Multivariate State-Space Approach to Variance Reduction in Series with Level and Variance Breaks

Oksana Bollineni-Balabay

Jan van den Brakel

Franz Palm

1 September 2014

Contents

Foreword

The present work illustrates how time-series model-based techniques can improve the production of official statistical figures for repeatedly conducted surveys. Apart from reducing the variance of design-based estimates, time series models provide estimates on the magnitude of discontinuities caused by survey redesigns. Finally, in surveys with a subdivision into several domains, such models can be effectively applied in the context of small area estimation problems. The authors thank dr. Sabine Krieg (Statistics Netherlands) for valuable comments and Rob Willems (Statistics Netherlands) for the technical help with calculating design variances from the micro-data.

Summary

Statistics Netherlands applies a design-based estimation procedure to produce Road Transportation figures. Frequent survey redesigns caused discontinuities in these series which obstructs the comparability of figures over time. Moreover, several reductions of the sample size and changes in the sample design resulted in variance breaks and unacceptably large sampling errors in the last part of the series. This paper shows how both problems are solved simultaneously using a multivariate structural time series model that borrows strength over time and space and models the aforementioned discontinuities. The paper illustrates an increased precision when one moves from univariate models to a multivariate one where the domains are jointly modelled. This increase is especially significant in the most recent period when sample sizes get smaller, with standard errors of the design-based estimator of the target variables being reduced by 40 to 70 percent with respect to standard errors of the model-based estimates.

1. Introduction

National statistical institutes (NSIs) mostly confine themselves to the design-based approach when compiling official statistics, using the Horvitz-Thompson estimator or general regression estimator (Särndal *et al*., 1992). An important problem with repeated surveys that is often encountered in official statistics is that survey process modifications may have systematic effects on the outcomes resulting in level breaks, sometimes called discontinuities, as well as in variance breaks. Another issue NSIs frequently have to deal with is providing reliable estimates for subpopulations or domains. Sample sizes are often too small to use design-based estimators for detailed subdivisions of the population. This is where structural time series modelling comes into play in full force due to its ability to efficiently solve both problems simultaneously.

First of all, structural time series modelling can be seen as a form of small area estimation (SAE) where a model is used to increase the precision of a domain estimate with sample information observed in preceding periods or other domains. With a structural time series model, the signal of a target variable can be filtered from the time series that contains sampling and measurement errors. The signal comprises the sum of several unobserved components, such as trend, seasonal and regression components (Harvey, 1989, Durbin and Koopman, 2001). Such components can derive benefit from sample information accumulated in the past, which is sometimes called borrowing strength over time. Modelling the correlation between such components of different domains in a multivariate structural time series model can further improve the model-based domain estimates by taking advantage of sample information from other domains (Pfeffermann and Burck, 1990, Pfeffermann and Bleuer, 1993). This is often referred to as borrowing strength over space. Harvey and Chung (2000) simultaneously modelled the series of survey estimates with another correlated series to improve the precision of the survey estimates. In the case of common trends or other common components, the so-called common factor models can be employed to parsimoniously account for cross-correlation inherent in the system. An application of a common factor model in the context of SAE is given by Krieg and van den Brakel (2012). Pfeffermann and Tiller (2006) proposed to benchmark time series estimates of domains to sufficiently precise direct estimates at the national level as an alternative method to borrow strength over space. This method also provides a form of robustness against model misspecification.

