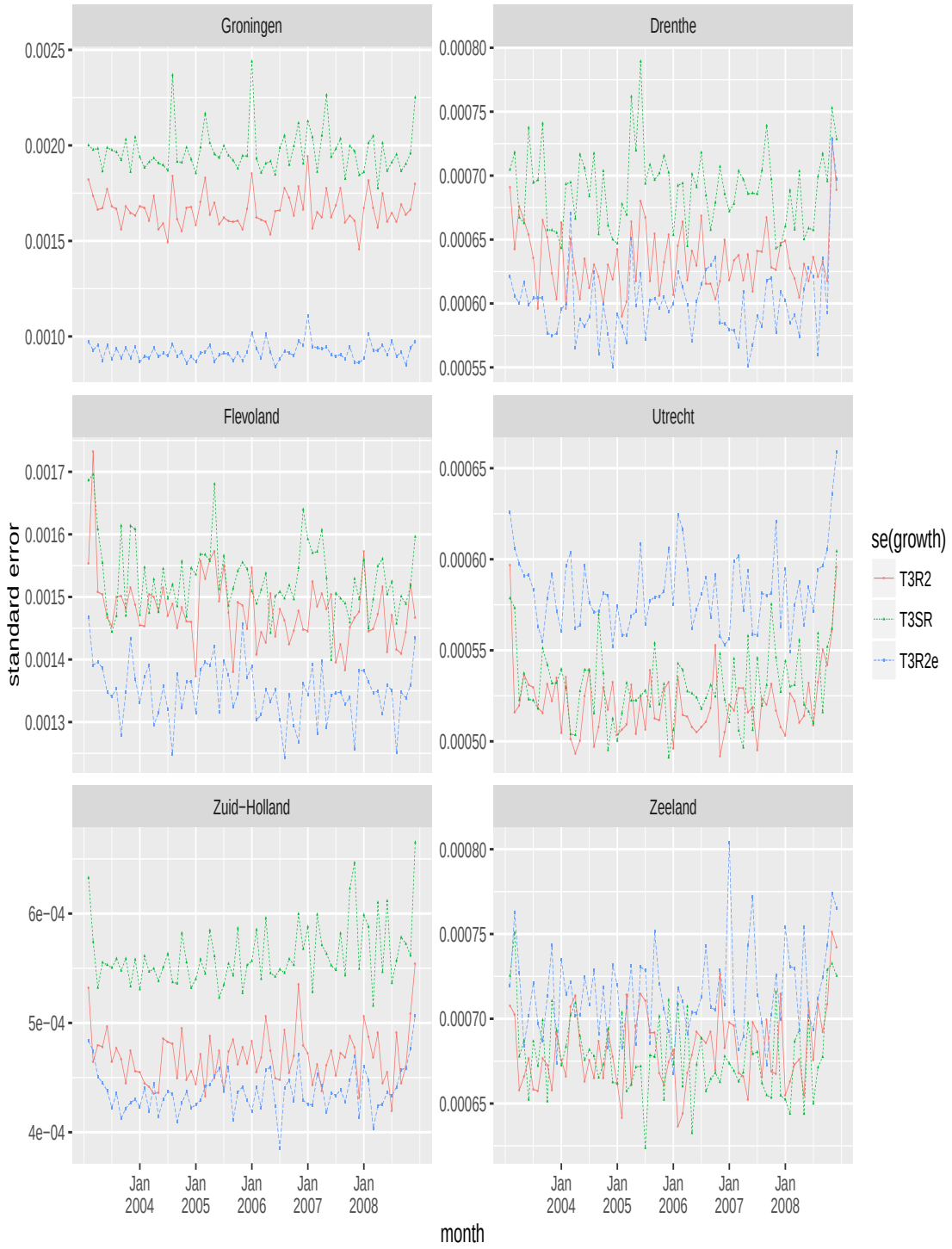

Figure 40 Comparison of standard errors.

