Statistics Netherlands

Discussion Paper

Multilevel Hierarchical Bayesian vs. State Space Approach in Time Series Small Area Estimation: the Dutch Travel Survey

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

2016 | 03

Oksana Bollineni-Balabay
Jan van den Brakel
Franz Palm
Harm-Jan Boonstra

Contents

This study compares two different techniques in a time series small area application: state space models estimated with the Kalman filter with a frequentist approach to hyperparameter estimation, and multilevel time series models estimated within the hierarchical Bayesian framework. The application chosen is the Dutch Travel Survey featuring level breaks caused by the survey redesigns, as well as small sample sizes for the main publication domains. Both models require variances of the design-based domain estimates as prior information. In practice, however, only unstable estimates of the design-based variances are available. In this paper, excessive volatility and a possible bias in design-based variance estimates are removed with the help of a state space model. The multilevel and state space modelling approaches deliver similar results. Slight differences in model-based variance estimates appear mostly in small-scale domains and are due to neglecting uncertainty around the hyperparameter estimates in state space models, and to a lesser extent due to skewness in the posterior distributions of the parameters of interest. The results suggest that the reduction in design-based standard errors with the hierarchical Bayesian approach is over 50% at the provincial level, and over 30% at the national level, averaged over the domains and time.

1 Introduction

The problem of small area estimation (SAE) emerges when reliable estimates for small areas cannot be produced relying solely on a design-based inference procedure due to an insufficient number of sampled units. Either the number of units sampled in the domain of interest is too small, or the variable of interest has a rare occurrence, in cases when the break-up into domains is defined post hoc, after the survey has been done. An increasing demand for small area statistics boosted the development of numerous SAE techniques. Most of these are meant for cross-sectional data, where the so-called *strength over space* could be borrowed. This implies that sample information from other similar domains could be used to improve the design estimate for a small area of interest. Sometimes, an area may not be sampled at all, in which case regression synthetic estimators (Gonzalez (1973)) produce figures using some auxiliary variables. Composite estimators may be used to combine direct estimates with their synthetic counterparts. This is often done in the framework of the Fay-Herriot model (Fay and Herriot (1979)), which is estimated using the empirical best linear unbiased predictor (EBLUP) or hierarchical Bayesian (HB) approach. The literature on these methods is extensive, see, e.g., Ghosh and Lahiri (1987), Prasad and Rao (1990), Rao (2003).

While small area estimates may be improved by borrowing strength over space, adding the time dimension further offers a huge potential for improvement. *Borrowing strength over time* is realised by using sampling information accumulated over time in the respective, as well as in other domains. Time series techniques, such as multilevel times series and structural (unobserved component) time series (STS) models, are powerful tools in producing more reliable estimates for repeatedly conducted surveys, both for small and non-small areas. Thanks to their ability to borrow strength over time and domain space, such models can extract the so-called signal by removing a great part of the sampling noise from design-based point-estimates. As a result, standard errors (SEs) of such model-based point-estimates are usually substantially lower than the design SEs. Furthermore, time series models can help compare figures prior to and after survey redesigns by estimating level breaks. Time series extensions of the Fay-Herriot model can be found in Rao and Yu (1994), Datta et al. (1999), and

You (2008) (see Rao (2003) for an overview). For small area applications modelled with the STS techniques, see Harvey and Koopman (1997), Pfeffermann (1991), Pfeffermann and Bleuer (1993), Pfeffermann and Tiller (2006), Pfeffermann et al. (2014), Krieg and Van den Brakel (2012) and Bollineni-Balabay et al. (2016). Level break estimation within the STS framework is illustrated in Van den Brakel and Roels (2010).

With this paper, the authors would like to contribute to empirical evidence and practical considerations in times series SAE. This study compares two different time series techniques in a real life time series small area application - the Dutch Travel Survey (DTS). This survey features unacceptably volatile estimates in its more than 600 small domains. The reasons for that are small sample sizes at the provincial level and several survey redesigns, which make the official figures hard to compare over time. Therefore, this survey could largely benefit from the time series model-based approach in producing official statistical figures. The two time series approaches considered in this paper are multilevel time series modelling within the hierarchical Bayesian (HB) framework and the frequentist-based STS approach facilitated by the Kalman filter. The initial intention was to check if specifications with random regression coefficients and/or random intercepts, like in multilevel models, perform better than STS models where such components are conventionally modelled as fixed effects. Further, another model is considered within the multilevel Bayesian framework - a model with a nearly identical to that of the STS model likelihood function (differing only in the way the trend model is specified). This comparison is performed in order to check if the quality of the STS model-based estimates produced by frequentist estimation routines is good enough compared to the full Bayesian multilevel framework, as most practitioners, including Statistics Netherlands, estimate STS models in the frequentist framework.

All model specifications within both modelling approaches are made comparable in terms of the same pooling dimension, as well as in that no spatial correlation is assumed across domains. Rationales behind multivariate against univariate modelling dimensions are also addressed. The multilevel time series and STS approaches are compared in terms of adequacy of model-based point-estimates for the trend and for multiple level breaks due to the survey redesigns, as well as in terms of gain in precision that can be reached within these two modelling frameworks. It is well known, though, that variance estimates often get a negative bias in the frequentist STS models due to neglected hyperparameter uncertainty (see, e.g., Pfeffermann and Tiller (2005), Bollineni-Balabay et al. (2015)), so a direct comparison of variance estimates obtained from HB multilevel and frequentist STS models is not completely fair. In this work, we quantify the uncertainty around hyperparameters to make the comparison of the two models fairer.

Another novelty of this paper lies in exploiting the STS approach to refine the multilevel one in the case of unreliable or missing variance estimates of the design estimator (further referred to as design-based variance estimates). These are used as input in multilevel models and are treated as the true known sampling error variances. This paper shows how volatility, missing values, as well as a possible bias in design-based variance estimates can be alleviated.

Section 2 of this paper presents a brief description of the DTS. Section 3 is fully devoted to the multilevel model setting. It presents the model itself, as well as estimation details for the HB approach in two subsections. Section 4 presents the STS model and has the same structure. Section 5 explains how unreliable design-based variance estimates can be improved using the STS technique. Sections 6 and 7 present empirical results for the DTS at the provincial and national level, respectively. In each of these sections, the multivariate trend structure is examined with regard to possible pooling across one or more factor variables. Further, the performance of the multilevel HB model is compared to that of the STS model. Section 8 contains conclusions and discussion.

Table 2.1 Domain classification of the DTS along three factors

Provinces	Mobility motives ("Mot")	Transport modalities ("Mod")
1. Groningen	1. Work	1. Car driver
2. Friesland	2. Business	2. Car passenger
3. Drenthe	3. Visits	3. Train
4. Overijssel	4. Shopping	4. Bus/tram/metro
5. Flevoland	5. Education	5. Scooter/moped
6. Gelderland	6. Recreative	6. Bicycle
7. Utrecht	7. Other motives	7. Walking
8. Noord-Holland		8. Other modalities
9. Zuid-Holland		
10. Zeeland		
11. Noord-Brabant		
12. Limburg		

2 Dutch Travel Survey

The DTS is a stratified two-stage survey that attempts to measure the travel behavior of the Dutch population. Every year, a cross-sectional sample is drawn with sampling units being defined either as households (before 2010), or persons (since 2010). The variable of interest considered in this study is the number of kilometers per person per day (km-pppd) covered per transport modality and motive either at the national, or provincial level. Design estimates for km-pppd are obtained with the general regression (GREG) estimator (equation (6.4.1) in Särndal et al. (1992)) by dividing the estimated total number of kilometers covered within a year by the population with a certain transport modality/motive by the population size and the number of days in a year. Point- and variance GREG-estimates are produced on an annual basis and for different break-downs into domains. The time span considered in this paper is $T = 29$, covering the years 1985-2013. Since 1985, the survey has undergone several redesigns. Until 1994, the population of interest consisted of residents aged 12 years and older. Since 1994, children under 12 years old have also been included in the population of interest. In 1995, the sample became six times as large as before (a net-sample size increase from 10 000 to about 65 000 households). Since 2003-2004, however, net sample sizes have been considerably lower constituting slightly more than 20 000 households (about 43 000 persons between 2010 and 2013). In 1999, the DTS went through the second major redesign that featured some response-motivation and follow-up measures. The published series were corrected for this level break within Statistics Netherlands. In 2004, the data collection for the survey was transferred to another agency. Finally, the survey was redesigned again in 2010 and since then has been conducted by Statistics Netherlands. These two redesigns caused major discontinuities in the time series and have not been corrected for so far.

Domains of the DTS are formed by an intersection of three factors (see Table 2.1): transport modalities (Mod) indexed as i, mobility motives (Mot) indexed as j, and territorial units (provinces) indexed as p. The total number of provinces is $P = 12$, and each province p is broken into $J = 7$ motives, and each motive j is broken into $I = 8$ transport modalities.

The model described in the subsequent section will be applied at two aggregation levels of the DTS survey. Firstly, time series of national averages of distances covered, i.e. km-pppd, can be obtained, thus forming an intersection of modalities and motives. Secondly, one can go one level deeper where the target variable is defined as time series of provincial averages, i.e. km-pppd at the intersection of all the three above-mentioned factors.

3 Multilevel Time Series Approach Applied to the DTS

3.1 Model Specification

Application of multilevel models is common practice in small area estimation (SAE). The model presented in this section is a time series extension of the well known Fay-Herriot model for a cross-section.
In the model below, small areas (domains) are indexed with $m \epsilon \{1, ..., M\}$, and the number of years is T. At the national level, $M = I \times J = 56$; at the provincial level, $M = I \times J \times P = 672$. The M-dimensional vector of GREG estimates $\boldsymbol{Y_t}$ can be expressed as the sum of signal $\boldsymbol{\theta_t}$ and survey errors contained in vector $\boldsymbol{e_t}$:

$$\boldsymbol{Y_t} = \boldsymbol{\theta_t} + \boldsymbol{e_t}, \quad \boldsymbol{e_t} \overset{id}{\sim} N(\boldsymbol{0}, \boldsymbol{\Phi_t}), \quad t \, \epsilon \, \{1, ..., T\}, \tag{3.1}$$

where the signal $\boldsymbol{\theta_t}$ consists of the true population parameter and a measurement error that is stable over time under a particular survey design and changes when a redesign occurs. A change in the measurement error is represented by a level break (or discontinuity). Level breaks are included in the signal on purpose, so that the reader can decide which of those to keep or remove from the signal, depending on whether or not a certain survey modification is viewed as an improvement. Survey errors $\boldsymbol{e_t}$ are independently distributed over time with a M-dimensional covariance matrix $\boldsymbol{\Phi_t}$. Although sampling units may occur in several domains at the same time (e.g., in domains "Work-Bicycle" and "Work-Train"), correlations between such domains are assumed to be negligible, which renders the $\boldsymbol{\Phi_t}$-matrix diagonal with design-based variances for every domain m. These variances are generally unknown, so design-based variance estimates $Var(Y_{m,t})$ (Särndal et al. (1992))are used instead and are treated as the true known variances.
Here and further in this paper, vectors and matrices are printed in bold. Vectors with a subscript t or m are T- or M-dimensional, respectively. No correlation is assumed between different stochastic terms in this paper, unless explicitly mentioned.
The signal development over time is described by a combination of a stochastic and deterministic trend together with level breaks, as mentioned above. All potential cyclical movements in the economy are covered by the stochastic trend, hence the absence of the cyclical component. There is no seasonal component because the data is annual. The following time-series model is assumed for the M-dimensional vector of signals $\boldsymbol{\theta_t}$:

$$\begin{aligned}\boldsymbol{\theta_t} =& c\boldsymbol{\iota} + \kappa t\boldsymbol{\iota} + \delta_{1,t}^{RE}(\beta_1^{RE}\boldsymbol{\iota} + \boldsymbol{\beta_1^{RE}}) + ... + \delta_{K_{RE},t}^{RE}(\beta_{K_{RE}}^{RE}\boldsymbol{\iota} + \boldsymbol{\beta_{K_{RE}}^{RE}}) + \boldsymbol{\nu} + \boldsymbol{\kappa}t + \\ & \delta_{1,t}^{FE}\boldsymbol{\beta_1^{FE}} + ... + \delta_{K_{FE},t}^{FE}\boldsymbol{\beta_{K_{FE}}^{FE}} + \boldsymbol{u_t}, \quad t \, \epsilon \, \{1, ..., T\},\end{aligned} \tag{3.2}$$

where c and κ are the overall intercept and linear time trend coefficient, respectively, $\boldsymbol{\iota}$ is an M-dimensional vector of ones, $\boldsymbol{\nu} = (\nu_1, ..., \nu_M)'$ are independent domain effects that serve as random intercepts in the model, $\boldsymbol{\kappa} = (\kappa_1, ..., \kappa_M)'$ are domain-specific random linear time trend coefficients; K_{RE} and K_{FE} denote numbers of level interventions modelled as random or fixed effects, respectively; $\boldsymbol{\beta_k^{RE}}$ and $\boldsymbol{\beta_k^{FE}}$ are vectors with level break coefficients assumed to be random or fixed effects, respectively. Scalars β_k^{RE} represent the mean of the k-th level break random coefficient across domains; $\delta_{k,t}^{RE}$ and $\delta_{k,t}^{FE}$ are indicator variables for level breaks modelled as random and fixed effects, respectively. These indicators switch from zero to one at the moment of a redesign. With an entry of a new dummy variable, all previous dummies remain equal to one. Note that scalar terms c, κ, β_k^{RE} are estimated as fixed effects. They are contained in a vector, say $\boldsymbol{\beta}$, along with $\boldsymbol{\beta_k^{FE}}$.

Terms in $\boldsymbol{u_t} = (u_{1,t}, \dots, u_{M,t})'$ are domain-by-time effects modelled as random walks. The pattern of the DTS series suggests a non-stationary model for the trend $\boldsymbol{u_t}$. Therefore, one could assume either a random walk ($I(1)$), or an integrated random walk ($I(2)$, the so-called smooth-trend model, as in Harvey (2001)). Here, the latter formulation is preferred due to its well-known flexibility and parsimony (Durbin and Koopman (2012), Ch. 3, Harvey (2001)). This application, as well as the one in Bollineni-Balabay et al. (2016) provide evidence for data overfitting when other specifications (I(1)-random walk or the local linear trend model) for the stochastic trends are applied. When applied to the DTS, these alternative trend specifications resulted in severe overfitting too. Therefore, only the smooth-trend model is considered in this paper. The disturbance terms $\boldsymbol{\epsilon_t}$ of the $\boldsymbol{u_t}$-terms are assumed to be normally, identically and independently distributed over time and across domains:

$$\boldsymbol{u_t} = \boldsymbol{u_{t-1}} + \boldsymbol{r_{t-1}}, \;\; \boldsymbol{r_t} = \boldsymbol{r_{t-1}} + \boldsymbol{\epsilon_t}, \;\; \boldsymbol{\epsilon_t} \overset{iid}{\sim} N(\boldsymbol{0}, \boldsymbol{\Sigma_u}), \quad t \in \{1, \dots, T\}, \tag{3.3}$$

where $\boldsymbol{\Sigma_u}$ is a diagonal IJ- or PIJ-dimensional matrix at the national or provincial level, respectively.

The following identifiability constraints are imposed on the stochastic trend estimates $\hat{u}_{m,t}$ to insure that, after accounting for the level break interventions, all the deterministic time variation in signal $\theta_{m,t}$ is accounted for by the linear trend $(\kappa + \kappa_m)t$, and the stochastic time variation around this deterministic trend and the remaining time-average level $c + \nu_m$ is accounted for solely by $\hat{u}_{m,t}$:

$$\sum_{t=1}^{T} \hat{u}_{m,t} = 0, \quad \sum_{t=1}^{T} \hat{u}_{m,t} t = 0. \tag{3.4}$$

Since there is no reason to assume that the stochastic trends have similar dynamics across either motives, or transport modalities, as will be verified in Subsection 7.1, every diagonal element of a IJ-dimensional square matrix $\boldsymbol{\Sigma_u}$ is assigned a unique value at the national level. As for the provincial level, one could assume that the trend disturbances for motive j and transport modality i from P provinces come from the same distribution. This would make the model parametrisation much more parsimonious. Then matrix $\boldsymbol{\Sigma_u}$ will consist of P block replicas of a IJ-dimensional covariance matrix $\boldsymbol{\Sigma_{[IJ]}}$ for motive-modality intersections:

$$\boldsymbol{\Sigma_u} = \boldsymbol{I_{[P]}} \otimes \boldsymbol{\Sigma_{[IJ]}}. \tag{3.5}$$

Whether or not this assumption is feasible, will be verified in Subsection 6.1.

The area-specific terms ν_m, κ_m and $\beta_{m,k}^{RE}$ are assumed to share the same variance across domains and to be normally and independently distributed over the domain space and between each other. By construction, they are distributed around zero due to the presence of the overall elements c, κ, and β_k^{RE}:

$$\begin{aligned} \boldsymbol{\nu} &\overset{iid}{\sim} N(\boldsymbol{0}, \sigma_\nu^2 \boldsymbol{I}), \\ \boldsymbol{\kappa} &\overset{iid}{\sim} N(\boldsymbol{0}, \sigma_\kappa^2 \boldsymbol{I}), \\ \boldsymbol{\beta_k^{RE}} &\overset{iid}{\sim} N(\boldsymbol{0}, \sigma_{\beta_k^{RE}}^2 \boldsymbol{I}), \end{aligned} \tag{3.6}$$

where $\boldsymbol{0}$ and $\boldsymbol{I}$ denote an M-dimensional vector and identity matrix, respectively. Such model specification allows us to draw terms ν_m, κ_m and $\beta_{m,k}^{RE}$ from distributions centred around zero, which will be necessary in the Bayesian estimation, but it makes these terms, as well as their means unidentifiable. The sums $(c + \nu_m)$, $(\kappa + \kappa_m)$ and $(\beta_k^{RE} + \beta_{m,k}^{RE})$ are, however, identifiable and constitute random intercepts, random linear trend slopes and random level break coefficients, respectively. Note that $\boldsymbol{\beta_k^{RE}}$-coefficients can be estimated as fixed effects by setting their variance $\sigma_{\beta_k^{RE}}^2$ equal to infinity. There is always a trade-off between regression coefficients' bias and variance when a random or fixed effects assumption is chosen, see Clark and Linzer (2015) for a discussion.