Secondly, structural time series models can handle discontinuities induced by a redesign of the survey. A model can embrace different forms of interventions which account for systematic effects of a survey redesign on the outcomes (Van den Brakel and Roels, 2010). This is a direct application of the state-space intervention approach proposed by Harvey and Durbin (1986). Redesigns can also affect the variance of the direct estimates both directly (through changes in the sample size) and indirectly. For instance, changes in the data collection process can affect the standard errors of the direct estimator through a change in the variance of the measurement errors. If the design variances are available at the NSI, they can be used as prior information in the time series model. This will automatically account for shock-effects and other forms of heteroscedasticity in the variances induced by the survey redesigns (see, e.g., Binder and Dick, 1990; Durbin and Quenneville, 1997). The information on design variances, however, is not always available. In such cases, time series models must account for changes in design variances by making the model variance time-dependent. Effective structural time series modelling could improve the accuracy of time series published by NSIs that rely on the traditional design-based approach from sampling theory. Harvey and Chung (2000) provide an illustrative example for the Labour Force Survey in the UK. However, despite all the advantages of this model-based approach, NSIs are still reluctant to apply these techniques in the production of official statistics, mainly because model misspecification easily results in severely biased estimates. Careful model evaluation and selection is required, which is an additional labourious stage in the production process, particularly for multipurpose surveys, since separate models are required for different variables. To our knowledge, only two governmental statistical institutes use a state-space model in the production of their official figures: Statistics Netherlands with their Dutch Labour Force Survey model (Van den Brakel and Krieg, 2009), and the U.S. Bureau of Labor Statistics (Tiller, 1992, Pfeffermann and Tiller, 2006).

In this paper, structural time series modelling is applied to the Dutch Road Transportation Survey (the DRTS). This survey is interesting from both practical and academic points of view because it features several design transitions that caused multiple level breaks in the series, as well as some design variance breaks visible in the published figures. Moreover, due to reduced budgets, the DRTS faces decreasing sample sizes, which results in an increasing loss of precision in the direct estimates. This paper presents a simultaneous solution to both problems. An additional complication that is addressed in this application is that sampling units can belong to more than one domain, resulting in additional correlation between the domains.

The main contribution of this paper is that it demonstrates that survey estimates can be substantially improved within the framework of a multivariate state-space approach where multiple survey redesigns and other survey process changes are treated as exogenous events by modelling various level and variance breaks. The model also allows for cointegrated trends and a contemporaneous inter-domain correlation in the sampling errors. As the DRTS time series are published both at the national level and at the underlying domain level, the question of how the aggregated series should be treated within the state-space approach is also addressed. Section 2 describes the data and its major discontinuities induced by survey redesigns. Section 3 focuses on the structural time series models employed in the DRTS. Estimation results are presented in Section 4. Section 5 summarises the main findings and offers some possible further improvements.

2. The Dutch Road Transportation Survey

This Section describes the DRTS set-up along with the most important changes it has faced over time. The DRTS is a survey that analyses freight transportation in terms of tons, kilometres and ton-kilometres. The target variables are constructed for international and domestic segments separately. Further, these variables are divided into Hire-and-Reward (HR) and Own-Account (OA) categories according to whether or not transportation is carried out at a cost of the vehicle owner. The application of the present study is domestic OA road freight transportation carried out by vehicles registered at the Dutch Admission Authority for Vehicles. These series are measured in thousands of tons on a quarterly basis from 1976(1) until 2010(4) (where numbers in brackets denote quarters), and is divided into nine categories according to the so-called NSTR-classification (*Nomenclature uniforme des marchandises pour les Statistiques de Transport, Revisée*). This classification is based on the type of goods transported and includes ten categories (short names used hereinafter are given in brackets): 0. agricultural products and live animals (agriculture); 1. foodstuff and animal fodder (food); 2. solid mineral fuels; 3. petroleum oils and petroleum; 4. ores, metal scrap, roasted iron pyrites (ores); 5. iron, steel and non-ferrous metals (including intermediates) (metals); 6. crude and manufactured minerals, building materials (minerals); 7. fertilizers; 8. chemicals; 9. vehicles, machinery and other goods (including cargo) (other goods). The enumeration in the present paper begins from 1, with NSTR 2 and 3 being combined in domain 3 (oil). The analysis is therefore based on nine target variable series

which are called domains in this paper. Since some vehicles transport goods from different categories, they may appear in more than one domain. As a result, an additional correlation arises between such domains. This makes this application different from the traditional situation where sampling units belong to one domain only.