6 Discussion

A time-series small area estimation model has been applied to a large amount of survey data, comprising 6 years of Dutch LFS data, to estimate monthly unemployment fractions for 12 provinces over this period. Estimation is carried out in two stages in order to limit the computational burden and model complexity. First, the unit-level data are aggregated to the month by province level by means of a survey regression estimator that reduces non-response bias. The Dutch LFS is a rotating panel survey with five waves differing according to their time-in-panel, and initial survey regression estimates are computed for each wave. The resulting set of initial estimates together with estimates of their standard errors and correlations are input for the time-series SAE model, which borrows strength over provinces and over time to yield improved estimates. Besides, the model accounts for rotation group bias relative to the first wave by modeling it as a measurement error, and accounts for the serial correlation in the survey errors. Two modelling approaches are applied and compared. The first one is a state-space model estimated with a Kalman filter where the unknown hyperparameters are replaced by their maximum likelihood estimates. The second one is a multilevel time-series approach. A Bayesian approach is taken, and the model is fit using a Gibbs sampler. Sparse matrices are used for fast generation of draws from the full conditional distributions, while multiplicative redundant reparameterizations are used to speed up convergence of the sampler.

The time-series models that do not account for cross-sectional correlations and borrow strength over time only, already show a major reduction of the standard errors compared to the direct estimates. A further small decrease of the standard errors is obtained by borrowing strength over space through cross-sectional correlations in the time-series models. In this paper analyses are based on smoothed estimates, which implies that time-series model estimates are based on the completely observed series. However, a regular production process will normally produce estimates for the current, latest period only. In that case the reduction of standard errors brought about by the time-series model is smaller, because strength over time is only borrowed from the past. Another great advantage of the time-series model approach concerns the estimation of change. Under the multilevel model estimates of change and their standard errors can be easily computed, especially when the model fit is in the form of an MCMC simulation. Under the state-space approach, estimates of change follow directly from the Kalman filter recursion by keeping the required state variables from the past in the state vector. The desired estimate for the change, including its standard error, follows from the contrast of the specific state variables. Time-series models are particularly appropriate for making estimates of change. Month-to-month and year-to-year change of monthly data are very stable and precise, which is a consequence of the strong positive correlation between level estimates. Nevertheless, the stability of the estimates of change strongly depends on the choice of the trend model. Local level models result in more volatile trend estimates and thus also more volatile estimates of change and naturally have a higher standard error compared to smooth trend models.

Annual estimates and year-to-year estimates for annual change can be derived as linear combinations from the monthly figures. A topic for further research is to compare these estimates obtained from the time-series modelling approach with direct estimates, which are in general also very reliable at national and provincial level.

In this paper different trend models are considered that model correlation between domains with the purpose to borrow strength over time. The most complex approach is to specify a full covariance matrix for the disturbance terms of the trend component. One way to construct parsimonious models is to take advantage of cointegration. In the case of strong correlation between domains the covariance matrix will be of reduced rank, which means that the trends of the m_A domains are driven by less than m_A common trends. In this application two common trends are sufficient to model the dynamics of the twelve provinces, resulting in a strong reduction of the number of hyperparameters required to model the cross-sectional correlations between the domains. In order to further reduce the number of state and hyperparameters alternative trend models are considered that implicitly account for cross-sectional correlations. Under this approach all domains share the same trend. Each domain has a domain specific trend to account the deviation between the overall trend and the domain specific trend. This can be seen as a simplified form of a common trend model. In this application the trends obtained under this alternative trend model results in comparable estimates for the trends and standard errors. So this approach might be a practical attractive alternative for common trend models. For example if the number of domains is large or the number of common factors is larger, then the proposed trend models are less complex compared to general common trend models.

A difference between the multilevel models and state space models is that under the former, model components are more often found to be time varying, while under the state-space approach most components, with the exception of the trend, are estimated as time invariant. This is a result of the method of model fitting. Under the frequentist approach, used to fit state-space models, maximum likelihood estimates for most of the hyperparameters are estimated zero. Under the hierarchical Bayesian approach the entire distribution of the variance parameters is simulated resulting in mean values for these hyperparameters that are always positive. A consequence of this feature is that the covariances between the trend disturbances are smaller under the hierarchical Bayesian approach, which makes it difficult to detect cointegration and construct common factor models. An advantage of the hierarchical Bayesian approach is that the standard errors of the domain predictions account for the uncertainty of replacing the unknown hyperparameters for their estimates. As a result the standard errors obtained under the hierarchical Bayesian approach of comparable models are slightly higher and more realistic compared to the state-space approach. For the state-space approach several bootstrap methods are available to account for hyperparameter uncertainty (Pfeffermann and Tiller, 2005). These procedures are not applied in this paper.