Model (3.2) assumes that area-specific random terms ν_m originate from one distribution (the same applies to κ_m and $\beta_{m,k}^{RE}$). It is, however, quite plausible that differences exist between means of random terms belonging to different motives, transport modalities or provinces. If the pattern of the series for a certain intersection of the motives with transport modalities suggests that provinces do not differ much from each other, the model could be written as in (3.7) with only two, instead of three additional terms per random component. Then the remaining random effect variation around the motive and modality means would be described by an M-dimensional random-effect vector. All this gives a rise to the following model, which is an extension of Model (3.2):

$$
\begin{aligned}
\boldsymbol{\theta_t} =& c\boldsymbol{\iota} + \boldsymbol{\iota}_{[M/I]} \otimes \boldsymbol{\nu_{Mod}} + \boldsymbol{\nu_{Mot}} \otimes \boldsymbol{\iota}_{[M/J]} + \boldsymbol{\nu} \\
&+ \kappa t\boldsymbol{\iota} + \boldsymbol{\iota}_{[M/I]} \otimes \boldsymbol{\kappa_{Mod}} t + \boldsymbol{\kappa_{Mot}} t \otimes \boldsymbol{\iota}_{[M/J]} + \boldsymbol{\kappa} t \\
&+ \delta_{1,t}^{RE}(\beta_1^{RE}\boldsymbol{\iota} + \boldsymbol{\iota}_{[M/I]} \otimes \boldsymbol{\beta_{1,Mod}^{RE}} + \boldsymbol{\beta_{1,Mot}^{RE}} \otimes \boldsymbol{\iota}_{[M/J]} + \boldsymbol{\beta_1^{RE}}) + \\
&\dots \\
&+ \delta_{K_{RE},t}^{RE}(\beta_{K_{RE}}^{RE}\boldsymbol{\iota} + \boldsymbol{\iota}_{[M/I]} \otimes \boldsymbol{\beta_{K_{RE},Mod}^{RE}} + \boldsymbol{\beta_{K_{RE},Mot}^{RE}} \otimes \boldsymbol{\iota}_{[M/J]} + \boldsymbol{\beta_{K_{RE}}^{RE}}) \\
&+ \delta_{1,t}^{FE}\boldsymbol{\beta_1^{FE}} + \dots + \delta_{K_{FE},t}^{FE}\boldsymbol{\beta_{K_{FE}}^{FE}} + \boldsymbol{u_t}, \quad t \in \{1, \dots, T\},
\end{aligned}
\tag{3.7}
$$

where each of the ν, κ, β^{RE}-terms sub-indexed with Mod contains random term means taken across motives and provinces for each of I transport modalities and is distributed as

$$
\begin{aligned}
&\boldsymbol{\nu_{Mod}} \overset{iid}{\sim} N(\boldsymbol{0}, \sigma_{\nu_{Mod}}^2 \boldsymbol{I}), \\
&\boldsymbol{\kappa_{Mod}} \overset{iid}{\sim} N(\boldsymbol{0}, \sigma_{\kappa_{Mod}}^2 \boldsymbol{I}), \\
&\boldsymbol{\beta_{Mod}^{RE}} \overset{iid}{\sim} N(\boldsymbol{0}, \sigma_{\beta_{Mod}^{RE}}^2 \boldsymbol{I}).
\end{aligned}
\tag{3.8}
$$

The same applies to random terms labelled with Mot. These are J-dimensional. Random terms $\boldsymbol{\nu}, \boldsymbol{\kappa}, \boldsymbol{\beta_k^{RE}}$ are M-dimensional and take care of the variation around their respective motive and modality means.

Note that the signal $\theta_{m,t}$ has no area-by-time white noise in the multilevel models presented in this paper. It is assumed that all the stochastic variation in $\theta_{m,t}$ over time is picked up by the stochastic structure of the trend. In fact, allowing for an additional noise in the signal resulted in data overfit in many, especially small domains. The absence of the additional noise term makes it furthermore easier to compare the multilevel approach with the structural time series one.

Special cases of models (3.2) and (3.7) could also be considered. For instance, all level breaks could be modelled as fixed effects. Further, if distribution assumptions about the random terms ν_m and κ_m do not seem to hold, one could consider modelling these terms as fixed effects. In the latter case, estimation at the national level will be reduced to a univariate setting.

3.2 Estimation Details on the Multilevel Model

The multilevel modelling technique constitutes a composite estimator, i.e. a combination of the design and synthetic estimators. The model in (3.1)-(3.2) can be estimated with the EBLUP (empirical best unbiased predictor), or within the hierarchical Bayesian (HB) approach. For this application, the HB approach is chosen with non-informative priors using the Gibbs-sampler pre-programmed in the *mcmcsae* R-package (Boonstra (2015)). The posterior means of the signals $\theta_{m,t}$ are taken to be domain point-estimates, and the posterior standard deviations of $\theta_{m,t}$s serve as their measures of uncertainty. These posterior standard deviations are further in this paper referred to as SEs for brevity.

Random terms ν_m, κ_m and $\beta_{m,k}^{RE}$ are independent and assigned normal priors as in (3.6) within the Bayesian estimation framework. The random walk term has a normal prior distribution as well: $\boldsymbol{u_m} \overset{iid}{\sim} N(\boldsymbol{0}, \sigma_{u,m}^2 \boldsymbol{A})$, $\boldsymbol{A}$ being a T-dimensional covariance matrix. In case when the initial

u_{m1} in all domains are assigned a non-informative prior, matrix $\boldsymbol{A}$ becomes unbounded. However, the limit of the precision matrix $\boldsymbol{A^{-1}}$ used in the Gibbs sampler is well-defined, see, e.g., Rue and Held (2005). Apart from that, the precision matrix $\boldsymbol{A^{-1}}$ is sparse (i.e., mostly contains zeros and, in this case, has five diagonals), which makes the Gibbs sampler efficient. The variances of the random-effect components are assigned inverse chi-squared priors with degrees of freedom v and a scale parameter s^2: $\sigma_\nu^2 \sim Inv-\chi^2(v_\nu, s_\nu^2)$, $\sigma_\kappa^2 \sim Inv-\chi^2(v_\kappa, s_\kappa^2)$, $\sigma_{\beta RE}^2 \sim Inv-\chi^2(v_{\beta_{RE}}, s_{\beta RE}^2)$, $\sigma_{u,m}^2 \sim Inv-\chi^2(v_{u,m}, s_{u,m}^2)$. Non-informative priors on these variance parameters would require small values (close to zero) for v and s^2 hyperparameters. If variances of the stochastic terms ν_m, κ_m etc. are small (relative to the design-based variances), the convergence of the Gibbs sampler could be very slow. Apart from that, as Gelman et al. (2008) and Polson et al. (2012) suggest, the inverse-gamma (or inverse chi-squared) parametrization for variance parameters is often not robust and should be replaced by distributions for standard deviation parameters from the folded noncentral t-family, of which the half-Cauchy distribution is a special case. In order to solve both problems, Gelman et al. (2008) suggest applying a redundant multiplicative parametrisation. In this paper, this parameter expansion is applied to any stochastic term components for which non-informative priors are chosen: $\nu_m = \xi_\nu \tilde{\nu}_m, \kappa_m = \xi_\kappa \tilde{\kappa}_m$, etc., where ξ-terms are multiplicative scalar parameters, and the terms with a tilde are distributed like in (3.6) and (3.3), but with variances $\tilde{\sigma}_\nu^2, \tilde{\sigma}_\kappa^2$, etc., for which inverse-Gamma (or inverse chi-squared) priors can be chosen. Such parametrization is used for standard deviation parameters that are expressed as: $\sigma_\nu = |\xi_\nu|\tilde{\sigma}_\nu, \sigma_\kappa = |\xi_\kappa|\tilde{\sigma}_\kappa$, etc. The ξ-terms are independent of each other and are assigned normal priors:

$$\xi_\nu \sim N(\alpha_\nu, \gamma_\nu),\ \ \xi_\kappa \sim N(\alpha_\kappa, \gamma_\kappa),\ \ \xi_{\beta RE} \sim N(\alpha_{\beta RE}, \gamma_{\beta RE}),\ \ \xi_u \sim N(\alpha_u, \gamma_u). \tag{3.9}$$

Setting $\alpha = 0$ (and $\gamma = 1$, without loss of generality) implies that priors on the standard deviation parameters come from the half t-family, see Gelman (2006). Setting $\alpha = 1$ and $\gamma = 0$ is equivalent to the original (non-expanded) parametrization of the model. Combining $\xi \sim N(0, 1)$ with 1 degree of freedom in $\tilde{\sigma}^2 \sim Inv-\chi^2(1, s^2)$ results in a half-Cauchy prior for parameter σ. For numerical reasons, the scale parameters s^2 in this application are restricted to the standard deviation of the variable of interest in vector $\boldsymbol{Y}$.

The overall intercept c, linear trend coefficient κ and expectations β_k^{RE} of random level break coefficients are estimated as regression coefficients and are contained in $\boldsymbol{\beta}$, along with vectors $\boldsymbol{\beta_k^{FE}}$. The prior for all the regression coefficients is flat (normal with mean $\boldsymbol{\beta_0} = \boldsymbol{0}$ and a large variance $\boldsymbol{\Omega_{\beta_0}}$).

Denoting the parameter vector by $\boldsymbol{\psi}$:

$\boldsymbol{\psi} = (\boldsymbol{\nu}, \boldsymbol{\kappa}, \boldsymbol{\beta^{RE'}}, \boldsymbol{u}, \boldsymbol{\beta}', \sigma_\nu^2, \sigma_\kappa^2, \sigma_{\beta RE}^2, \sigma_u^2, \xi_\nu, \xi_\kappa, \xi_{\beta RE}, \xi_u)$,

the likelihood function can be written as:

$$p(\boldsymbol{Y}|\boldsymbol{\psi}) = N(\xi_\nu \tilde{\boldsymbol{\nu}} \otimes \boldsymbol{\iota}_{[T]} + \xi_\kappa \tilde{\boldsymbol{\kappa}} \otimes \boldsymbol{t} + \boldsymbol{X^{RE}\beta^{RE}} + \boldsymbol{X\beta} + \xi_u \tilde{\boldsymbol{u}}, \boldsymbol{\Phi}) \tag{3.10}$$

where $\boldsymbol{t}$ denotes a vertical vector with time indicators $(0, 1, \ldots, T-1)'$, $\boldsymbol{\iota_{[T]}}$ is a T-dimensional column vector of ones, $\boldsymbol{\Phi}$ is a $[MT \times MT]$ matrix, matrix $\boldsymbol{X^{RE}}$ is $[MT \times MK_{RE}]$ and contains dummy regressors for the vector with random level break effects $\boldsymbol{\beta^{RE}} = (\boldsymbol{\beta_1^{RE'}}, \ldots, \boldsymbol{\beta_{K_{RE}}^{RE'}})'$, and $\boldsymbol{X}$ is a $[MT \times (2 + K_{RE} + d^{FE})]$-dimensional matrix, d^{FE} being the number of level break coefficients modelled as fixed effects. The dimension of matrix $\boldsymbol{X}$ is due to the presence of the overall effects c, κ, and K_{RE} β_k^{RE}s in the vector $\boldsymbol{\beta}$, as well as due to the presence of the d^{FE}-dimensional vector $\boldsymbol{\beta^{FE}}$ of fixed level break coefficients.

The parameters in the prior distribution below are assumed to be mutually independently distributed. The joint prior is then a product of each parameter's marginal prior distribution:

$$p(\boldsymbol{\psi}) = \left[\prod_{l=\nu,\kappa,\beta^{RE}} \left[\prod_{m=1}^{M} N(0,\tilde{\sigma}_l^2)\right]\right]\left[\prod_{m=1}^{M} N(\boldsymbol{0},\tilde{\sigma}_{u,m}^2\boldsymbol{A})\right] \times$$
$$\left[\prod_{\substack{l=\nu,\kappa,\\ \beta^{RE},u}} Inv-\chi^2_{\sigma_l^2}(v_l,s_l^2)\right]\left[\prod_{\substack{l=\nu,\kappa,\\ \beta^{RE},u}} N_{\xi_l}(\alpha_l,\gamma_l)\right] N(\boldsymbol{\beta_0},\boldsymbol{\Omega_{\beta_0}}). \tag{3.11}$$

Then the conditional posterior density of the parameter vector $\boldsymbol{\psi}$ is proportional to the following joint density: $p(\boldsymbol{\psi}|\boldsymbol{Y}) \propto p(\boldsymbol{\psi})p(\boldsymbol{Y}|\boldsymbol{\psi})$. See Appendix A for each parameter's conditional posterior distribution used in the Gibbs-sampler.

4 Structural Time Series (STS) Unobserved Component Modelling in the Case of the DTS

4.1 STS Model Specification

The general theory on STS models is presented in Durbin and Koopman (2012) and Harvey (1989). Similarly to the multilevel framework presented in Section 3.1, the series of the DTS design estimates $Y_{m,t}$ in a STS model can be decomposed into signal $\theta_{m,t}$ and the survey error component, as in (3.1). The $\theta_{m,t}$-term, in turn, is decomposed into several unobserved components:

$$\theta_{m,t} = L_{m,t} + \delta_{1,t}\beta_{m,1} + \dots + \delta_{K,t}\beta_{m,K} + \varepsilon_{m,t}, \quad t \in \{1,\dots,T\}, \tag{4.1}$$

where $L_{m,t}$ is the stochastic trend component, K is the total number of level breaks in domain m, and $\delta_{k,t}$ is an indicator variable for the k-th level-break regression coefficient $\beta_{m,k}$. The population parameter error term is $\varepsilon_{m,t} \overset{iid}{\sim} N(0,\sigma_\varepsilon^2)$. In cross-sectional surveys like the DTS, it is difficult to separate this term from the sampling error $e_{m,t}$, especially if the variance of $\varepsilon_{m,t}$ is small relatively to the sampling variance. Therefore, the two terms are combined into one composite error term $v_{m,t}$ that is assumed to be largely dominated by the sampling error. In order to incorporate the design-based variance estimates $Var(Y_{m,t})$ in a STS model, the composite error term $v_{m,t}$ can be modelled as $\tilde{e}_{m,t}\sqrt{Var(Y_{m,t})}$, $\tilde{e}_{m,t} \overset{iid}{\sim} N(0,\sigma^2_{\tilde{e}_m})$. For the variance of this product to be close to the design-based variance estimate $Var(Y_{m,t})$, $\sigma^2_{\tilde{e}_m}$ should be close to unity. Deviations from unity should correct for a possible under- or overestimation of design-based variance estimates. In this way, STS models feature more flexibility, unlike the multilevel ones, where the survey error variance estimates, used as prior input into the model, are assumed to be true and are thus fixed.
For one domain m, the model in (4.1) is referred to as a univariate STS model. If several domains have to be estimated simultaneously so that they can borrow cross-sectional sample information from each other, univariate models can be stacked under each other to constitute a multivariate STS model. In this application, no spatial correlation is assumed, just like in the multilevel model presented in the previous section.

The obtained multivariate STS model largely resembles the multilevel one in (3.1), with each domain being represented by the following equation:

$$Y_{m,t} = L_{m,t} + \delta_{1,t}\beta_{m,1} + \dots + \delta_{K,t}\beta_{m,K} + \tilde{e}_{m,t}\sqrt{Var(Y_{m,t})}, \quad t \in \{1, \dots, T\}. \tag{4.2}$$

As mentioned in Section 3.1, the smooth-trend model is assumed for the trend in the multilevel setting. Here, the local linear trend model and a random walk with a drift both resulted in a severe data overfit (see Harvey (2001) or Durbin and Koopman (2012) for different trend model specifications). Therefore, the same trend model is assumed for the structural time series (STS) model setting as in (3.3), though without the identification constraints for the multilevel setting mentioned in (3.4). The trend component $L_{m,t}$, apart from the random-walk component $u_{m,t}$ with its deterministic part $t(\kappa + \kappa_m)$, also implicitly contains an intercept. However, this intercept is defined solely on the basis of the corresponding domain's input series $Y_{m,t}$ and therefore would be equivalent to intercept $c + \nu_m$ in (3.2) only if the latter is modelled as a fixed, rather than random effect. The covariance matrix of the disturbance terms belonging to the trend $L_{m,t}$ is diagonal as in (3.5).

4.2 STS Model Estimation Details

Linear structural time series models with unobserved components are usually put into a state-space form and analysed with the Kalman filter. First, the model hyperparameters (here, σ_u^2s and $\tilde{\sigma}_{m,t}^2$s) are estimated using the maximum-likelihood (ML) approach by iteratively running the Kalman filter. The hyperparameters are treated as known and set equal to their ML-estimates, whereafter state variables ($L_{m,t}$, $\beta_{m,k}$s) can be extracted by the Kalman filter recursions. These recursions are initialised with diffuse priors for non-stationary variables (see Koopman (1997)). One has to be aware of the fact, that mean square error (MSE) estimates of the state variables produced by the Kalman filter are negatively biased, since the uncertainty around the hyperparameter estimates is not taken into account. See Pfeffermann and Tiller (2005) for the true MSE estimation approaches, as well as Bollineni-Balabay et al. (2015) for a simulation-based comparison of different approaches existing in the literature. The STS models presented in this paper are estimated in OxMetrics 7 (Doornik (2007)) in combination with the *SsfPack 3.0* package (Koopman et al. (2008)). One could think of a full Bayesian approach to STS model estimation with prior distributions for the model hyperparameters instead of hyperparameter ML estimates. However, the computational capacity required for models like the DTS one would make this approach unfeasible. Apart from that, since most practitioners, including Statistics Netherlands, estimate STS models in the frequentist framework, the authors aim to find if the quality of frequentist STS model-based estimates is sufficiently good compared to when estimation is carried out within the full Bayesian multilevel framework.