Another advantage of the multilevel approach is the possibility to use DIC or the conditional AIC that uses the effective degrees of freedom, defined as the trace of the Hat matrix that maps the observed data to the predicted values, as a penalty for model complexity. This also accounts for the size of the random effects in the multilevel models.

From a computational point of view, there are some differences between the methods too. The STM Kalman filter approach can be used online, producing new filtered estimates by updating previous predictions when data for a new month arrives. The Gibbs sampler multilevel approach used here produces estimates for the whole time-series at once. It must be re-estimated completely when data for a new month arrives. However, due to the use of sparse matrices and redundant reparameterization the multilevel approach is quite competitive computationally, see also Knorr-Held and Rue (2002). An advantage of the simultaneous multilevel estimation is that constraints over time can easily be imposed. For example,

imposing sum-to-zero constraints over time allows to include local level provincial trends for all provinces in addition to a global smooth trend with no resulting identification issues.

The time-series standard errors are very smooth, and a more thorough model evaluation is necessary to find out whether that is appropriate or whether the time-series model underfits the unemployment data or is open to improvement in other ways. There are many ways in which the time-series SAE model may be extended to further improve the estimates and standard errors. For example, it may be an improvement to use a logarithmic link function in the model formulation as in You (2008). Effects would then be multiplicative instead of additive. Another improvement would come from a more extensive modeling of the sampling variances (You and Chapman, 2006; You, 2008; Gómez-Rubio et al., 2010). The models can also be improved by including additional auxiliary information at the province by month level, for instance registered unemployment. In Datta et al. (1999) similar effects associated with unemployment insurance are modeled as varying over areas, although not over time.

Another issue is that survey redesigns can occur. The Dutch LFS has undergone redesigns in 2010 and 2012, which have led to changes in non-sampling error and consequent level changes in the time-series of direct unemployment estimates. To account for such changes, intervention variables may be added (van den Brakel and Krieg, 2015) as additional components in the measurement part of model (10). A last practical issue is that of benchmarking the detailed estimates obtained from the multilevel time-series model so as to agree with estimates at a higher level of aggregation. The estimates at a high level of aggregation are often estimated prior to the detailed figures, using different methods; in the case of the Dutch LFS, national unemployment figures are currently estimated using a structural time-series model. To obtain a consistent set of published estimates the detailed figures can be minimally adjusted so as to agree with these national estimates. Another possibility is to derive national monthly figures from the domain estimates by aggregation. If estimates for multiple related target variables are based on separate models, more consistency relations may have to be imposed. For the purpose of benchmarking and imposing more general consistency relations various methods are available, see for example Bell et al. (2013) and references therein.

References

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

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

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

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

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

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

Bollineni-Balabay, O., J. van den Brakel, F. Palm, and H. Boonstra (2016). Multilevel hierarchical bayesian vs. state-space approach in time series small area estimation; the dutch travel survey. Technical Report Discussion paper 2016-03, Statistics Netherlands.

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

Boonstra, H. J. (2016). *mcmcsae: MCMC Small Area Estimation*. R package version 0.7.

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

Boonstra, H. J. and J. Michiels (2013). Evaluatie van kleine-domeinschatters (kds) voor arbeidspositie op regionaal niveau, in dutch, internal report ppm-2013-03-11-hbta, statistics netherlands.

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

Box, G., G. Jenkins, and G. Reinsel (2008). *Time Series Analysis; forecasting and control*. Wiley and Sons.

Chan, J. and I. Jeliazkov (2009). Efficient simulation and integrated likelihood estimation in state space models. *International Journal of Mathematical Modelling and Numerical Optimisation 1*, 101--120.

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

Doornik, J. (2009). *An Object-oriented Matrix Programming Language Ox 6*. Timberlake Consultants Press.

Durbin, J. and S. Koopman (2001). *Time Series Analysis by State Space Methods*. Oxford University Press.

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

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

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

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

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

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

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

Gómez-Rubio, V., N. Best, S. Richardson, G. Li, and P. Clarke (2010). Bayesian statistics for small area estimation. Technical Report http://www.bias-project.org.uk/research.htm.

Harvey, A. (1989). *Forecasting, structural time series models and the Kalman filter*. Cambridge University Press.