5 Tackling unreliability and missing values in design-based variance estimates of the DTS

The problem with the DTS is that the design-based variance estimates are missing in 2004-2009. For 2010-2013, they are only available at the intersections of provinces with modalities and of modalities with motives at the national level. The missingness in all domains in 2004-2009 can be imputed with the help of the Kalman filter, as shown a bit later in this subsection. As for the variances missing only at certain intersections at the end of the sample (2010-2013), these may be be approximated by using the variances available at the other intersections. The following approach is applied in order to approximate the standard errors ($SE(Y_{m,t})$) at the three-dimensional intersection:

$$SE(Y_{Mod,Mot,p,t}) = SE(Y_{Mod,Mot,t}) \cdot SE(Y_{Mod,p,t})/SE(Y_{Mod,t}), \tag{5.1}$$

where the second factor in the right-hand side reflects an inflation of the standard error of the GREG estimator (further referred to as the design-based standard error) in modality Mod when switching from the national to the provincial level. The performance of this method is verified using the 2003 data and is strikingly good for every province for the domain intersection in the left panel of Fig. 5.1 (here, relative margins of error are plotted, i.e. the margins of error at the 95% confidence level divided by the point-estimate). In most other domain intersections, the approximation performs similarly. The right panel shows how the approximation performs in the worst case.

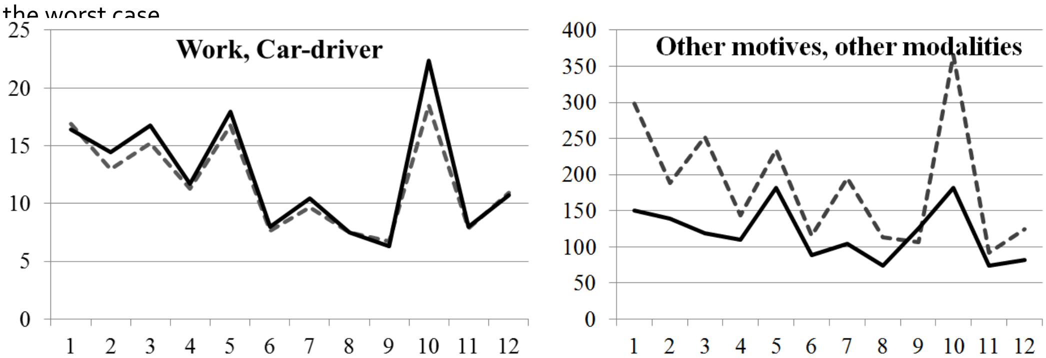


Figure 5.1 Original (solid line) and approximated (dashed line) relative margins of error at the 95% confidence level for 12 provinces, 2003, percent

The Kalman filter (KF) is applied to impute the missing design-based standard error estimates and to smooth out sampling fluctuations. Therefore, *smoothed* trends of the design-based standard error estimates from the following univariate STS model are used as input information in multilevel and STS models for the domain of interest, i.e. km-pppd:

$$\begin{aligned} SE(Y_{m,t}) &= L^{SE}_{m,t} + \varepsilon_{m,t}, \quad \varepsilon_{m,t} \overset{iid}{\sim} N(0,\sigma^2_{\varepsilon}), \\ L^{SE}_{m,t} &= L^{SE}_{m,t-1} + R^{SE}_{m,t-1}, \\ R^{SE}_{m,t} &= R^{SE}_{m,t-1} + \eta_{m,t}, \quad \eta_{m,t} \overset{iid}{\sim} N(0,\sigma^2_{\eta}), \quad t \in \{1, \dots, T\}, \end{aligned} \tag{5.2}$$

where $L^{SE}_{m,t}$ and $R^{SE}_{m,t}$ are the level and slope of the trend for the design-based standard errors, respectively. Although standard deviations of normally distributed variables are Chi-distributed, such Chi-distributions can be approximated with a normal distribution given moderate to large numbers of degrees of freedom. Model 5.2 could also be augmented with sample size information. However, in this case such augmentations resulted in data overfit. Some of the smoothed SE estimates are plotted against their original counterparts in Fig. 5.2.

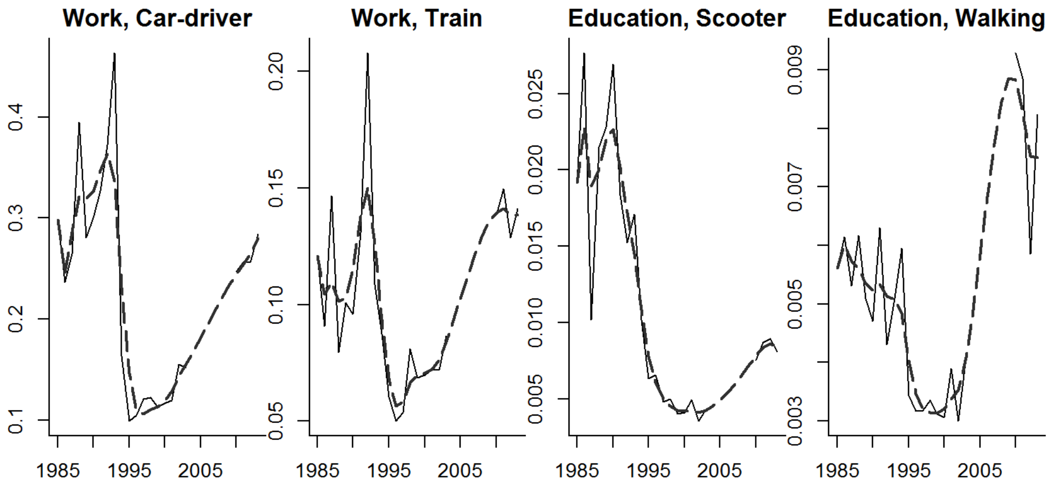


Figure 5.2 Original (solid line) and smoothed (dashed line) design-based standard errors, Zuid-Holland.

If these smoothed estimates are used as an input in multilevel models (and thus treated as the known true standard errors of the design estimator), some domains still experience too volatile multilevel model point-estimates (see Fig.5.3, solid lines). This occurs only in domains where the number of km-pppd is small (further referred to as small-scale domains, e.g., "Walking", "Bus/Tram/Metro") and makes one suspect that the design-based variance estimates are biased in these domains. For this reason, one could consider using the smoothed design-based standard errors scaled within the univariate STS framework. This means that the sampling error variances would be represented both in the STS and multilevel models by the product of the scaling parameter $\sigma^2_{\tilde{e}_m}$ and the smoothed design-based variance estimates $Var(Y_{m,t})$. Fig. 5.3 shows point-estimates resulting from the two approaches applied to the same multilevel model. The differences in other domains are negligible because most of these scaling parameters are concentrated around unity, as Fig. 5.4 shows. Further in this paper, all the multilevel analysis will be based on the second approach which applies scaled smoothed design-based variance estimates. Not only does it prevent overfitting in domains whose design-based variance estimates are not reliable, but it also eliminates another factor responsible for differences between the outcomes from multilevel and STS models, because in this way both models use design-based variance estimates corrected in the same way.

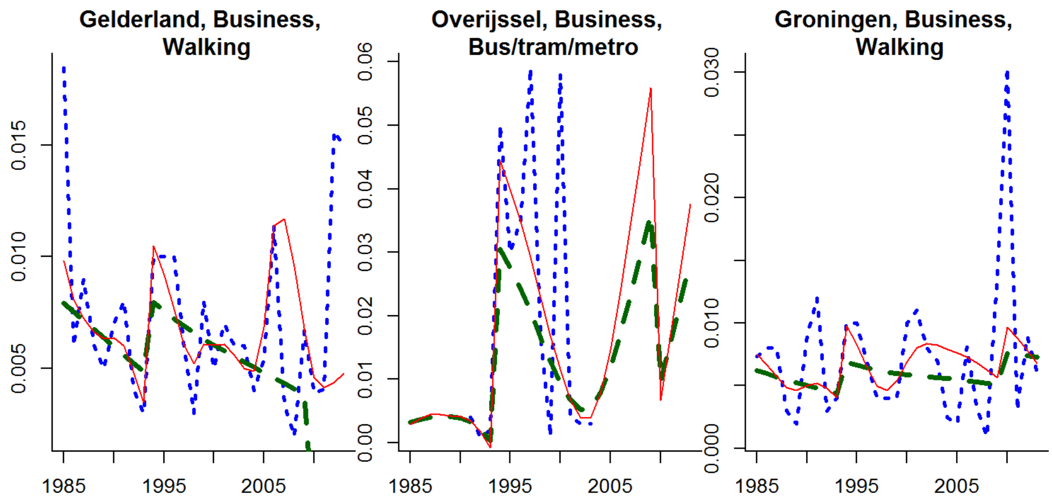


Figure 5.3 Design-based point-estimates (dotted lines) and multilevel model posterior means that use smoothed design SE estimates (solid line) and smoothed design SE estimates scaled within the univariate STS framework (dashed line).

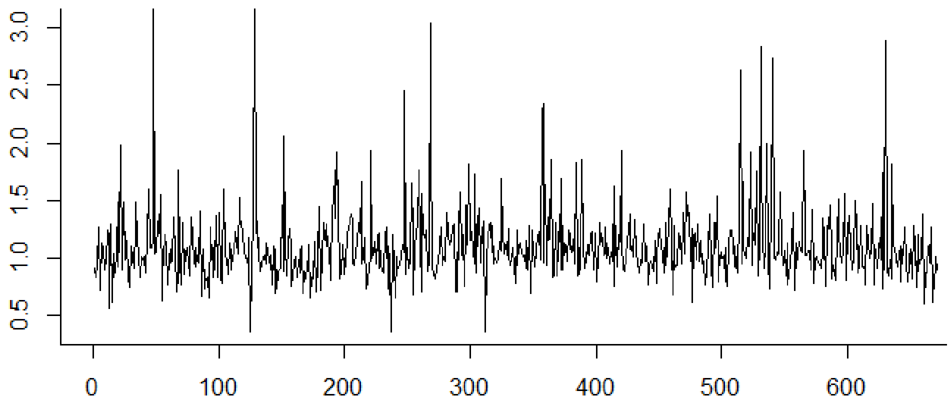


Figure 5.4 Scaling factor for design-based standard errors in univariate STS models.

6 The DTS at the Provincial Level

Before we estimate both models, a careful inspection for outliers is needed. The trend disturbance variance $\sigma^2_{u,m}$ may get overestimated due to one single outlier, especially if the design-based variance of this outlying point is negatively biased. Excessively large trend disturbance variances would cause model point-estimates to overfit the data. In order to eliminate the effect of outliers, we inspect estimation results from univariate STS models for each domain and assign a large value to the design-based variance of every outlying point. There are five outliers in total at the provincial level, which belong to small-scale domains involving ”Scooter”, ”Train” or ”Other modalities”. These outliers are treated in the same way in the multilevel and STS models.

6.1 The Multivariate Variance Structure of Stochastic Trends

Small areas (domains) at the provincial level are formed by an intersection of 12 provinces with 8 modalities and 7 motives ($M = 672$). Subsection 3.1 presented a general variance structure of the trend disturbances. To see if some domains with certain irregularities can benefit from other domains by means of pooling by either provinces, motives or modalities, it is best to compare the trend disturbance variance estimates $\hat{\sigma}_{u,p}$ obtained from univariate STS models (4.2). By looking at the top panel of Fig. B.1 with box-plots for these $\sigma_{u,m}$-estimates, it becomes clear that they are very heterogeneously distributed across the modalities. With heterogeneity less pronounced across the motives (the lower panel), there is no systematic way in which some motives resemble the others, and the existing differences are still too big to be able to pool the trend disturbance variances across the motives.
Viewing provinces as panel units with responses on $\sigma_{u,m}$ of motives and modalities provides an opportunity to use larger amounts of sample information in estimating the hyperparameters. This would make the hyperparameters estimates more reliable if the processes in provinces are similar indeed. In order to check that, one could construct scatter- or barplots of $\sigma_{u,p,j,i}$-estimates from univariate STS models for the twelve provinces for every intersection of the motives and transport modalities (find the scatterplot and barplots in Fig. 6.1 and B.2, respectively). These plots reveal that $\sigma_{u,p,j,i}$-estimates of provinces 2 (Friesland) and 5 (Flevoland) often exhibit very different values from the rest of the provinces - at eight and seventeen Mot/Mod intersections, respectively. Large $\sigma_{u,m}$-values in province 5 (e.g., in Mod 1/Mot 1-3,6, Mod 2/Mot 3, Mod 4/Mot 1) are mostly caused by a bigger scale of the series, i.e. with more km-pppd being covered in this province compared to the other provinces. As for Province 2 (e.g., in Mod 2/Mot 1, Mod 5/Mot 1, Mod 8/Mot 6), large $\sigma_{u,m}$-values are caused by either a bigger scale, or a more volatile pattern of the series. $\sigma_{u,p,j,i}$-estimates of province 1 exhibit the largest values (compared to the rest of the provinces) for about five times, and those of province 10 and 12 - four and three times, respectively. It is worthwhile to observe the scale of the latter provinces' series (1, 10 and 12): it hardly ever exceeds the scale of the rest of the provinces, and is sometimes even smaller. Therefore, these provinces would be better off if pooled together with the other ones. Further, one can see that the disturbance variances in many cases tend to take on very small values when estimated within the STS approach. It means that the trend in such cases resembles a straight line. Keeping in mind that the ML estimator tends to underestimate hyperparameters when their distribution is right-skewed, it would also be desirable to estimate such variance hyperparameters in a pool with other domains (provinces in this case). Thus, we abstain from pooling provinces 2 and 5 with the other ten. Therefore, the twelve provinces are divided into a cluster of two and a cluster of ten provinces, within which variances are pooled. This applies to the multilevel, as well as to the multivariate STS analysis.

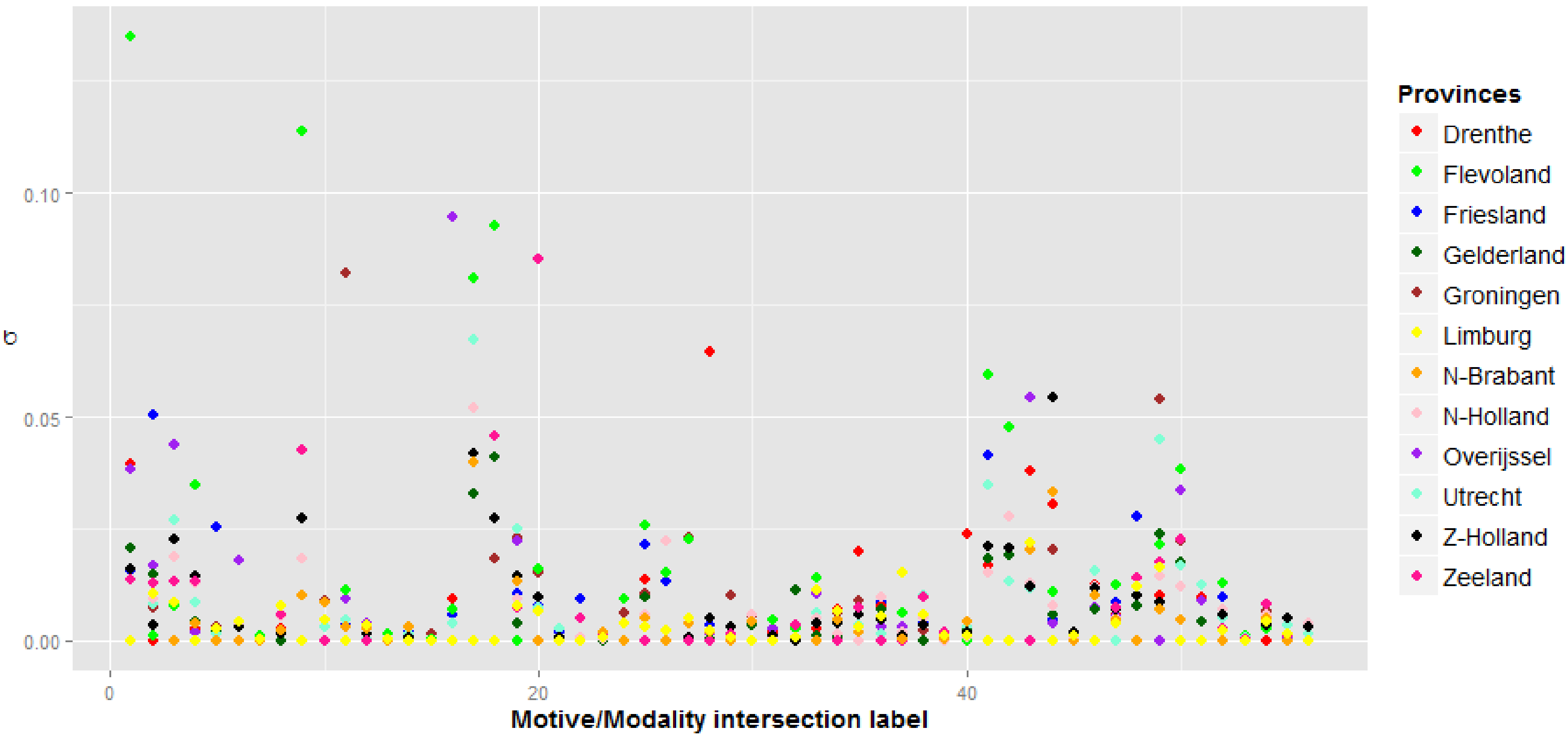


Figure 6.1 Provincial σ_u from univariate STS models for all 56 intersections of motives and transport modalities.

The DTS point-estimates are very similar when comparing the univariate and multivariate STS settings with each other. There are, however, a few domains that exhibit significant differences (depicted in Fig. 6.2). It is worth considering the consequences that pooling may have on model estimates. If one province turns out to be very different from the others it has been pooled with, then an underestimate of the stochastic trend variance can force the model-based point-estimates to take strange slopes, for example, when κ_ms are not sufficiently scattered due to a small σ_κ^2 (as, e.g., in Noord-Holland/Shopping/Car-passenger). The advantage of pooling $\sigma_{u,m}$ across provinces is that certain idiosyncracies occurring only in one province can be eliminated. If there is, for instance, no real factor behind the sudden surge in point-estimates of the Limburg/Business/Bus-Tram-Metro domain, pooling could be considered in order to get rid of such irregularities. Further, excessive volatility in the univariate model point-estimates of one province due to underestimation of design-based variances in this province can be overcome by borrowing information from other provinces for an identical intersection of motives and transport modalities (as in Zuid-Holland/Shopping/Scooter).

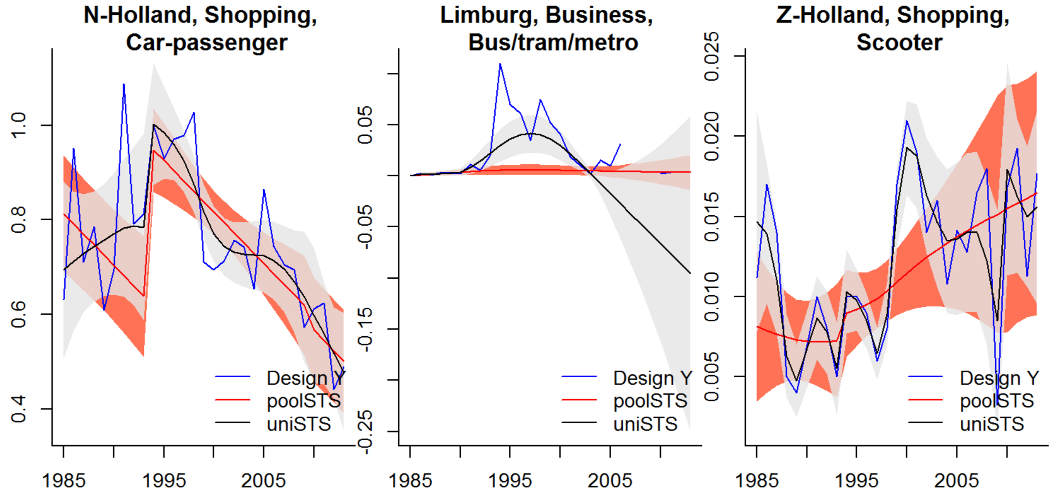


Figure 6.2 STS model-based point-estimates obtained from univariate STS models (black line) and from a multivariate STS model where σ_u^2s are pooled across provinces (coral line); the colour of the confidence intervals corresponds to the colour of the line.

6.2 Multilevel Model Estimation Results

Several model formulations have been considered for the multilevel framework. These formulations have already been mentioned at the end of Section 3.1 and are summarised in Table 6.1. Model selection is based on the deviance information criterion (DIC) (see Spiegelhalter et al. (2002)) which is a generalisation of the well-known AIC and BIC for hierarchical models. Adequacy of point- and variance estimates is also taken into account as an informal criterion. Table 6.1 also contains information on the minimum sufficient number of iterations, burn-in iterations, as well as the thinning interval h. The latter means that only every h-th draw from the parameter posterior distribution is saved in the MCMC chain. Thinning helps overcome memory constraints when a chain is strongly autocorrelated (see Gelfand and Smith (1990)). Further, in order to ensure that the stationary distribution of a parameter has been reached, more than one chain is needed, each chain starting with draws from an overdispersed distribution (Gelman and Rubin (1992)). Three parallel chains are constructed in this application for each parameter. We use the Gelman-Rubin convergence statistic for multiple chains (R-hat) to be able to judge about the sufficient length of a burn-in period and about whether the chains have mixed well, which allows us to arrive at a reasonable estimate of the true posterior distribution.

Table 6.1 Multilevel models considered for the DTS at the provincial level

Model labels	Year and domain indication for level breaks	Itera-tions	Burn-in	Thinning	DIC
HB-bRE	β^{RE}: 1994, 2010 for all m; β^{FE}: 2004 for m in Mot 6,7	75000	50000	50	-43668
HB-bFE	β^{FE}: 1994, 2010 for all m; 2004 for m in Mot 6,7	55000	30000	50	-43606
HB-FE	β^{FE}: 1994, 2010 for all m; 2004 for m in Mot 6,7; $(c + \nu_m)$, $(\kappa + \kappa_m)$ as FE	5000	1000	10	-43859

HB-bRE stands for a model with random effects as in (3.1)-(3.2). HB-bFE differs from HB-bRE in that all β_m-coefficients are estimated as fixed effects. HB-FE differs from HB-bFE in that ν_m and κ_m terms are estimated as fixed effects, too. Different model modifications of the kind in (3.7) have been tried for HB-bRE and HB-bFE models. However, these modifications either resulted in a numerical failure (in model HB-bRE variations), or failed to secure multiple chain convergence in some variance parameters ($\sigma^2_{\kappa_{Mod}}, \sigma^2_{\nu_{Mot}}$) in model HB-bFE variations. The only model modification that could be fitted successfully was the HB-bFE model with an additional $\boldsymbol{\nu_{Mod}}$-term. However, this specification did not result in model improvement, with the DIC value being even lower than that of the basic HB-bFE model.

Fig. 6.3 shows point-estimates for the three multilevel models mentioned in Table 6.1. While most domains get comparable point-estimates from the three models, a combination of "Car-driver" with "Work" and "Business" motives features some most striking problems inherent to models HB-bRE and HB-bFE. It is worth mentioning that the 1994-break is largest exactly at the intersection of modality "Car-driver" with motive "Busines". Only five provinces are depicted, but the results are similar for the rest of the provinces. As the second row of Fig. 6.3 demonstrates, the model with level breaks modelled as random effects is incapable of fitting the level breaks for the year 1994. Neither of the HB-bRE modifications can cure this problem (some of them even result in less adequate level break estimates for 1994). The level break for the years 2010-2013 is not captured by the HB-bRE model in these domains either. Apparently, the assumptions about common distributions for β^{RE}_ms are violated in the case of the DTS. The first row of Fig. 6.3 suggests that the assumptions about common distributions for κ_ms and/or for ν_ms are often not applicable either.

The model HB-bFE exhibits a greatly improved fit. However, the fit for the years 1985-1994 is not optimal for the "Work"-motive (see the first row). Apart from that, the level of the signal in "Business" seems to be too low (see the second row of Fig. 6.3), suggesting that the common distributions imposed on the random intercepts do not allow the ν_m-terms to take sufficiently large values.

Finally, the fit of the model with only fixed effects (HB-FE) seems to be most adequate. This model also gets the lowest DIC-value, see Table 6.1.

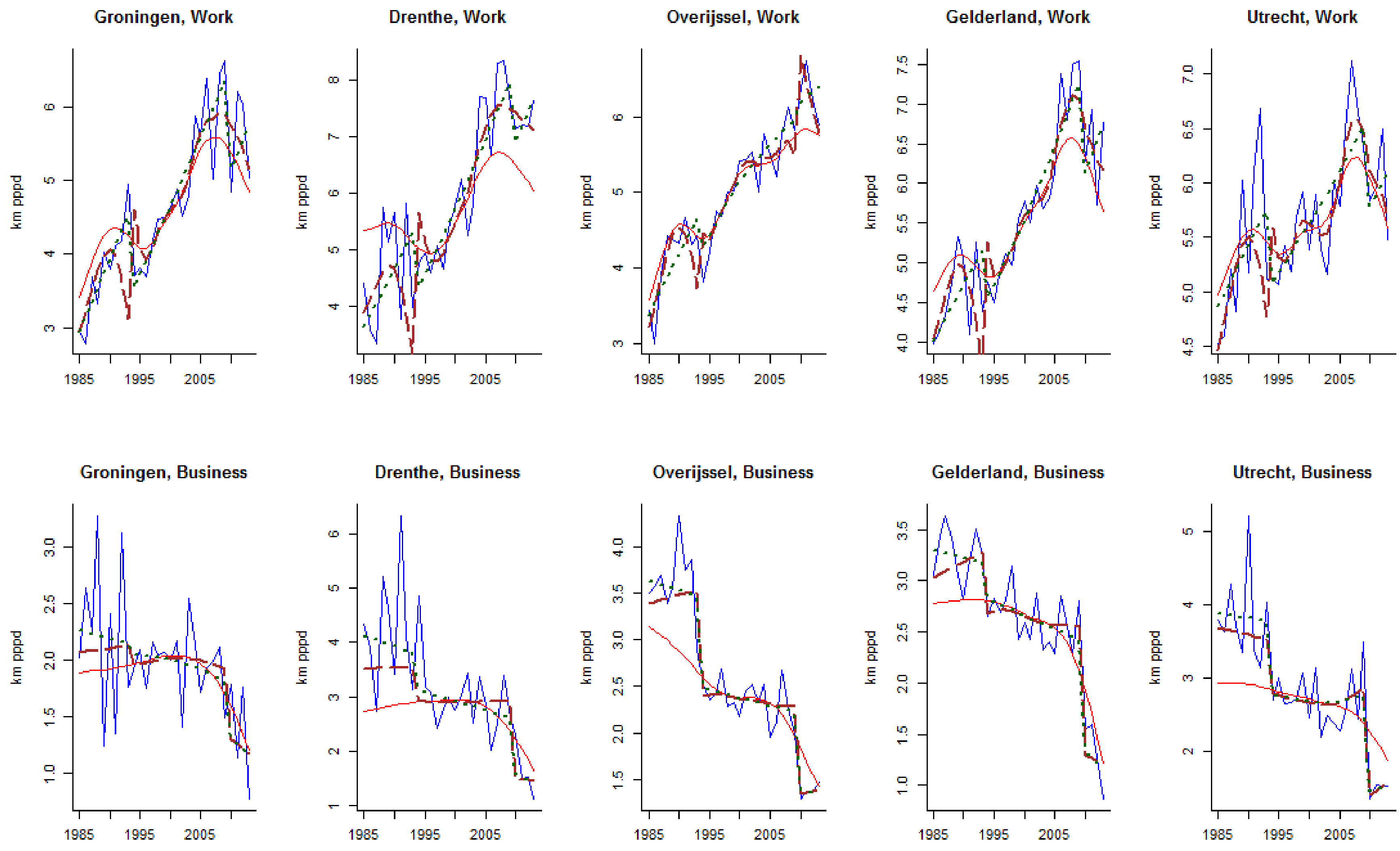


Figure 6.3 $Mod1$ **(Car-passenger) point-estimates produced by the design-based estimator (blue solid line) and HB-bRE (red solid line), HB-bFE (dashed line), HB-FE (dotted line) models**

6.3 Multilevel and STS Model Estimation Results Compared

Not only does the multilevel model HB-FE turn out to be the best one out of the multilevel models considered in this paper, but it also offers a more straightforward comparison with STS models, since intercepts and regression coefficients in the latter type of models are in fact fixed effects. However, the performance of the two modelling approaches in terms of signal (trend etc.) variance estimates cannot be fairly compared at this stage. There are at least two reasons for that. We will try to quantify the effect of each of them shortly.

The first reason is the fact that the true (unknown) hyperparameters of a STS model (in this case, σ_u^2 and $\tilde{\sigma}_e^2$) are replaced by their maximum likelihood estimates and are treated as known. Within the HB approach, the uncertainty about these hyperparameter estimates is summarised by the variance of their posterior distribution. The signal MSE bias in STS models due to hyperparameter uncertainty can be negligible if the distribution of the ML estimator is symmetric and well concentrated around its mean. But, if the series is short or if the hyperparameters are close to their boundary values, the ML estimator distributions of such hyperparametes (and their posteriors too) are not symmetric. In this case, the uncertainty around the hyperparameter estimates could be very large, resulting in a considerable signal MSE bias in STS models. Therefore, lower signal variance estimates delivered by an STS model, in comparison with posterior signal variances, do not necessarily mean that the STS approach is superior to the HB one. In fact, the extent to which HB model-based SE estimates exceed those from a comparable STS model gives an indication for the scale of the MSE negative bias in such STS model. The negative bias in STS models can be accounted for by the bootstrap method of Pfeffermann and Tiller (2005). Bollineni-Balabay et al. (2015) show why this method is superior to other existing methods for MSE estimation in STS models.

Another source of differences when comparing the two modelling approaches may appear when the signal posterior distributions are skewed, thus producing larger posterior standard deviations for domain predictors than the signal standard errors produced by a linear STS model. In fact, if the normality assumption about the data/disturbances is not satisfied, linear STS models like the ones considered in this paper should not be applied.

Point-estimates produced by the HB-FE and multivariate STS model are depicted in Fig. 6.4, together with the 95% confidence intervals from the STS model superimposed on the credible intervals from the HB-FE model. The latter are symmetric as they are constructed using the posterior standard deviations, rather than being quantile-based. In most domains, the posterior distribution of the domain predictors is symmetric. It is sometimes slightly skewed in small-scale domains. This is where differences in point- and variance estimates become visible. The STS model-based point-estimates tend to be smoother (as in the lower panel of Fig. 6.4), since they are based on smaller (close to the boundary space) values of the trend disturbance variances. The HB model-based point-estimates in such domains, in turn, stem from a set of draws from a heavy right-tailed distribution of the trend variance, which results in more flexible HB-based trends. Further, the differences in point-estimates from the STS and HB models are also partially due to the fact that posterior means of $\theta_{m,t}$, rather than medians, are taken as domain m predictors.

The signal standard error estimates produced by the two approaches could be compared and summarised in terms of relative deviation (RD) of the STS-derived signal SEs from the HB posterior standard deviations:

$$RD_{HB,t}^{STS} = \frac{SE_t^{STS} - SE_t^{HB}}{SE_t^{HB}} \cdot 100\%. \tag{6.1}$$

For the multivariate setting, the overall average of this measure across time and domains is equal to -9.4%, with a median of -3.1%. This itself does not imply serious differences between the outcomes of the two approaches, but shows that the distribution of the RD-terms is very skewed, with the mean being pulled to the left by extreme SE-differences in small-scale domains.

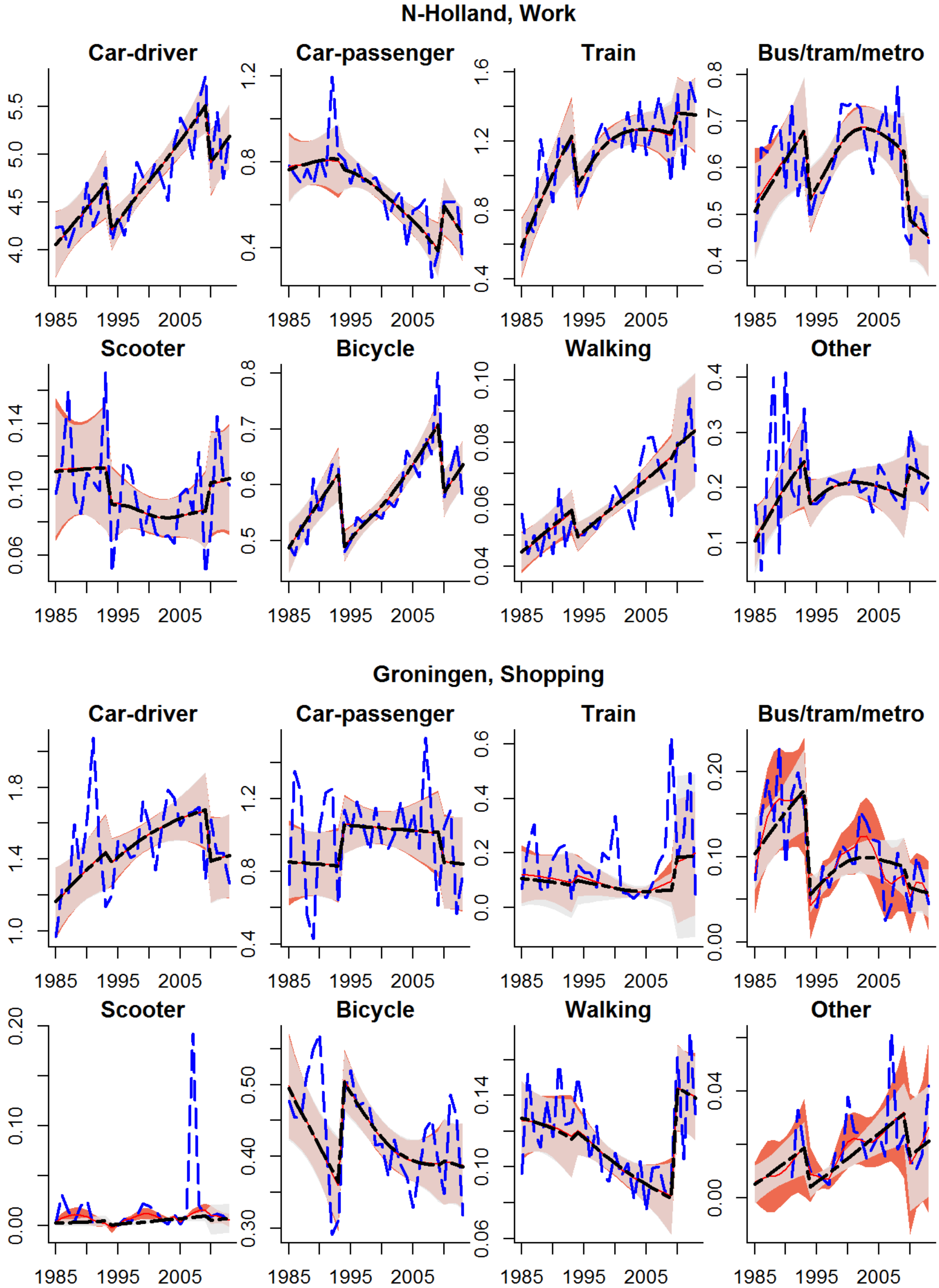


Figure 6.4 The provincial level design point-estimates (thin dashed line), point-estimates from the multivariate STS model (thick dashed) and posterior means from the HB-FE model (solid line), km-pppd; 95% confidence intervals from the STS model superimposed on the credible intervals from the HB-FE model, the colour corresponds to the colour of the point-estimates

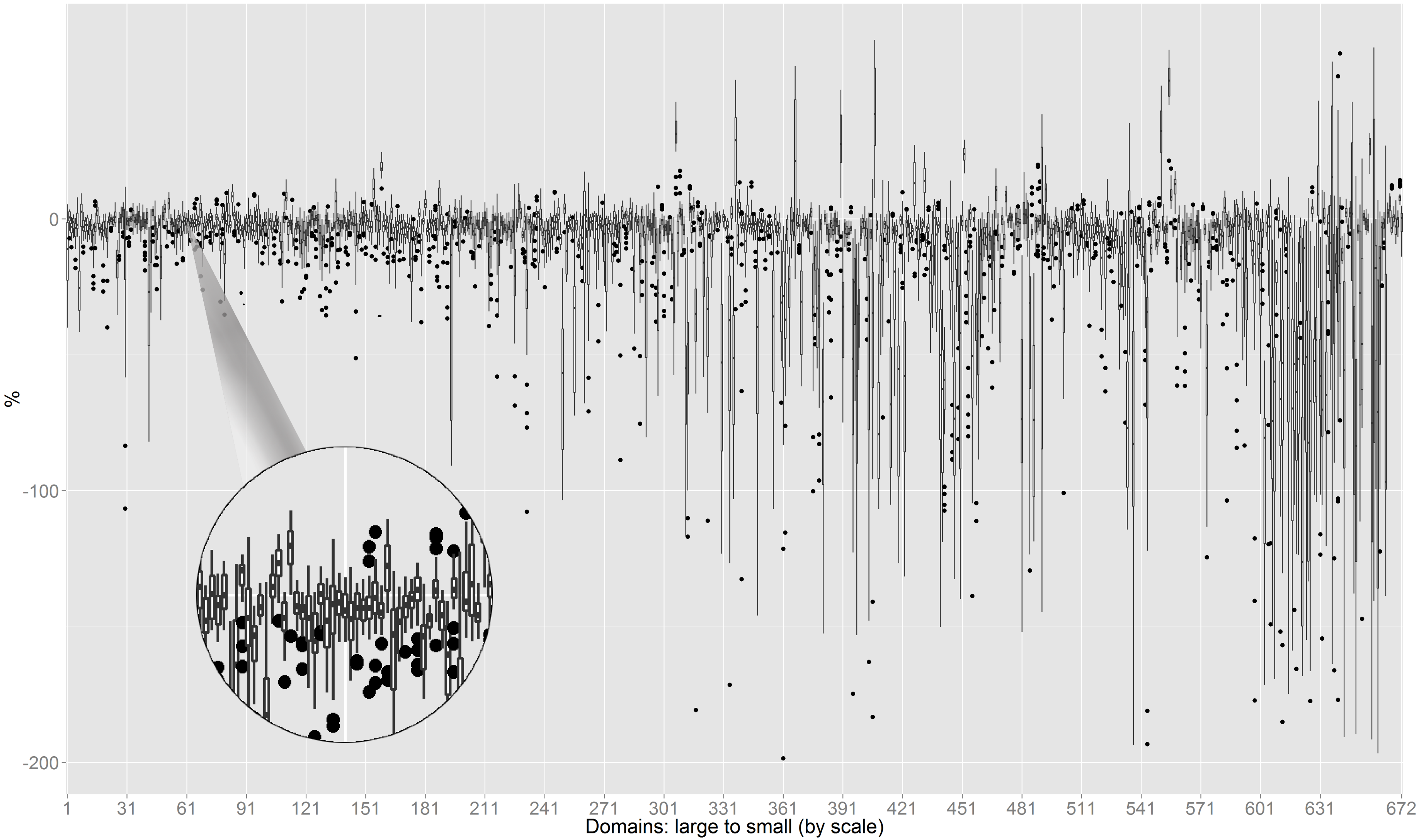


Figure 6.5 Box-plots for the provincial level $RD_{HB,t}^{STS}$ - relative deviation of signal SEs produced in the multivariate STS model from signal SEs produced by the HB-FE model, %.