Harvey, A. (2006). Seasonality and unobserved components models: an overview. Technical report, Conference on seasonality, seasonal adjustment and their implications for short-term analysis and forecasting, 10-12 May 2006, Luxembourg.

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

Knorr-Held, L. and H. Rue (2002). On block updating in markov random field models for disease mapping. *Scandinavian Journal of Statistics 29*(4), 597--614.

Koopman, S. (1997). Exact initial kalman filtering and smoothing for non-stationary time series models. *Journal of the American Statistical Association 92*, 1630--1638.

Koopman, S., A. Harvey, J. Doornik, and A. Shephard (1999). *STAMP: Structural Time Series Analyser, Modeller and Predicter*. Timberlake Consultants, Press London.

Koopman, S., A. Harvey, J. Doornik, and A. Shephard (2007). *STAMP 8: Structural Time Series Analyser, Modeller and Predictor*. Timberlake Consultants, Press London.

Koopman, S., A. Shephard, and J. Doornik (2008). *Ssfpack 3.0: Statistical algorithms for models in state-space form*. Timberlake Consultants, Press London.

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

McCausland, W., S. Miller, and D. Pelletier (2011). Simulation smoothing for state-space models: A computational efficiency analysis. *Computational Statistics and Data Analysis 55*, 199--212.

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

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

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

Pfeffermann, D. and R. Tiller (2005). Bootstrap approximation to prediction mse for state-space models with estimated parameters. *Journal of Time Series Analysis 26*, 893--916.

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

Piepho, H.-P. and J. Ogutu (2014). Simple state-space models in a mixed model framework. *The American Statistician 61*(3), 224--232.

Proietti, T. (2000). Comparing seasonal components for structural time series models. *International Journal of Forecasting 16*, 247--260.

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

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

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

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

Ruiz-Cárdenas, R., E. Krainski, and H. Rue (2012). Direct fitting of dynamic models using integrated nested laplace approximations - inla. *Computational Statistics and Data Analysis 56*, 1808--1828.

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

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

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

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

Woodruff, R. (1966). Use of a regression technique to produce area breakdowns of the monthly national estimates of retail trade. *Journal of the American Statistical Association 61*(314), 496--504.

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

You, Y. and B. Chapman (2006). Small area estimation using area level models and estimated sampling variances. *Survey Methodology 32*(1), 97--103.

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

Appendices

I Multilevel and structural formulations of time-series models

To clarify the relation between the multilevel and structural approaches to time-series models, we use the example of the simple local level model, which, in its structural formulation is given by (see e.g. Durbin and Koopman (2001))

$$y_t = \alpha_t + \epsilon_t\,, \qquad \epsilon_t \sim N(0, \sigma_\epsilon^2)\,, \tag{30}$$

$$\alpha_{t+1} = \alpha_t + \eta_t\,, \qquad \eta_t \sim N(0, \sigma_\eta^2)\,, \tag{31}$$

for $t = 1, \ldots, T$ where ϵ_t and η_t are all independent.

A multilevel specification of the local level model can be stated in terms of the vectors $\alpha = (\alpha_1, \ldots, \alpha_T)'$ and similarly defined y and ϵ, as

$$y = \alpha + \epsilon\,, \qquad \epsilon \sim N(0, \sigma_\epsilon^2 I_T)\,, \tag{32}$$

$$\alpha \sim N(0, \sigma_\eta^2 Q^{-1})\,, \tag{33}$$

where I_T is the T-dimensional identity matrix, and $\frac{1}{\sigma_\eta^2}Q$ is the precision matrix of α given by

$$Q = \begin{pmatrix} 1 & -1 & & & & \\ -1 & 2 & -1 & & & \\ & -1 & 2 & -1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & -1 & 2 & -1 \\ & & & & -1 & 1 \end{pmatrix}. \tag{34}$$

The reason for expressing the multivariate normal distribution for α in terms of its precision matrix instead of its covariance matrix is that it is directly related to the conditional independence structure of the random walk. The precision matrix in this case is a simple band matrix with bandwidth equal to 1, and thus very sparse for large T.

To see that the multilevel and structural specifications define the same model, consider the covariance matrix of α implied by (31). Given α_1 we have $\mathrm{cov}(\alpha_s, \alpha_t | \alpha_1) = \min(s-1, t-1)\sigma_\eta^2$. If an independent prior $\alpha_1 \sim N(a_1, v_1\sigma_\eta^2)$ is assigned, then the elements of the covariance matrix are $V_{\alpha;st} = cov(\alpha_s, \alpha_t) = (v_1 + \min(s-1, t-1))\,\sigma_\eta^2$. This matrix can be inverted to give

$$Q_\alpha = \sigma_\eta^2 V_\alpha^{-1} = \begin{pmatrix} 1+1/v_1 & -1 & & & & \\ -1 & 2 & -1 & & & \\ & -1 & 2 & -1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & -1 & 2 & -1 \\ & & & & -1 & 1 \end{pmatrix}. \tag{35}$$

In the diffuse prior limit $v_1 \to \infty$, which in Durbin and Koopman (2001) gives the so-called exact initial Kalman filter, Q_α reduces to Q, thus restoring the symmetry over time. In this limit the precision matrix remains well-defined, although it becomes singular, as $Q\iota_T = 0_T$, where ι_T and 0_T are T-vectors of 1s and 0s. The resulting prior (33) is improper, and the associated random walk model component is an intrinsic autoregressive model (Besag and Kooperberg, 1995; Rue and Held, 2005). The prior impropriety corresponds to the non-informativeness about the

overall level of the random walk. It is customary to add an intercept to the multilevel model to capture the mean level of the series,

$$y_t = \mu + \alpha_t + \epsilon_t\,, \qquad \epsilon_t \sim N(0, \sigma_\epsilon^2)\,, \tag{36}$$

with a non-informative prior on μ. For identification and interpretability reasons, a sum-to-zero constraint is then imposed on α:

$$\iota_T'\alpha = 0\,. \tag{37}$$

Now α is more easily recognized as a random effect term distributed around zero, although temporally correlated.

II Multilevel model specifications

Here we specify the fixed and random effect components used in the multilevel models. In the ten models T1SR to T4R the first two characters determine the random trend component(s) and the remaining ones determine the fixed effects used in the model. The trend components are summarized in Table 13, the fixed effects in Table 14.

The random effect components are described in terms of the general specification described in Section 4.2, in paricular equation (24), i.e. in terms of their parameterized variance matrix V, precision matrix $Q_A = A^{-1}$, and constraint matrix R that is imposed on the random effect coefficient vector. The matrix V can be a diagonal matrix, a fully parameterized unstructured covariance matrix or just a scalar variance parameter. This is indicated in the corresponding column in Table 13 with the dimension of V appended between parentheses. Recall that m_A is the number of areas, i.e. provinces. The precision matrix Q_A is given in the next column, in which Q_{RW1} stands for a first-order random walk precision matrix defined in equation (34) and Q_{RW2} for a second-order random walk precision matrix defined in equation (25). The dimension is appended between parentheses and as all random walks are defined over time, it is always m_T. The last column lists the constraint matrices, which stem from the singular vectors of the precision matrices Q_A. Here $R_{RW1}(m_T)$ is a row-vector of m_T ones, and $R_{RW2}(m_T)$ is a $2 \times m_T$ matrix with ones on the first row and $1, 2, \ldots, m_T$ on the second row.

trend model	component	V	Q_A	R
T1	provincial trend	diagonal(m_A)	$Q_{RW2}(m_T)$	$I_{m_A} \otimes R_{RW2}(m_T)$
T2	provincial trend	unstructured(m_A)	$Q_{RW2}(m_T)$	$I_{m_A} \otimes R_{RW2}(m_T)$
T3	global trend	scalar(1)	$Q_{RW2}(m_T)$	$R_{RW2}(m_T)$
	provincial trend	diagonal(m_A)	$Q_{RW1}(m_T)$	$I_{m_A} \otimes R_{RW1}(m_T)$
T4	global trend	scalar(1)	$Q_{RW2}(m_T)$	$R_{RW2}(m_T)$
	provincial trend	diagonal(m_A)	$Q_{RW2}(m_T)$	$I_{m_A} \otimes R_{RW2}(m_T)$

Table 13 Random effect components defining the four different trend models

The fixed effects in Table 14 are specified in terms of the variables province (12 categories), time (a single quantitative variable taking values $1, \ldots m_T$), season (12 categories), wave (5 categories) and wave2 (2 categories, distinguishing only between first and follow-up wave).