It would be interesting to check if there is a tendency for the HB-based SEs to exceed the STS-based ones in small-scale domains, since such domains are suspected to feature skewed posteriors. By "scale" here is meant a time average of the domain's point-estimates produced by the multivariate STS model. In this way, the scale represents both the number of km-pppd and the effective sample size: the scarcity of respondents belonging to a certain motive/modality intersection translates into low per-person figures for kilometers travelled for that intersection. The 672 provincial domain numbers sorted by scale in the descending order are plotted on the x-axis of Fig. 6.5. It is clearly visible that, as the domain scale decreases, the STS signal standard errors tend to deviate more from the HB-FE posterior standard deviations in a negative direction. Most of the extreme deviations (those up to 200%) occur around level interventions. There, the HB-based signals obtain larger uncertainty than the STS-based ones.
Now we try to quantify the effect of hyperparameter uncertainty on the differences between the HB-FE-based and STS-based SEs. For that, an identical HB-FE model has been estimated (referred to as HB-FE-ML), for which informative priors on the trend disturbance variances have been set with a large number of degrees of freedom and the scale parameter taken equal to the ML estimates from the STS multivariate model. Relative deviations as in (6.1) have been calculated for the HB-FE and HB-FE-ML posterior standard deviations. Their average value suggests that skewness in the signal posterior distributions can be blamed for only -1.6% out of the above-mentioned -9.4% reduction/underestimation in the HB posterior standard deviations by the STS approach. The remaining -7.8% are due to the hyperparameter uncertainty around the trend disturbance variances, not accounted for in the STS approach.
It is of particular interest how much reduction in design-based standard errors can be obtained by time series modelling technique. The HB-FE model offers a 51% reduction in design-based standard errors at the provincial level on average, with a median of 54%. For the STS multivariate model, where provinces are pooled as described in Subsection 6.1, the mean and median percentage reduction are slightly bigger - 54% and 57%, respectively. These and the above-mentioned figures indicate that the HB-FE and STS approaches deliver very similar results, with sizeable differences appearing mostly in small domains due to neglected hyperparameter uncertainty in STS models.

7 The DTS at the National Level

7.1 The Multivariate Variance Structure of Stochastic Trends

The number of domains at the national level is $M = 56$, defined by an intersection of 8 transport modalities with 7 motives. As already described in Subsection 6.1, one can look at box-plots with the trend disturbance variance estimates obtained from univariate STS models in order to see if the parametrisation of the trend variance matrix Σ_u can be made more parsimonious. Similarly to the provincial level, $\sigma_{u,m}$s do not exhibit resemblance either across motives, or (and in particular) across the modalities, as can be seen in Fig. 7.1. The seeming resemblance between some of the box-plots (e.g., $Mod5$-$Mod8$) disappears when zooming in the plots. Endowing each domain with its own $\sigma_{u,m}$ at the national level limits the STS approach to the univariate setting. Within the multilevel approach, a multivariate structure hinges on the assumption of common distributions for the random effects in the HB-bRE and HB-bFE models. The HB-FE model at the national level constitutes a set of univariate HB time-series models.

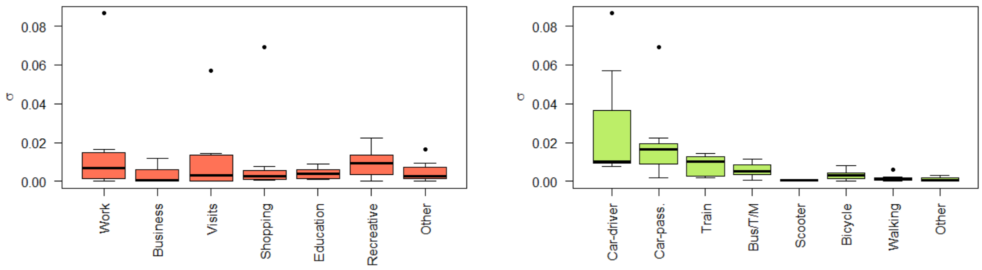


Figure 7.1 σ_u **from univariate state-space models at the national level**

7.2 Multilevel and STS Model Estimation Results Compared

For the national level series, the same multivariate model variations described in Subsection 6.2 are considered. According to Table 7.1, the HB-FE model has the lowest DIC-value, but the differences among the three HB model modifications are very small. As Fig. 7.2 shows, the HB-bRE model does not experience any difficulties with the fit, as was the case at the provincial level. Point-estimates from the HB-bFE, HB-FE and STS models almost coincide. Appendix C presents point-estimates and their credible/confidence intervals for the HB-FE and univariate STS models, as well as for the design-based estimator. The striking similarity between the HB-FE and STS model-based point- and variance estimates is less strong in small-scale domains, with the STS point-estimates being slightly smoother (e.g., in "Visits/Bicycle", "Visits/Other modalities", "Recreative/Bicycle", "Recreative/Other modalities").

Table 7.1 Multilevel models considered for the DTS at the national level

Model labels	Iterations	Burn-in	Thinning	DIC
HB-bRE	75000	50000	50	-8114
HB-bFE	55000	30000	50	-8166
HB-FE	5000	1000	10	-8189

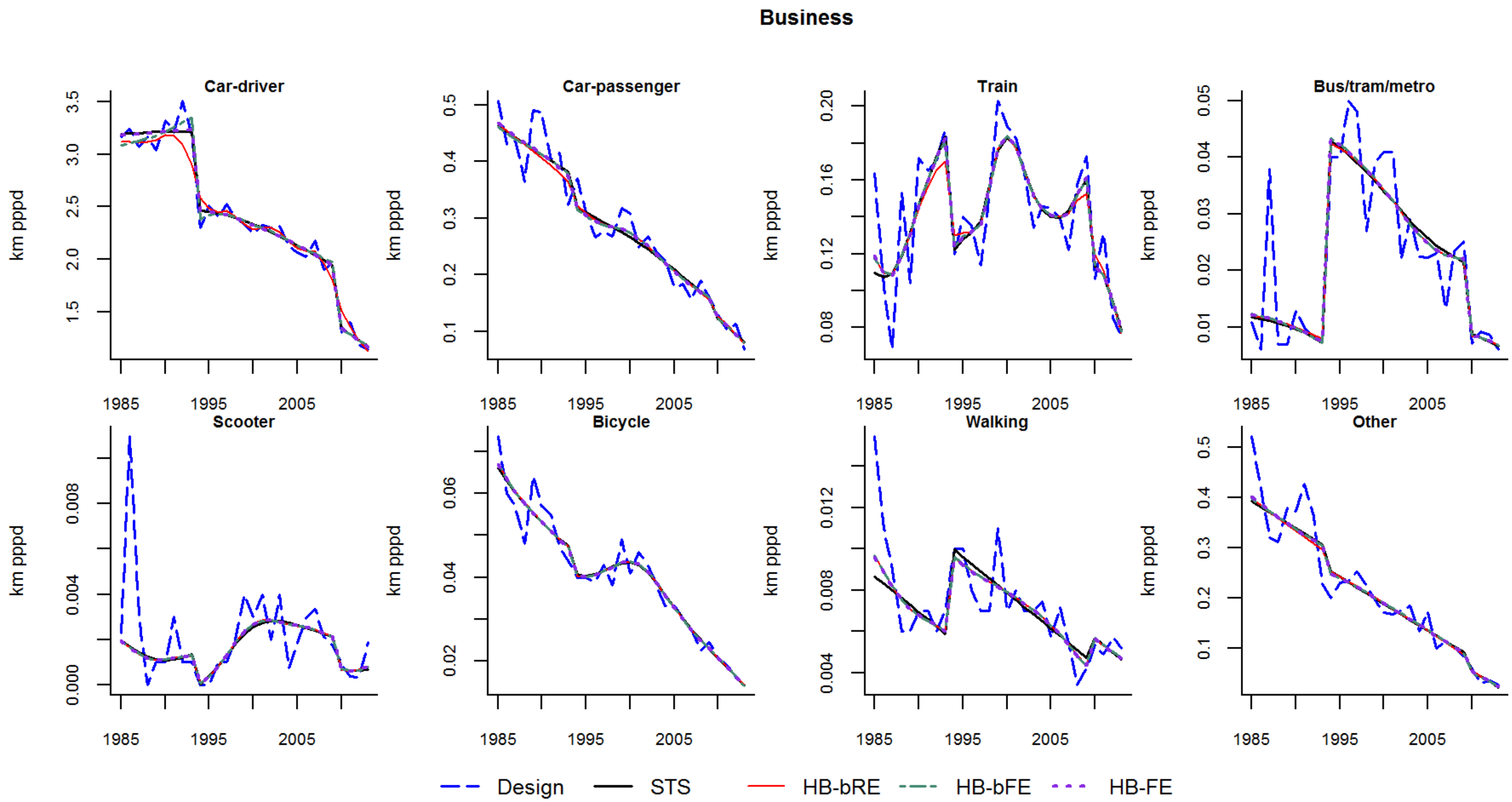


Figure 7.2 National level design- and model-based point-estimates, km-pppd.

For the comparison of the two approaches in terms of SEs, the reader is referred to equation (6.1). The overall average of $RD_{HB,t}^{STS}$-terms across time and domains at the national level is equal to -10.9%, and the median equals -9.1%. This negative mean is clearly visible in Fig. 7.3.

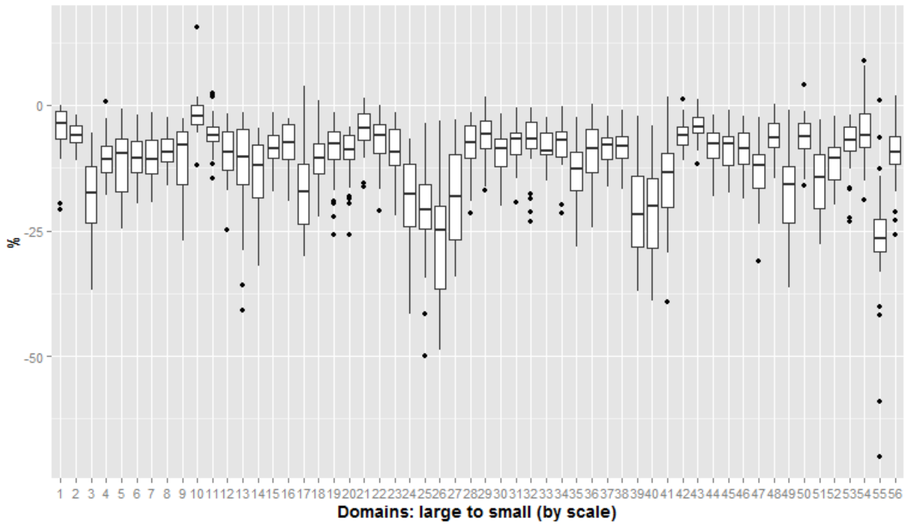


Figure 7.3 The national level $RD_{HB,t}^{STS}$ - relative deviation of STS SEs from HB-FE SEs for signals,%.

Though much less pronounced than at the provincial level, some association of low RD-terms with the decreasing domain scale can still be seen at the national level (Fig. 7.3). Indeed, extreme RD-values of more than 20% occur in several small domains (e.g., "Visits/Other modalities", "Visits/Bicycle"). Some of the largest domains still feature low RD-values (e.g., "Visits/Car-passenger", the third one from left). An inspection of the point-estimates of these domains in Appendix C shows that the low $RD_{HB,t}^{STS}$-values correspond to those domains where the signal point-estimates from the HB models are visibly more volatile than those produced by STS models. As with the provincial data, a similar HB-FE-ML model has been estimated at the national level. It turns out that the difference between the HB- and STS-based SEs is almost entirely due to the hyperparameter uncertainty around the trend disturbance variances (an average -10.1% out of the above-mentioned -10.9%).
Mean reduction in design-based standard errors with the model-based approach is less than at the provincial level, but still quite appreciable with the overall average of 31.7%, and the median of 34.3% for the HB-FE model, (39% and 41%, respectively, for the STS approach).

8 Discussion

Time series models are well known for their power in reducing design-based variances and in making point-estimates more stable, be it multilevel or structural time series models. Apart from that, time series techniques are indispensable when it comes to estimating level breaks due to survey redesigns. This paper aimed at establishing which of the two modelling approaches - STS or multilevel - should be preferred.
The multilevel model estimated with the Hierarchical Bayesian approach (HB model) is a time-series extension of the Fay-Herriot model. Apart from featuring hyperpriors for its parameters, it differs from the conventional STS model (in the sense of Harvey, Koopman, and

in the way it is used at Statistic Netherlands) in that time-invariant effects (e.g. intercepts, level break or linear trend coefficients) can be treated as random. In this application, however, assumptions about random effects sharing the same variance across domains turned out to be invalid at the provincial level. Therefore, several HB-model variations have been considered where (some of) these random components are modeled as fixed ones. Not only does the fixed-effect specification of the HB-model (HB-FE) make the comparison of the HB and STS approaches more straightforward, but it is also the only specification that provides an adequate fit to the data.
The comparison between the STS and the HB-FE model in terms of estimated signal variances is still not completely fair. First of all, this is due to the fact that the true (unknown) hyperparameters of a STS model are replaced by their maximum likelihood estimates and are treated as known. To account for this additional uncertainty, one would in addition have to resort to bootstrapping techniques, see Pfeffermann and Tiller (2005), Bollineni-Balabay et al. (2015). Within the HB-FE approach, the uncertainty around hyperparameter estimates is summarised by the variance of their posterior distribution. Therefore, lower signal variance estimates delivered by an STS model do not necessarily mean that the STS approach is superior to the HB-FE one. In fact, the extent to which HB-FE model-based variance estimates exceed those from a comparable STS model gives us an indication for the scale of the negative bias in the signal MSEs of such STS model. This bias could be particularly large if variance hyperparameters are close to their boundary values, with heavy right tails in the posterior distributions/distributions of the ML estimator. This is the case in many small-scale domains of the DTS, which makes the hyperparameter uncertainty the primary source of differences between the two modelling approaches in the case of the DTS. Secondly, another source of differences is skewed signal posterior distributions that produce larger posterior standard deviations for domain predictors than the signal standard errors produced by a linear STS model relying on the assumption of normality in the data/disturbances. In the DTS, slight skewness in the signal posterior is observed at the provincial level, mainly in those domains whose trend disturbance variance posteriors are skewed enough to feed some degree of asymmetry through to the signal posterior.
It turns out that both point- and variance model-based estimates produced with the STS and HB-FE techniques are very similar. Differences become visible mainly in small domains. Standard errors produced by the STS model are smaller than posterior standard deviations from the HB-FE model by 9.4% on average (across time and domains) at the provincial level (median 3.1%), and by 10.9% at the national level (median 9.1%). At the national level of the DTS, skewness turned out to have a negligible effect on the standard error difference between the two approaches, but at the provincial level it is responsible for about 1.5% out of the above-mentioned 9.4% of the HB-FE model-based posterior standard deviations on average. The rest is due to the hyperparameter uncertainty unaccounted for within the STS approach. For these above-mentioned reasons, one should be aware of negative biases in frequentist STS-based variances in short time series of series that feature small variance hyperparameters. In such cases, the negative biases should be accounted for by means of an additional procedure, such as the bootstrap of Pfeffermann and Tiller (2005).
As an important by-product, the results of this paper give an idea about how much reduction in design-based standard errors can be obtained by these time series modelling techniques. The mean reduction in design-based standard errors with the HB-FE model is 51% at the provincial level, with the median of 54%. In other words, in order to reduce the design-based variance to this extent, one would have to increase the sample size more than four-fold (conditional on the point-estimates). For the STS model, the mean and median percentage reduction in the design-based standard errors is slightly bigger - 54% and 57%, respectively, - due to the reasons

already mentioned above. Mean reduction in design-based standard errors at the national level is smaller than at the provincial one, but is still quite appreciable with the overall average of 31.7% and the median of 34.3% for the HB-FE model, and 39% and 41%, respectively, for the STS model.
Another aspect we look at is unreliable design-based variance estimates. First of all, these estimates are subject to sampling volatility. With the help of a simple STS model, design-based standard errors can be smoothed with the Kalman filter. These smoothed standard error estimates are further used as input in the multilevel and STS models. Secondly, design-based variance estimates could be biased in the case of small domains. If the bias is negative, for instance, then treating these estimates as the true variances in a multilevel setting results in model overfit by putting too much weight on the design-based estimates. We suggest using the STS univariate analysis to scale the design-based variance estimates in the right direction for further use in multilevel models. The uncertainty around design-based variance estimates could also be taken into account by imposing a prior distribution on them, as in You and Chapman (2006). The comparative analysis of the two time series modelling approaches, however, should not be affected much by the way the design-based variance estimates are treated in this paper. The reported reduction in design-based standard errors is conditional on these approximated design-based variance estimates.
The techniques presented here can be used nearly in any repeatedly conducted SAE application, especially if a survey suffers from discontinuities due to redesigns. Unlike in the application considered in Bollineni-Balabay et al. (2016), the presence of design-based variance estimates here makes it possible to continuously apply this model-based approach in the production of official statistical figures. Accounting for each new survey redesign will, however, be possible only with a delay of at least one period. In addition, in the first time periods after a survey redesign, estimated figures are likely to undergo substantial revisions, as soon as new data become available under the new design. Yet, this problem is of a temporary nature and there does not seem to exist any other solution except for a parallel run that increases the survey expenses. See Van den Brakel and Krieg (2015) for an example where a parallel run is conducted to obtain design-based estimates for discontinuities. These estimates are then used as a priori information in a structural time series model to avoid the problem of revisions.
In this paper, the time dimension has been exploited for variance and volatility reduction in the point-estimates, as well as for level break estimation. As for the spatial dimension, so far it has been used only for getting rid of some idiosyncracies through pooling trend variances across provinces. However, it can also be used as another source of variance reduction in model-based estimates by exploiting spatial correlations between domains belonging to different provinces, e.g., by allowing for f common stochastic trends shared by more than f domain trends, the way it was done in e.g. Bollineni-Balabay et al. (2016), Krieg and Van den Brakel (2012). In the case of the DTS, this approach seems to be worth exploring, since the pattern of trends belonging to a certain motive-modality intersection is very similar among the twelve provinces of the Netherlands.

Acknowledgements

The authors thank Bart Buelens and Barry Schouten (Statistics Netherlands) for thorough reading and valuable comments on a draft of this paper.

References

Bollineni-Balabay, O., Van den Brakel, J., and Palm, F. (2015). Accounting for Hyperparameter Uncertainty in Small Area Estimation Based on State-Space Models. *Discussion paper, Statistics Netherlands, Heerlen*. `http://www.cbs.nl/NR/rdonlyres/B2E54E0F-C791-4665-9830-95EDF4D5A198/0/2015accountinghyperparameteruncertaintyinsmallsarea.pdf`.

Bollineni-Balabay, O., Van den Brakel, J., and Palm, F. (2016). Multivariate state space approach to variance reduction in series with level and variance breaks due to survey redesigns. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*.

Boonstra, H. J. (2015). *mcmcsae: MCMC small area estimation*. R package version 0.5.

Clark, T. S. and Linzer, D. A. (2015). Should I use fixed or random effects? *Political Science Research and Methods* 3, 399--408.

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

Doornik, J. (2007). *An object-oriented matrix programming language Ox 5*. Timberlake Consultants Press, London.

Durbin, J. and Koopman, S. J. (2012). *Time series analysis by state space methods*. Number 38. Oxford University Press.

Fay, R. and Herriot, R. (1979). Estimates of income for small places: an application of James-Stein procedures to census data. *Journal of the American Statistical Association* 74 (366a), 269--277.

Gelfand, A. E. and Smith, A. F. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* 85 (410), 398--409.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis* 1 (3), 515--534.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2014). *Bayesian data analysis*, Volume 2. Taylor & Francis.

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

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

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (6), 721--741.

Ghosh, M. and Lahiri, P. (1987). Robust empirical Bayes estimation of means from stratified samples. *Journal of the American Statistical Association* 82 (400), 1153--1162.

Gonzalez, M. (1973). Use and evaluation of synthetic estimates. *Proceedings of the Social Statistics Section, American Statistical Association*, 33--36.

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

Harvey, A. C. (2001). Trend analysis. *Encyclopedia of Environmetrics*.

Harvey, A. C. and Koopman, S. J. (1997). Multivariate Structural Time Series Models. In C. Heij, H. Schumacher, B. Hanzon, and C. Praagman (Eds.), *Systematic dynamics in economic and financial models*, pp. 269--298. Chichester: Wiley and Sons.

Koopman, S. J. (1997). Exact initial Kalman filtering and smoothing for nonstationary time series models. *Journal of the American Statistical Association* 92 (440), 1630--1638.

Koopman, S. J., Shephard, N., and Doornik, J. (2008). *SsfPack 3.0: statistical algorithms for models in state space form*. Timberlake Consultants Press, London.

Krieg, S. and Van den Brakel, J. (2012). Estimation of the monthly unemployment rate for six domains through structural time series modelling with cointegrated trends. *Computational Statistics & Data Analysis* 56 (10), 2918--2933.

Pfeffermann, D. (1991). Estimation and seasonal adjustment of population means using data from repeated surveys. *Journal of Business and Economic Statistics* 9, 163--175.

Pfeffermann, D. and Bleuer, S. (1993). Robust joint modelling of labour force series of small areas. *Survey Methodology* 19, 149--163.

Pfeffermann, D., Sikov, A., and Tiller, R. (2014). Single- and two-stage cross-sectional and time series benchmarking procedures for small area estimation. *TEST* 23 (4), 631--666.

Pfeffermann, D. and Tiller, R. (2005). Bootstrap Approximation to Prediction MSE for State-Space Models with Estimated Parameters. *Journal of Time Series Analysis* 26, 893--916.

Pfeffermann, D. and Tiller, R. (2006). Small-area estimation with state-space models subject to benchmark constraints. *Journal of the American Statistical Association* 101 (476), 1387--1397.

Polson, N. G., Scott, J. G., et al. (2012). On the half-Cauchy prior for a global scale parameter. *Bayesian Analysis* 7 (4), 887--902.

Prasad, N. and Rao, J. (1990). The estimation of the mean squared error of small-area estimators. *Journal of the American statistical Association* 85 (409), 163--171.

Rao, J. (2003). *Small area estimation*. John Wiley & Sons.

Rao, J. and Yu, M. (1994). Small-area estimation by combining time-series and cross-sectional data. *Canadian Journal of Statistics* 22 (4), 511--528.

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

Särndal, C.-E., Swensson, B., and Wretman, J. (1992). *Model assisted survey sampling*. Springer.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64 (4), 583--639.

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

Van den Brakel, J. and Roels, J. (2010). Intervention analysis with state-space models to estimate discontinuities due to a survey redesign. *Annals of Applied Statistics* 4, 1105--1138.

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

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

Appendices

A Full Conditional Distributions for the Gibbs-Sampler

The Gibbs sampler was first described by Geman and Geman (1984). Here we present unnormalised conditional posterior densities for each parameter b in $\boldsymbol{\psi} = (\boldsymbol{\nu}, \boldsymbol{\kappa}, \boldsymbol{\beta}^{\boldsymbol{RE}'}, \boldsymbol{u}, \boldsymbol{\beta}', \sigma_\nu^2, \sigma_\kappa^2, \sigma_{\beta^{RE}}^2, \sigma_u^2, \xi_\nu, \xi_\kappa, \xi_{\beta^{RE}}, \xi_u)$ for Model (3.2). Let $\boldsymbol{\psi}^{(-\boldsymbol{b})}$ denote the parameter vector where element b is deleted. Within the Gibbs sampler, the b-th parameter values are drawn conditionally on the data $\boldsymbol{Y}$ and the rest of the parameters $\boldsymbol{\psi}^{(-\boldsymbol{b})}$.
The conditional posterior density for parameters $c, \kappa, \beta_1^{RE}, \ldots, \beta_{K^{RE}}^{RE}$ and $\boldsymbol{\beta}^{\boldsymbol{FE}}$ contained in vector $\boldsymbol{\beta}$ originate from the product of the densities that contain these parameters:

$$\begin{aligned} p(\boldsymbol{\beta}|\boldsymbol{\psi}^{(-\boldsymbol{\beta})}, \boldsymbol{Y}) \propto & N(\boldsymbol{\beta_0}, \boldsymbol{\Omega_{\beta_0}})\times \\ & N(\xi_\nu \tilde{\boldsymbol{\nu}} \otimes \boldsymbol{\iota}_{[T]} + \xi_\kappa \tilde{\boldsymbol{\kappa}} \otimes \boldsymbol{t} + \xi_{\beta^{RE}} \boldsymbol{X^{RE}} \tilde{\boldsymbol{\beta}}^{\boldsymbol{RE}} + \boldsymbol{X\beta} + \xi_u \tilde{\boldsymbol{u}}, \boldsymbol{\Phi}), \end{aligned} \tag{A.1}$$

which, using the results on conjugate priors in Gelman et al. (2014), turns into a normal density $N(\boldsymbol{\mu_\beta}, \boldsymbol{\Omega_\beta})$ with the following mean and variance:

$$\begin{aligned} \boldsymbol{\mu_\beta} =& (\boldsymbol{X}'\boldsymbol{\Phi}^{-1}\boldsymbol{X} + \boldsymbol{\Omega}_{\boldsymbol{\beta_0}}^{-1})^{-1}\times \\ & \left[\boldsymbol{X}'\boldsymbol{\Phi}^{-1}(\boldsymbol{Y} - \xi_\nu \tilde{\boldsymbol{\nu}} \otimes \boldsymbol{\iota}_{[T]} - \xi_\kappa \tilde{\boldsymbol{\kappa}} \otimes \boldsymbol{t} - \xi_{\beta^{RE}} \boldsymbol{X^{RE}} \tilde{\boldsymbol{\beta}}^{\boldsymbol{RE}} - \xi_u \tilde{\boldsymbol{u}}) + \boldsymbol{\Omega}_{\boldsymbol{\beta_0}}^{-1}\boldsymbol{\beta_0}\right], \\ \boldsymbol{\Omega_\beta} =& (\boldsymbol{X}'\boldsymbol{\Phi}^{-1}\boldsymbol{X} + \boldsymbol{\Omega}_{\boldsymbol{\beta_0}}^{-1})^{-1}. \end{aligned}$$

For vector and matrix dimensions, refer to the description under the likelihood function presentation in (3.10).
The conditional posterior of the M-dimensional vector of scaled area effects $\tilde{\nu}_m$ is:

$$\begin{aligned} p(\tilde{\boldsymbol{\nu}}|\boldsymbol{\psi}^{(-\tilde{\boldsymbol{\nu}})}, \boldsymbol{Y}) =& N(\boldsymbol{\mu_{\tilde{\nu}}}, \boldsymbol{\Omega_{\tilde{\nu}}}) \propto N(\boldsymbol{0}_{[M]}, \tilde{\sigma}_\nu^2 \boldsymbol{I}_{[M]})\times \\ & N(\xi_\nu \tilde{\boldsymbol{\nu}} \otimes \boldsymbol{\iota}_{[T]} + \xi_\kappa \tilde{\boldsymbol{\kappa}} \otimes \boldsymbol{t} + \xi_{\beta^{RE}} \boldsymbol{X^{RE}} \tilde{\boldsymbol{\beta}}^{\boldsymbol{RE}} + \boldsymbol{X\beta} + \xi_u \tilde{\boldsymbol{u}}, \boldsymbol{\Phi}), \\ \boldsymbol{\mu_{\tilde{\nu}}} =& (\xi_\nu^2 \boldsymbol{I}_{[M]} \otimes \boldsymbol{\iota}'_{[T]} \boldsymbol{\Phi}^{-1} \boldsymbol{I}_{[M]} \otimes \boldsymbol{\iota}_{[T]} + 1/\tilde{\sigma}_\nu^2)^{-1}\times \\ & \xi_\nu^2 \boldsymbol{I}_{[M]} \otimes \boldsymbol{\iota}'_{[T]} \boldsymbol{\Phi}^{-1}(\boldsymbol{Y} - \xi_\kappa \tilde{\boldsymbol{\kappa}} \otimes \boldsymbol{t} - \xi_{\beta^{RE}} \boldsymbol{X^{RE}} \tilde{\boldsymbol{\beta}}^{\boldsymbol{RE}} - \boldsymbol{X\beta} - \xi_u \tilde{\boldsymbol{u}}), \\ \boldsymbol{\Omega_{\tilde{\nu}}} =& (\xi_\nu^2 \boldsymbol{I}_{[M]} \otimes \boldsymbol{\iota}'_{[T]} \boldsymbol{\Phi}^{-1} \boldsymbol{I}_{[M]} \otimes \boldsymbol{\iota}_{[T]} + 1/\tilde{\sigma}_\nu^2)^{-1}. \end{aligned} \tag{A.2}$$

A similar expression is valid for the linear trend random effects $\tilde{\kappa}_m$:

$$\begin{aligned}
p(\tilde{\boldsymbol{\kappa}}|\boldsymbol{\psi}^{(-\tilde{\kappa})},\boldsymbol{Y}) =& N(\boldsymbol{\mu}_{\tilde{\kappa}},\boldsymbol{\Omega}_{\tilde{\kappa}}) \propto N(\boldsymbol{0}_{[M]},\tilde{\sigma}_\kappa^2\boldsymbol{I}_{[M]})\times \\
& N(\xi_\nu\tilde{\boldsymbol{\nu}}\otimes\boldsymbol{\iota}_{[T]}+\xi_\kappa\tilde{\boldsymbol{\kappa}}\otimes\boldsymbol{t}+\xi_{\beta^{RE}}\boldsymbol{X^{RE}}\tilde{\boldsymbol{\beta}}^{\boldsymbol{RE}}+\boldsymbol{X\beta}+\xi_u\tilde{\boldsymbol{u}},\boldsymbol{\Phi}), \\
\boldsymbol{\mu}_{\tilde{\kappa}} =& (\xi_\kappa^2\boldsymbol{I}_{[M]}\otimes\boldsymbol{t'\Phi^{-1}}\boldsymbol{I}_{[M]}\otimes\boldsymbol{t}+1/\tilde{\sigma}_\kappa^2)^{-1}\times \\
& \xi_\kappa^2\boldsymbol{I}_{[M]}\otimes\boldsymbol{t'\Phi^{-1}}(\boldsymbol{Y}-\xi_\nu\tilde{\boldsymbol{\nu}}\otimes\boldsymbol{\iota}_{[T]}-\xi_{\beta^{RE}}\boldsymbol{X^{RE}}\tilde{\boldsymbol{\beta}}^{\boldsymbol{RE}}-\boldsymbol{X\beta}-\xi_u\tilde{\boldsymbol{u}}), \\
\boldsymbol{\Omega}_{\tilde{\kappa}} =& (\xi_\kappa^2\boldsymbol{I}_{[M]}\otimes\boldsymbol{t'\Phi^{-1}}\boldsymbol{I}_{[M]}\otimes\boldsymbol{t}+1/\tilde{\sigma}_\kappa^2)^{-1},
\end{aligned} \tag{A.3}$$

where $\boldsymbol{t}$ denotes a vertical vector with time indicators $(0,1,\dots,T-1)'$. The same logic applies to random level break coefficients $\tilde{\boldsymbol{\beta}}_{\boldsymbol{k}}^{\boldsymbol{RE}}$ (inter alia, vectors $\boldsymbol{t}$ are replaced with T-dimensional vectors with dummy regressors for the level break k).

The conditional distribution for the T-dimensional scaled stochastic trend random terms $\tilde{\boldsymbol{u}}_{\boldsymbol{m}}$ is based on the data and other terms that are related to domain m:

$$\begin{aligned}
p(\tilde{\boldsymbol{u}}_{\boldsymbol{m}}|\boldsymbol{\psi}^{(-\tilde{u}_m)},\boldsymbol{Y}) =& N(\boldsymbol{\mu}_{\tilde{\boldsymbol{u}}_{\boldsymbol{m}}},\boldsymbol{\Omega}_{\tilde{\boldsymbol{u}}_{\boldsymbol{m}}}) \propto N(\boldsymbol{0}_{[T]},\tilde{\sigma}_{u,m}^2\boldsymbol{A})\times \\
& N(\xi_\nu\tilde{\nu}_m\boldsymbol{\iota}_{[T]}+\xi_\kappa\tilde{\kappa}_m\boldsymbol{t}+\xi_{\beta^{RE}}\boldsymbol{X}_{\boldsymbol{m}}^{\boldsymbol{RE}}\tilde{\boldsymbol{\beta}}_{\boldsymbol{m}}^{\boldsymbol{RE}}+ \\
& \boldsymbol{X_m\beta_m}+\xi_u\tilde{\boldsymbol{u}}_{\boldsymbol{m}},\boldsymbol{\Phi}_{\boldsymbol{m}}), \\
\boldsymbol{\mu}_{\tilde{\boldsymbol{u}}_{\boldsymbol{m}}} =& (\xi_u^2\boldsymbol{\Phi}_{\boldsymbol{m}}^{-1}+\boldsymbol{A}^{-1}/\tilde{\sigma}_{u,m}^2)^{-1}\times \\
& \xi_u^2\boldsymbol{\Phi}_{\boldsymbol{m}}^{-1}(\boldsymbol{Y_m}-\xi_\nu\tilde{\nu}_m\boldsymbol{\iota}_{[T]}-\xi_\kappa\tilde{\kappa}_m\boldsymbol{t}-\xi_{\beta^{RE}}\boldsymbol{X}_{\boldsymbol{m}}^{\boldsymbol{RE}}\tilde{\boldsymbol{\beta}}_{\boldsymbol{m}}^{\boldsymbol{RE}}-\boldsymbol{X_m\beta_m}), \\
\boldsymbol{\Omega}_{\tilde{\boldsymbol{u}}_{\boldsymbol{m}}} =& (\xi_u^2\boldsymbol{\Phi}_{\boldsymbol{m}}^{-1}+\boldsymbol{A}^{-1}/\tilde{\sigma}_{u,m}^2)^{-1},
\end{aligned} \tag{A.4}$$

where vector $\boldsymbol{Y_m}$ and matrix $\boldsymbol{\Phi_m}$ are T-dimensional, matrix $\boldsymbol{X}_{\boldsymbol{m}}^{\boldsymbol{RE}}$ is $[T\times K_{RE}]$ and contains indicator variables for the scaled random level break coefficients $\tilde{\boldsymbol{\beta}}_{\boldsymbol{m}}^{\boldsymbol{RE}}$ in domain m, $\tilde{\boldsymbol{\beta}}_{\boldsymbol{m}}^{\boldsymbol{RE}}$ being $[K_{RE}\times 1]$. The matrix with fixed effects regressors $\boldsymbol{X_m}$ is build by the same logic, i.e. contains regressors only applicable to domain m and is therefore $[T\times(2+K_{RE}+d_m^{FE})]$, d_m^{FE} being the number of level breaks modelled as fixed effects in domain m. $\boldsymbol{\beta_m}$ contains the overall effects $c,\kappa,\beta_1^{RE},\dots\beta_{K_{RE}}^{RE}$ along with d_m^{FE} fixed level break coefficients. Drawing from (A.4) results in unconstrained draws of the stochastic trend terms. Due to restrictions presented in (3.4), these unconstrained draws should be adjusted to $\left(\boldsymbol{I}_{[T]}-\frac{\boldsymbol{\Omega}_{\tilde{\boldsymbol{u}}_{\boldsymbol{m}}}\boldsymbol{\iota}_{[T]}\boldsymbol{\iota}_{[T]}'}{\boldsymbol{\iota}_{[T]}'\boldsymbol{\Omega}_{\tilde{\boldsymbol{u}}_{\boldsymbol{m}}}\boldsymbol{\iota}_{[T]}}\right)\tilde{\boldsymbol{u}}_{\boldsymbol{m}}$, according to Rue and Held (2005).