fixed effects model	fixed effects
SR	province × (time + season + wave)
S	province × (time + season) + wave
R	province × (time + wave) + season
R2	province × (time + wave2) + season + wave
no S or R specified (as in model T3)	province × time + season + wave

Table 14 Fixed effects used in the models

III Gibbs sampler for the multilevel time-series model

For notational convenience we rewrite model (22) as

$$y = \eta + e \tag{38}$$

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

$$e \sim N(0, \bigoplus_{i=1}^{m_A} \lambda_i \Phi_i)\,, \tag{39}$$

in terms of a covariance matrices Φ_i depending on initial variance estimates treated as known, and scale factors λ_i, one for each area $i = 1, \ldots, m_A$.

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

$$p(\psi|y) \propto p(\psi)p(y|\psi)\,, \tag{40}$$

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

The Gibbs sampler does not directly sample from the joint posterior. Instead, it iteratively samples from the full conditional distributions $p(\psi_g|y, \psi_1, \ldots, \psi_{g-1}, \psi_{g+1}, \ldots, \psi_G)$, for a suitable decomposition of ψ in blocks ψ_g, $g = 1, \ldots, G$. The full conditionals for the class of linear multilevel models considered in this paper are easy to sample from: they are normal for all (fixed or random) coefficients and inverse chi-squared or inverse-Wishart for the variance parameters. With G the number of parameter blocks and K the number of simulations, the Gibbs sampling algorithm is as follows:

choose starting values $\psi_g^{(0)}$ for $g = 1, \ldots, G$
for k in 1 **to** K

for g in 1 **to** G

draw $\psi_g^{(k)}$ from $p(\psi_g|y, \psi_1^{(k)}, \ldots, \psi_{g-1}^{(k)}, \psi_{g+1}^{(k-1)}, \ldots, \psi_G^{(k-1)})$

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

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

III.0.2 Full conditional distributions

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

$$\lambda_i \overset{\text{ind}}{\sim} \text{Inv}{-}\chi^2(1,1)\,, \tag{41}$$

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

$$p(\lambda_i|y,.) \propto p(\lambda_i) N\,(e|0, \lambda_i \Phi_i)\,, \tag{42}$$

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

$$\begin{aligned} p(\lambda_i|y,.) &= \text{Inv}{-}\chi^2(\lambda_i|d_i, s_i)\,, \\ d_i &= n_i + 1\,, \\ s_i &= \frac{1}{d_i}\left(1 + e'_{\{i\}} \Phi_i^{-1} e_{\{i\}}\right)\,, \end{aligned} \tag{43}$$

where subscript $\{i\}$ denotes taking the elements associated with the ith level of factor f, and n_i is the number of those elements. In our case all $n_i = m/m_A$ for all i.

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

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

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

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

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

We use redundant multiplicative parameterization, which improves convergence of the Gibbs sampler (Gelman et al., 2008), and yields more robust prior distributions for the variance parameters (Gelman, 2006). For that purpose, a d-dimensional parameter vector ξ and a

q-vector $\tilde{v}$ of raw random effects are introduced, which combine to form the original coefficients as

$$v = \Delta_\xi \tilde{v} \qquad \Delta_\xi = I_l \otimes \mathrm{diag}(\xi) = \mathrm{diag}(W\xi)\,, \tag{44}$$

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

Priors on ξ and $\tilde{v}$ are

$$\begin{aligned} \xi &\sim N(0, I_d)\,, \\ \tilde{v} &\sim N(0, A \otimes \tilde{V})\,, \end{aligned} \tag{45}$$

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

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

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

$$\tilde{V} \sim \mathrm{Inv-Wish}(\nu_v, \Psi_v)\,, \tag{46}$$

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

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

$$\tilde{\sigma}^2_{v;k} \overset{\text{ind}}{\sim} \mathrm{Inv}{-}\chi^2(\nu_{v;k}, s^2_{v;k}) \tag{47}$$

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

$$\tilde{\sigma}^2_v \sim \mathrm{Inv}{-}\chi^2(\nu_v, s^2_v)\,. \tag{48}$$

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

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

$$v \sim N(0, A \otimes V) \qquad V = \mathrm{diag}(\xi)\, \tilde{V}\, \mathrm{diag}(\xi)\,. \tag{49}$$