Variance components of random effects are drawn from the $Inv-\chi^2$ density:

$$\begin{aligned}
p(\tilde{\sigma}_\nu^2|\boldsymbol{\psi}^{(-\tilde{\sigma}_\nu^2)},\boldsymbol{Y}) &\propto N(\boldsymbol{0}_{[M]},\tilde{\sigma}_\nu^2\boldsymbol{I}_{[M]})\times Inv-\chi^2(v_\nu,s_\nu^2), \\
p(\tilde{\sigma}_\nu^2|\boldsymbol{\psi}^{(-\tilde{\sigma}_\nu^2)},\boldsymbol{Y}) &= Inv-\chi^2(v_\nu+M,\frac{v_\nu s_\nu^2+\sum_m\tilde{\nu}_m^2}{v_\nu+M}).
\end{aligned} \tag{A.5}$$

The same goes for the other random effects, except for the stochastic trend terms (e.g., for random linear trend effects $\tilde{\kappa}_m$, terms v_ν, s_ν^2 and $\sum_m\tilde{\nu}_m^2$ would be replaced by v_κ, s_κ^2 and $\sum_m\tilde{\kappa}_m^2$, respectively).

If every domain m is assigned a unique value for its stochastic trend variance (as is the case at the national level of the DTS), then $\tilde{\sigma}_{u,m}^2$ is drawn from the following conditional:

$$\begin{aligned}
p(\tilde{\sigma}_{u,m}^2|\boldsymbol{\psi}^{(-\tilde{\sigma}_{u,m}^2)},\boldsymbol{Y}) &\propto Inv-\chi^2(v_u,s_u^2)N(\boldsymbol{0}_{[T]},\tilde{\sigma}_{u,m}^2\boldsymbol{A}), \\
p(\tilde{\sigma}_{u,m}^2|\boldsymbol{\psi}^{(-\tilde{\sigma}_{u,m}^2)},\boldsymbol{Y}) &= Inv-\chi^2(v_u+T-2,\frac{v_u s_u^2+\tilde{\boldsymbol{u}}_{\boldsymbol{m}}'\boldsymbol{A}^{-1}\tilde{\boldsymbol{u}}_{\boldsymbol{m}}}{v_u+T-2}),
\end{aligned} \tag{A.6}$$

where 2 is subtracted from T due to the two restrictions for the integrated random walk model.

The scaling ξ_ν-effects are drawn from the following distribution:

$$\begin{aligned}
p(\xi_\nu|\boldsymbol{\psi}^{(-\xi_\nu)},\boldsymbol{Y}) =& N(\mu_{\xi_\nu},\omega_{\xi_\nu}) \propto N(\alpha_\nu,\gamma_\nu)\times \\
& N(\xi_\nu\tilde{\boldsymbol{\nu}}\otimes\boldsymbol{\iota}_{[T]}+\xi_\kappa\tilde{\boldsymbol{\kappa}}\otimes\boldsymbol{t}+\xi_{\beta^{RE}}\boldsymbol{X^{RE}}\tilde{\boldsymbol{\beta}}^{\boldsymbol{RE}}+\boldsymbol{X\beta}+\xi_u\tilde{\boldsymbol{u}},\boldsymbol{\Phi}), \\
\mu_{\xi_\nu} =& (\tilde{\boldsymbol{\nu}}'\boldsymbol{I}_{[M]}\otimes\boldsymbol{\iota}_{[T]}'\boldsymbol{\Phi}^{-1}\boldsymbol{I}_{[M]}\otimes\boldsymbol{\iota}_{[T]}\tilde{\boldsymbol{\nu}}+1/\gamma_\nu)^{-1}\times \\
& \tilde{\boldsymbol{\nu}}'\boldsymbol{I}_{[M]}\otimes\boldsymbol{\iota}_{[T]}'\boldsymbol{\Phi}^{-1}(\boldsymbol{Y}-\xi_\kappa\tilde{\boldsymbol{\kappa}}\otimes\boldsymbol{t}-\xi_{\beta^{RE}}\boldsymbol{X^{RE}}\tilde{\boldsymbol{\beta}}^{\boldsymbol{RE}}-\boldsymbol{X\beta}-\xi_u\tilde{\boldsymbol{u}}), \\
\omega_{\xi_\nu} =& (\tilde{\boldsymbol{\nu}}'\boldsymbol{I}_{[M]}\otimes\boldsymbol{\iota}_{[T]}'\boldsymbol{\Phi}^{-1}\boldsymbol{I}_{[M]}\otimes\boldsymbol{\iota}_{[T]}\tilde{\boldsymbol{\nu}}+1/\gamma_\nu)^{-1}.
\end{aligned} \tag{A.7}$$

The same logic applies to the scaling parameters of the rest of the random terms.
As for scaling the stochastic trend terms, the number of unique ξ_u-parameters can be made equal to the number of unique variances on the main diagonal of $\boldsymbol{\Sigma_u}$. At the provincial level of the DTS, for instance, the number of stochastic trend variances, and thus ξ_u-parameters, is $Mot \times Mod$, producing a $Mot \times Mod$-dimensional $\boldsymbol{\xi_u}$ vector. In this case, $\xi_u \tilde{\boldsymbol{u}}$ terms in every expression of this appendix should be substituted with the following:

$$\xi_u \tilde{\boldsymbol{u}} \rightarrow \tilde{\boldsymbol{u}} \boldsymbol{I}_{[\boldsymbol{M}]} \otimes \boldsymbol{\iota}_{[\boldsymbol{T}]} \boldsymbol{\iota}_{[\boldsymbol{M}/\boldsymbol{Mot}/\boldsymbol{Mod}]} \otimes \boldsymbol{\xi_u}. \quad (A.8)$$

B Provincial level auxiliary estimation results

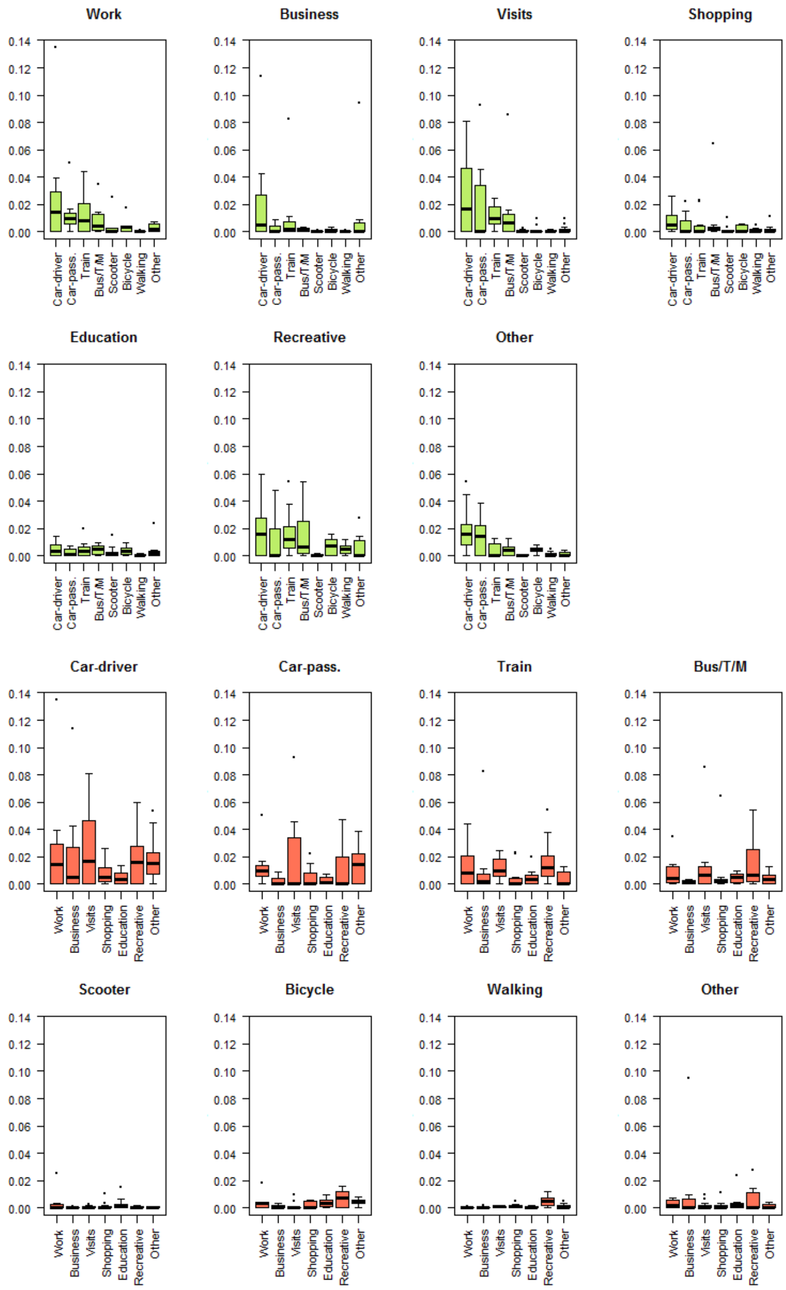
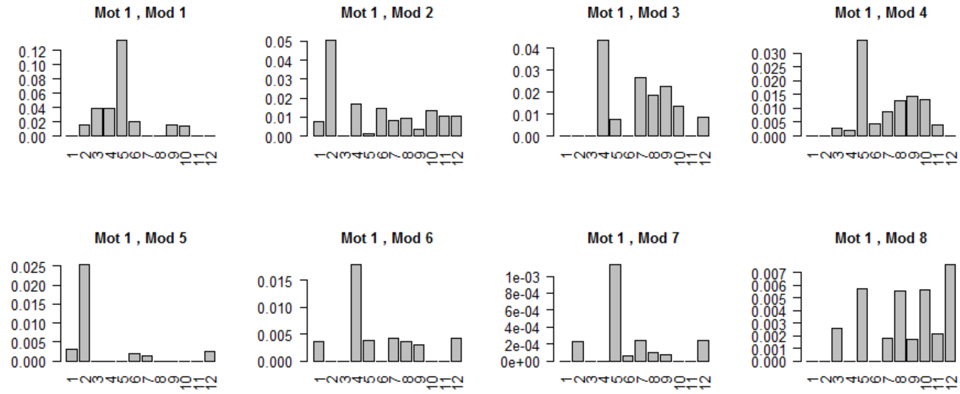
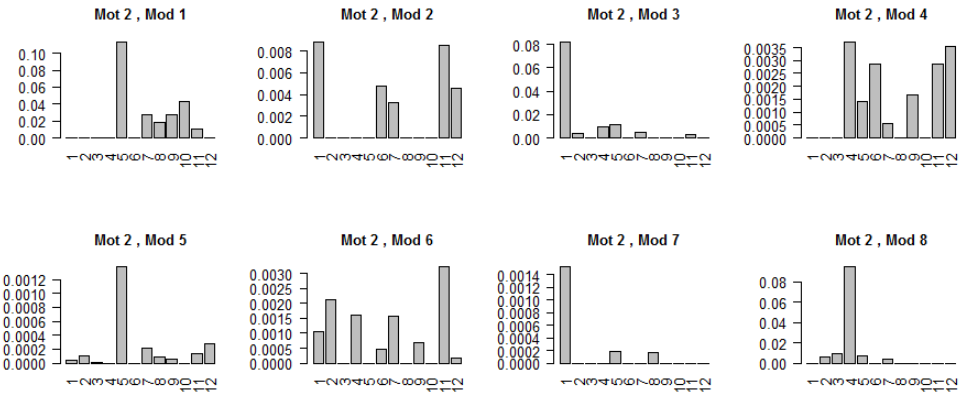
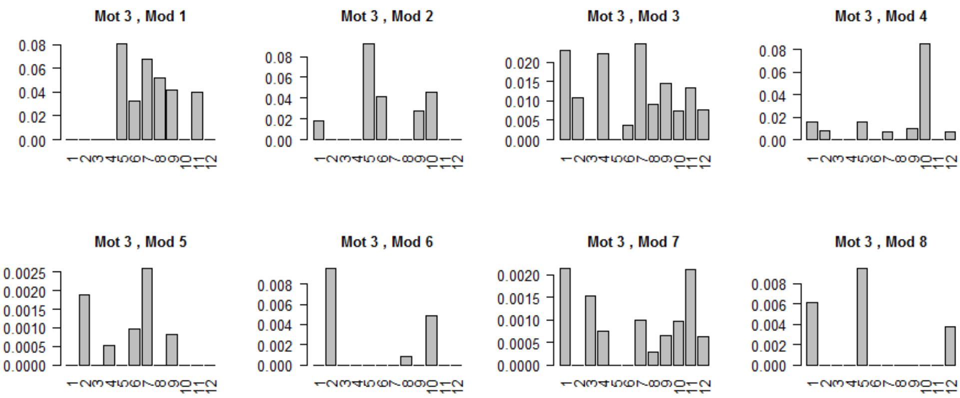
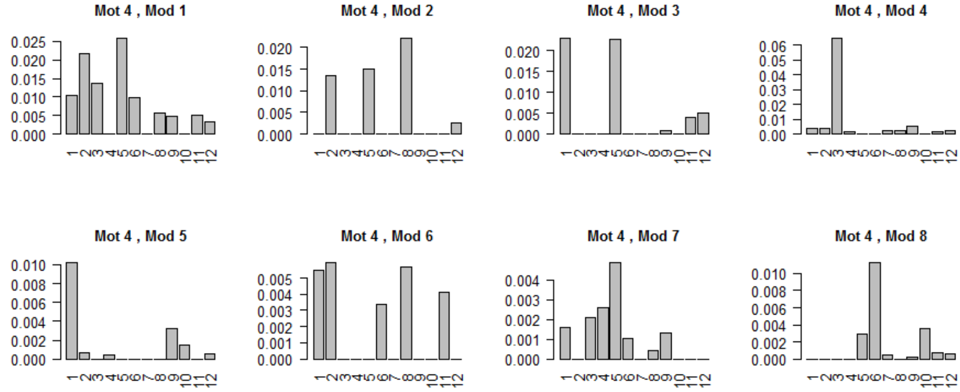


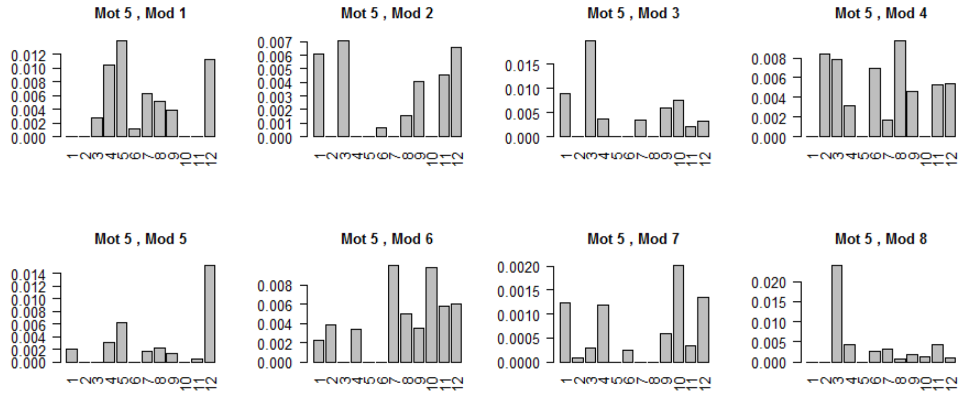
Figure B.1 Box-plots for σ_us from univariate STS models at the provincial level: modalities within motives (upper panel) and motives within modalities (lower panel)

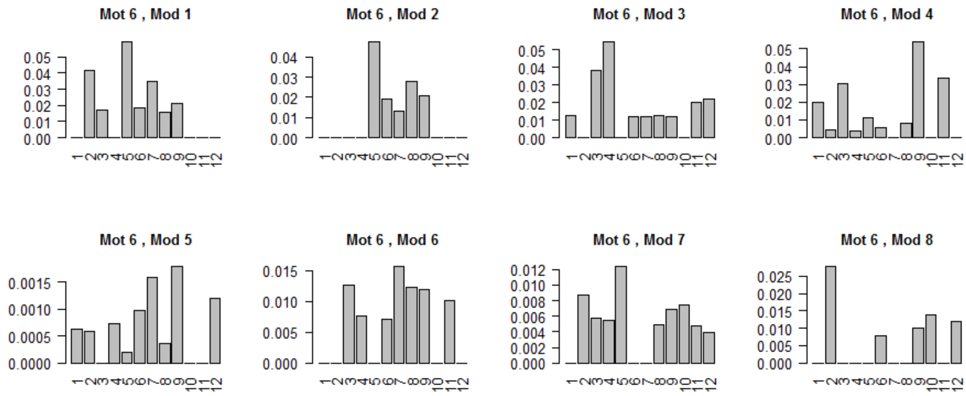

Mot 1 , Mod 1
Mot 1 , Mod 2
Mot 1 , Mod 3
Mot 1 , Mod 4
Mot 1 , Mod 5
Mot 1 , Mod 6
Mot 1 , Mod 7
Mot 1 , Mod 8


Mot 2 , Mod 1
Mot 2 , Mod 2
Mot 2 , Mod 3
Mot 2 , Mod 4
Mot 2 , Mod 5
Mot 2 , Mod 6
Mot 2 , Mod 7
Mot 2 , Mod 8


Mot 3 , Mod 1
Mot 3 , Mod 2
Mot 3 , Mod 3
Mot 3 , Mod 4
Mot 3 , Mod 5
Mot 3 , Mod 6
Mot 3 , Mod 7
Mot 3 , Mod 8


Mot 4 , Mod 1
Mot 4 , Mod 2
Mot 4 , Mod 3
Mot 4 , Mod 4
Mot 4 , Mod 5
Mot 4 , Mod 6
Mot 4 , Mod 7
Mot 4 , Mod 8


Mot 5 , Mod 1
Mot 5 , Mod 2
Mot 5 , Mod 3
Mot 5 , Mod 4
Mot 5 , Mod 5
Mot 5 , Mod 6
Mot 5 , Mod 7
Mot 5 , Mod 8


Mot 6 , Mod 1
Mot 6 , Mod 2
Mot 6 , Mod 3
Mot 6 , Mod 4
Mot 6 , Mod 5
Mot 6 , Mod 6
Mot 6 , Mod 7
Mot 6 , Mod 8

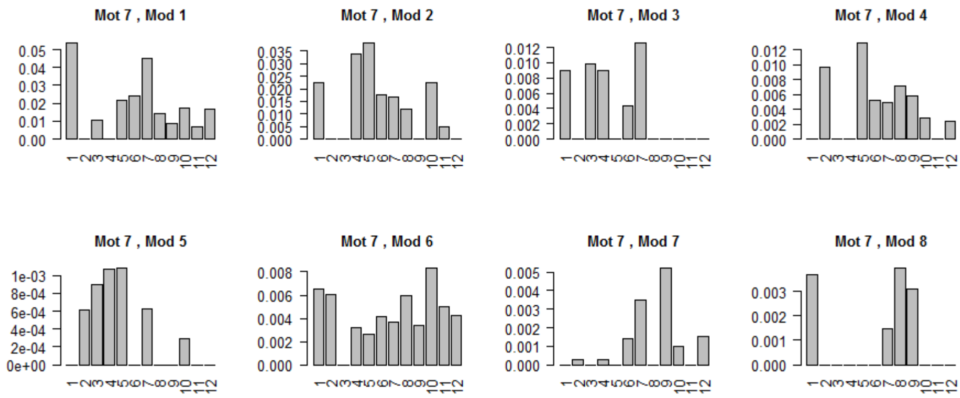
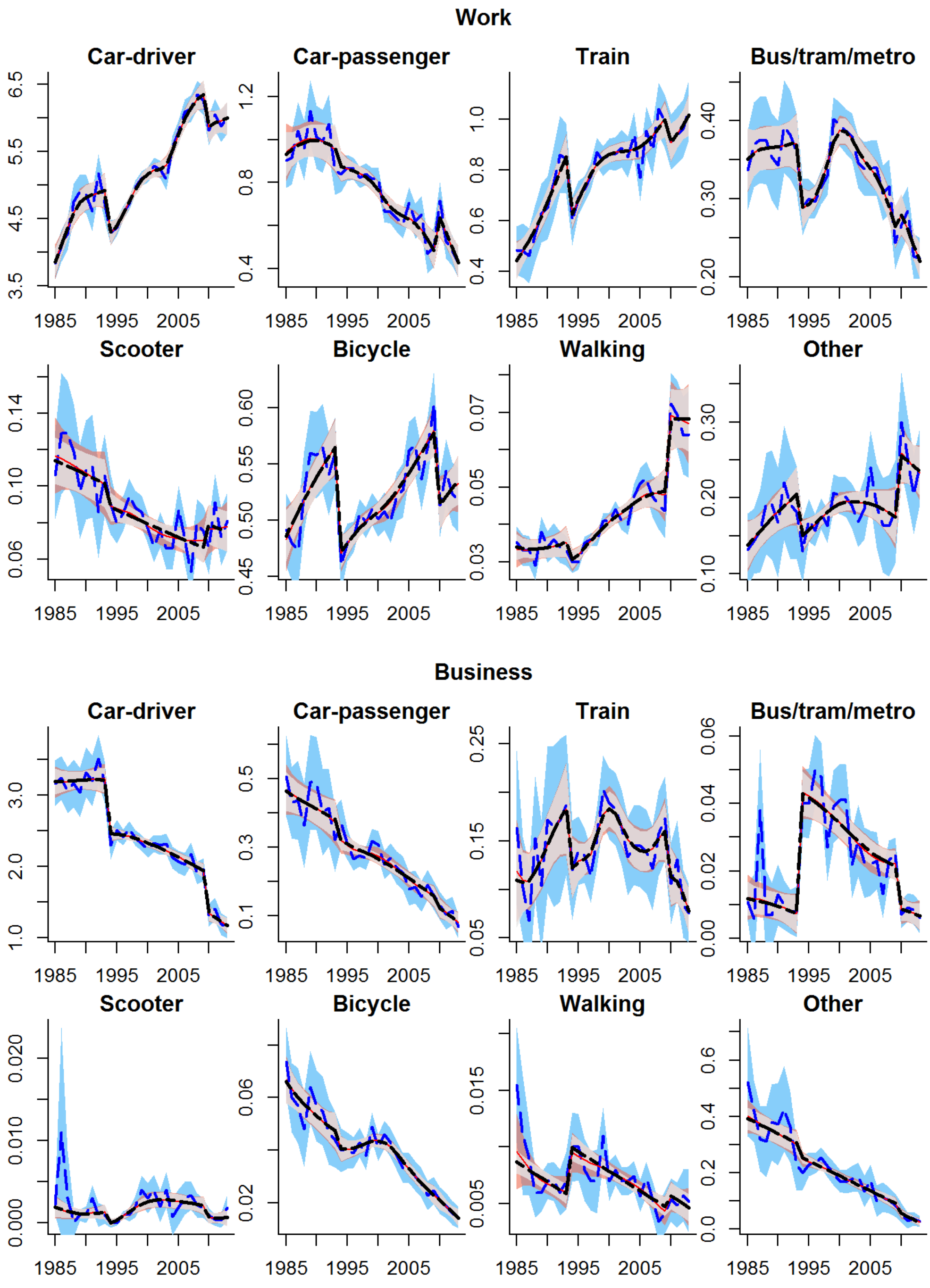
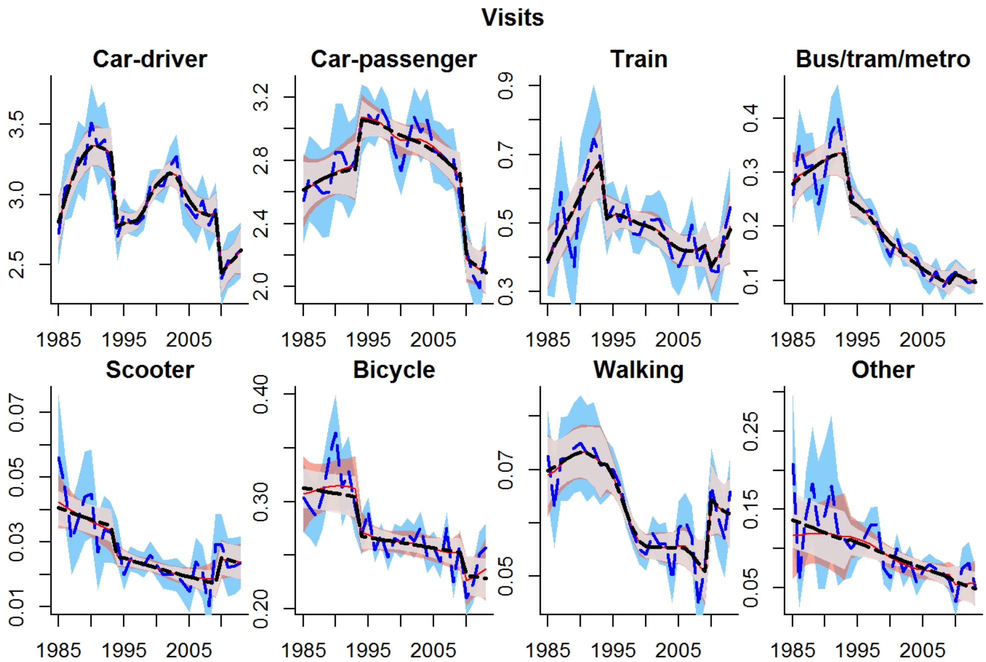
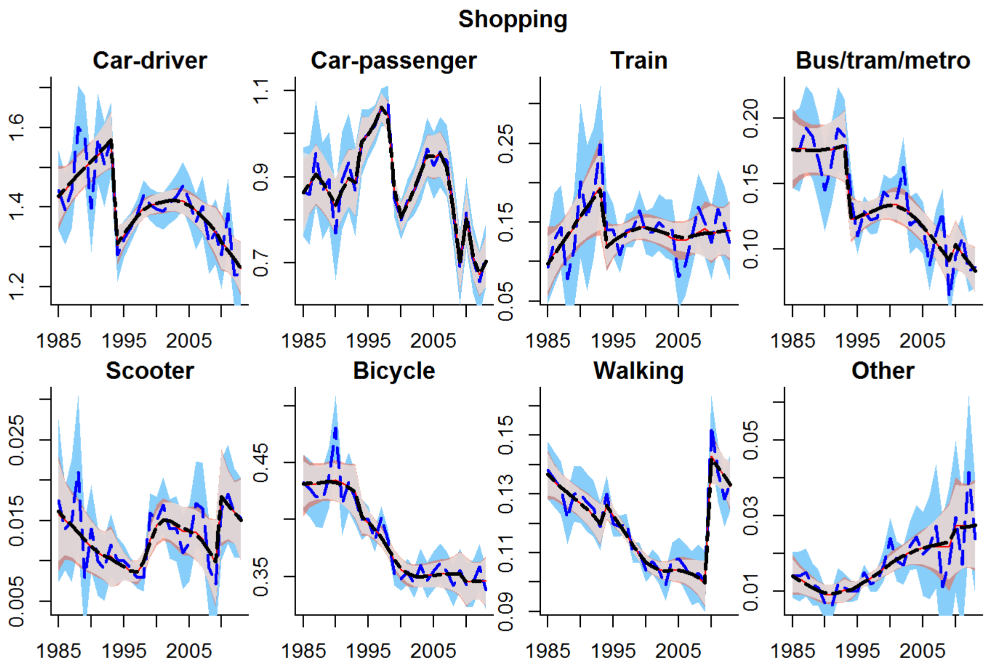


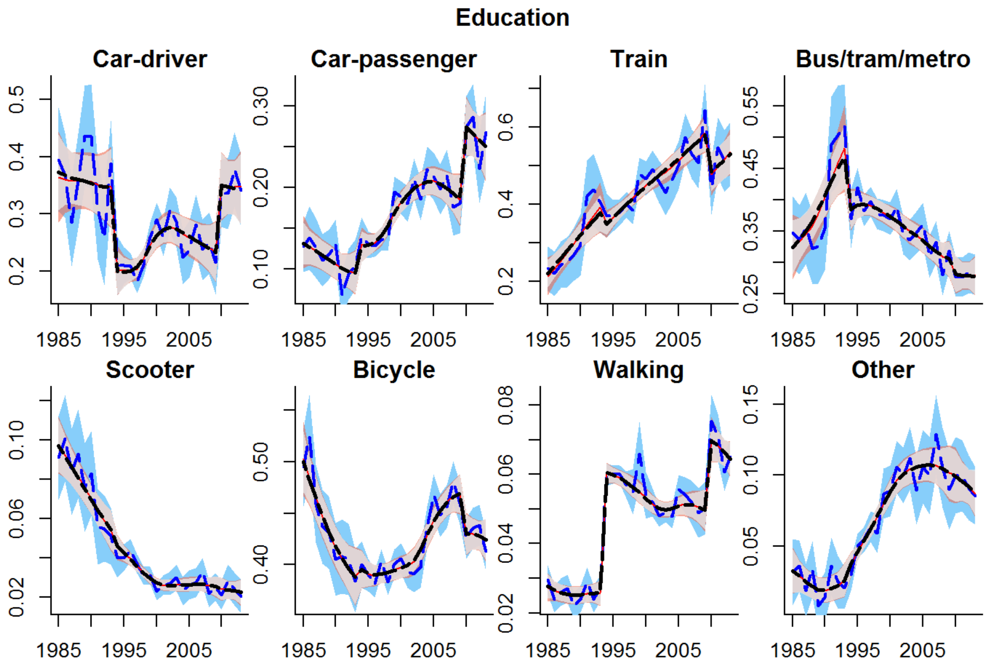
Figure B.2 Provincial σ_u from univariate STS models for every intersection of motives and transport modalities.

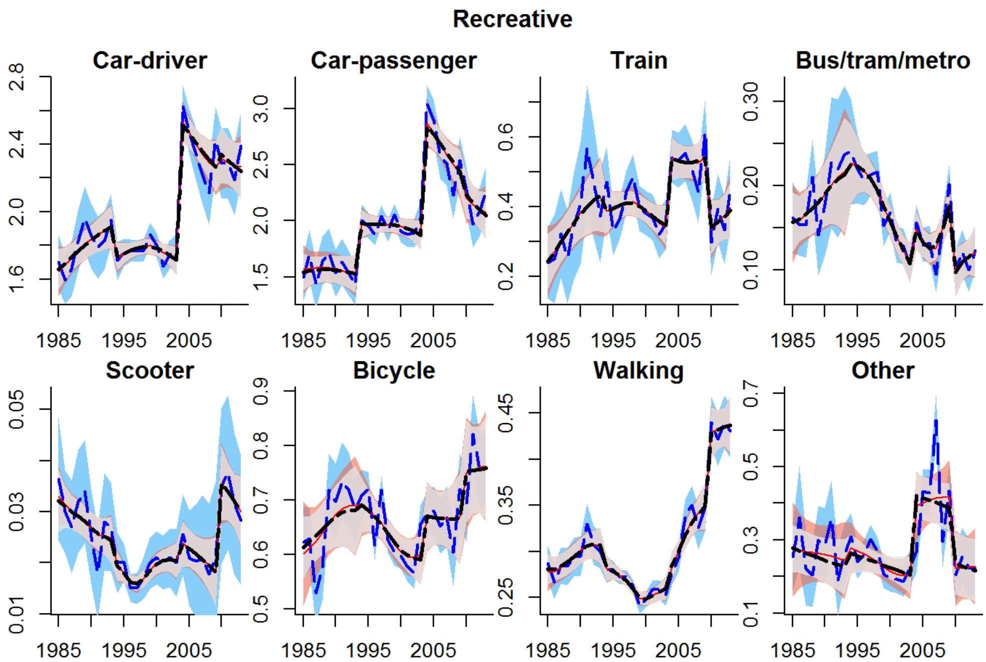
C National level estimation results




Visits
Car-driver
Car-passenger
Train
Bus/tram/metro
Scooter
Bicycle
Walking
Other
1985
1995
2005


Shopping
Car-driver
Car-passenger
Train
Bus/tram/metro
Scooter
Bicycle
Walking
Other
1985
1995
2005


Education
Car-driver
Car-passenger
Train
Bus/tram/metro
Scooter
Bicycle
Walking
Other
1985 1995 2005


Recreative
Car-driver
Car-passenger
Train
Bus/tram/metro
Scooter
Bicycle
Walking
Other
1985 1995 2005

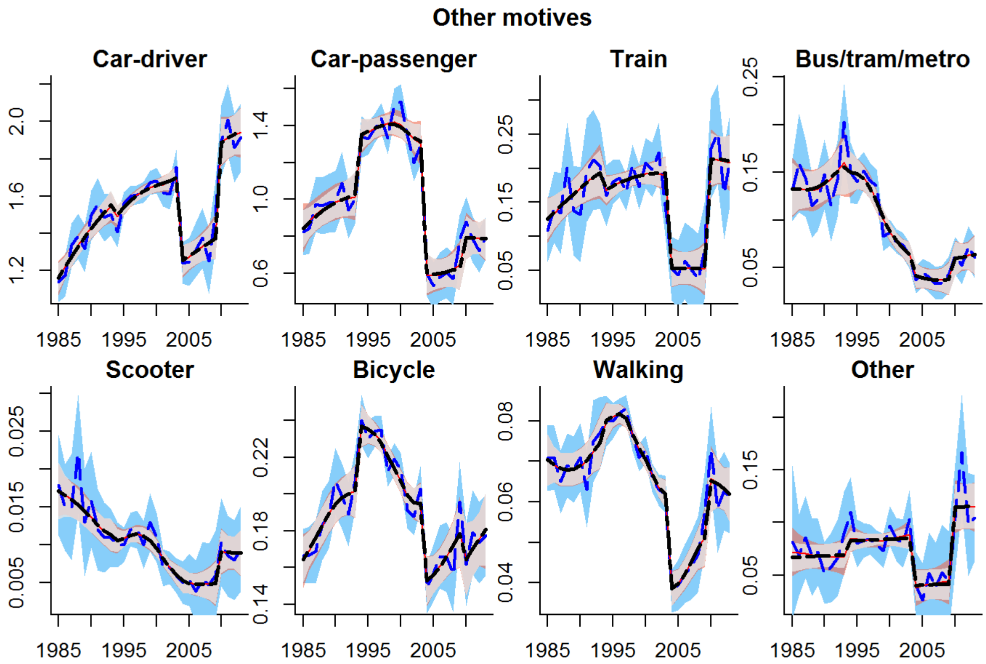


Figure C.1 The national level design-based point-estimates (thin blue dashed line), point-estimates from the multivariate STS model (thick black dashed) and posterior means from the HB-FE model (solid red); the colour of the 95% confidence (credible) intervals corresponds to the colour of the point-estimates