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

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

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

$$p(\xi|y,.) \propto N(\xi|0, I_d) N(y|Z\Delta_\xi \tilde{v} + \ldots, \Sigma) \tag{50}$$

Now

$$\Delta_\xi \tilde{v} = \mathrm{diag}(W\xi)\tilde{v} = \mathrm{diag}(\tilde{v})W\xi = \Delta_{\tilde{v}}\xi\,, \tag{51}$$

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

$$\begin{aligned} p(\xi|y,.) &= N(\xi|E_\xi, V_\xi) \\ V_\xi &= \left(\Delta'_{\tilde{v}} Z' \Sigma^{-1} Z \Delta_{\tilde{v}} + I_d\right)^{-1} \\ E_\xi &= \left(\Delta'_{\tilde{v}} Z' \Sigma^{-1} Z \Delta_{\tilde{v}} + I_d\right)^{-1} \Delta'_{\tilde{v}} Z' \Sigma^{-1} e_v\,, \end{aligned} \tag{52}$$

where $e_v = y - \eta + Zv$. Note that everything can be expressed in terms of v instead of $\tilde{v}$ by using $\tilde{v} = \Delta_\xi^{-1} v$. For $d = 1$ or in the case that $\tilde{V} = \sigma_v^2 I_d$ is defined in terms of a single variance parameter, (52) reduces to

$$\begin{aligned} p(\xi|y,.) &= N(\xi|E_\xi, V_\xi) \\ V_\xi &= \left(\tilde{v}'Z'\Sigma^{-1}Z\tilde{v} + 1\right)^{-1} \\ E_\xi &= \left(\tilde{v}'Z'\Sigma^{-1}Z\tilde{v} + 1\right)^{-1} \tilde{v}'Z'\Sigma^{-1}e_v \,. \end{aligned} \tag{53}$$

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

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

$$\begin{aligned} p(\tilde{V}|y,.) &\propto \text{Inv} - \text{Wish}(\tilde{V}|\nu_v, \Psi_v) N(\tilde{v}|0, A \otimes \tilde{V}) \\ &\propto \left|A \otimes \tilde{V}\right|^{-1/2} e^{-\frac{1}{2}\tilde{v}'(Q_A \otimes \tilde{V}^{-1})\tilde{v}} \\ &\times \left|\tilde{V}\right|^{-(\nu_v + d + 1)/2} e^{-\frac{1}{2}\text{tr}(\Psi_v \tilde{V}^{-1})} \\ &\propto \left|\tilde{V}\right|^{-(\nu_v + l^* + d + 1)/2} e^{-\frac{1}{2}\text{tr}\left[(\Psi_v + \tilde{v}_M' Q_A \tilde{v}_M)\tilde{V}^{-1}\right]} , \end{aligned} \tag{54}$$

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

$$|C \otimes D| = |C|^{\text{rank}(D)} |D|^{\text{rank}(C)} , \tag{55}$$

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

$$\begin{aligned} &\tilde{v}' \left(Q_A \otimes \tilde{V}^{-1}\right) \tilde{v} = \text{vec}(\tilde{v}_M')' \left(Q_A \otimes \tilde{V}^{-1}\right) \text{vec}(\tilde{v}_M') \\ &= \text{vec}(\tilde{v}_M')'\text{vec}(\tilde{V}^{-1}\tilde{v}_M' Q_A) = \text{tr}(\tilde{v}_M \tilde{V}^{-1} \tilde{v}_M' Q_A) \\ &= \text{tr}(\tilde{v}_M' Q_A \tilde{v}_M \tilde{V}^{-1}) \,. \end{aligned} \tag{56}$$

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

$$\begin{aligned} p(\tilde{V}|y,.) &= \text{Inv} - \text{Wish}(\tilde{V}|\nu_{v1}, \Psi_{v1}) \\ \nu_{v1} &= \nu_v + l^* \\ \Psi_{v1} &= \Psi_v + \tilde{v}_M' Q_A \tilde{v}_M \,. \end{aligned} \tag{57}$$

Note that V is obtained from (49), or more immediately by drawing from an inverse Wishart distribution with the same degrees of freedom, but with scale matrix

$$\text{diag}(\xi)\Psi_v \text{diag}(\xi) + v_M' Q_A v_M \,, \tag{58}$$

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

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

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

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

since in this case

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

Therefore,

$$\begin{aligned}p(\tilde{\sigma}_{v;1}^2\dots\tilde{\sigma}_{v;d}^2|y,.)&=\prod_{k=1}^{d}\text{Inv}{-}\chi^2(\tilde{\sigma}_{v;k}^2|\nu_{v1;k},s_{v1;k}^2)\\ \nu_{v1;k}&=\nu_{v;k}+l^*\\ s_{v1;k}^2&=\frac{1}{\nu_{v1;k}}\left(\nu_{v;k}s_{v;k}^2+(\tilde{v}_M'Q_A\tilde{v}_M)_{kk}\right)\,.\end{aligned}\tag{59}$$

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

$$\frac{1}{\nu_{v1;k}}\left(\nu_{v;k}s_{v;k}^2\xi_k^2+(v_M'Q_Av_M)_{kk}\right)\tag{60}$$

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

$$p(\tilde{\sigma}_v^2|y,.)\propto\left(\tilde{\sigma}_v^2\right)^{-\frac{\nu_v}{2}}e^{-\frac{\nu_vs_v^2}{2\tilde{\sigma}_v^2}}(\tilde{\sigma}_v^2)^{-dl^*/2}e^{-\frac{1}{2\tilde{\sigma}_v^2}\text{tr}(\tilde{v}_M'Q_A\tilde{v}_M)}\tag{61}$$

and so

$$\begin{aligned}p(\tilde{\sigma}_v^2|y,.)&=\text{Inv}{-}\chi^2(\tilde{\sigma}_v^2|\nu_{v1},s_{v1}^2)\\ \nu_{v1}&=\nu_v+dl^*\\ s_{v1}^2&=\frac{1}{\nu_{v1}}\left(\nu_vs_v^2+\text{tr}(\tilde{v}_M'Q_A\tilde{v}_M)\right)\,.\end{aligned}\tag{62}$$

In this case the original variance parameter is $\sigma_v^2=\xi^2\tilde{\sigma}_v^2$.

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

$$p(\tilde{v}|y,.)\propto N(\tilde{v}|0,A\otimes\tilde{V})N(y|Z\Delta_\xi\tilde{v}+\dots,\Sigma)\,,\tag{63}$$

implying

$$\begin{aligned}p(\tilde{v}|y,.)&=N(\tilde{v}|E_{\tilde{v}},V_{\tilde{v}})\\ V_{\tilde{v}}&=\left(\Delta_\xi'Z'\Sigma^{-1}Z\Delta_\xi+Q_A\otimes\tilde{V}^{-1}\right)^{-1}\\ E_{\tilde{v}}&=\left(\Delta_\xi'Z'\Sigma^{-1}Z\Delta_\xi+Q_A\otimes\tilde{V}^{-1}\right)^{-1}\Delta_\xi'Z'\Sigma^{-1}e_v\,.\end{aligned}\tag{64}$$

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

$$\begin{aligned}p(v|y,.)&=N(v|E_v,V_v)\\ V_v&=\left(Z'\Sigma^{-1}Z+Q_A\otimes V^{-1}\right)^{-1}\\ E_v&=\left(Z'\Sigma^{-1}Z+Q_A\otimes V^{-1}\right)^{-1}Z'\Sigma^{-1}e_v\,.\end{aligned}\tag{65}$$

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

$$v\to v^*=v-V_vR\left(R'V_vR\right)^{-1}R'v\,,\tag{66}$$

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

III.1 The Deviance Information Criterion

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

$$D(\psi) = -2 \log p(y|\psi) = m \log 2\pi + \log |\Sigma| + e' \Sigma^{-1} e \,,$$

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

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

Models with lower DIC values are preferred.

III.2 Implementation

The above Gibbs sampler has been implemented in R (R Development Core Team, 2009), and is being developed into a package called `mcmcsae` (Boonstra, 2016). The package makes extensive use of the sparse matrix facilities provided by package `Matrix` (Bates and Maechler, 2010) and also of some dense and sparse matrix routines of the C++ library Eigen, via packages `Rcpp` (Eddelbuettel and Francois, 2011) and `RcppEigen` (Bates and Eddelbuettel, 2013).