The estimation procedure of this survey is based on the Horvitz-Thompson (HT) estimator (Horvitz and Thompson, 1952, Narain, 1951). This is a design-based estimator that expands the observations by weights obtained as inverse inclusion probabilities of the sampled units. The HT point estimates of the own-account domestic transportation series, which are officially published by Statistics Netherlands, are shown in Fig. 1.1.1 (StatLine.cbs.nl). Over the years, the DRTS has undergone a number of methodological changes. Various amendments, more and less significant, were particularly frequent during the last decade. They caused breaks both in the level of the series and in their design variances. Only the most important changes that might have caused discontinuities are mentioned here. They are approached within the structural time series framework with the help of level interventions and by modelling heteroscedasticity in the survey errors. One of the major changes in the survey design is switching between one- and two-stage sampling schemes.

Currently, the DTRS has a one-stage stratified sampling design where vehicles (sampling units) are drawn directly from the database of the Dutch Admission Authority for Vehicles on a quarterly basis. The stratification scheme changed several times. It is based on variables related to the vehicle characteristics as well as to the total load capacity of the owning companies and industrial sectors these companies belong to. The vehicle transportation performance is observed during one week. The net-sample size in the most recent decennium has been fluctuating between 9,000 and 12,000 vehicles per quarter.

Until 2003, the DTRS was based on a one-stage stratified sample of vehicles. From 2003 to 2007, a stratified two-stage sampling design with the company as a primary sampling unit, and the vehicle as a secondary sampling unit was applied. Apart from that, several survey process changes took place in 2003, such as the introduction of a new questionnaire and the transition from a three-day to week-wise reporting period. These and other changes introduced in 2003 had a considerable effect on the level and design variance of the series point estimates. For instance, domain 7 (fertilisers) suffered an increase in the level in 2003 and an unexpected dip during eight quarters of 2007–2008. An analysis of other statistical figures related to this goods category did not reveal any particular factor that could have caused real changes in this domain of the size reflected in Fig. 1.1.1. For modelling purposes, the level shift in 2003 will be regarded as being caused by the survey redesign, whereas the dip in the level in 2007–2008 will be removed with an outlying intervention.

Since 2008, the sample design has been based on a stratified sample of vehicles. Since then, stratification has been based on the vehicle type and load capacity, as well as on the total load capacity of the company owning the vehicle and on the industry branch the company belongs to. Vehicles drawn at the beginning of every quarter are further clustered by their owners. The latter are, in turn, randomly assigned to different weeks in a quarter. This means that, since 2008, the sample design has still been based on a two-stage sampling. Apart from that, certain ambiguities arose in the late 2000's regarding the classification of shipments into the OA and HR categories. This caused a shift mainly in domestic transporters from the former to the latter category. Since 2008, big companies have gradually been joining the XML reporting system, which was used alongside with paper and electronic questionnaires.

Finally, the year 2010 saw an introduction of an automatic editing system and new electronic questionnaires. These change, however, did not have any salient effect on the pattern of the design estimates.

In 1994, small vehicles with a load capacity less than 1.5 tons were excluded from the OA series. From 1997 to 2002, these vehicles, most of which are vans, were included in the survey again. This caused a considerable upward shift in the level of domain 9 and is therefore modelled as a level intervention for this particular period of time. Another temporary level intervention is used to model a huge peak in 2003 in domain 9. Since 2003, vans, together with special vehicles, have been monitored using a separate smaller questionnaire. Starting from 2009, they are no longer surveyed. Since then, total weight transported by these vehicles has been estimated using register information from another national authority called 'Nationale AutoPas', which contains the number of kilometres the vehicle has covered each time it visits a vehicle service station. This observation method is obviously less precise, so the inclusion of van freight adds uncertainty to the series. Therefore, Statistics Netherlands decided to publish two different series: one that includes van freights, and the other that does not. However, due to problems in 2003 related to the identification of vans and other types of small vehicles, the series without van freights is available only from 2004, leaving the series with the peak in the four quarters of 2003. In different domains, the above-mentioned survey changes are reflected to a different extent, but it is clearly visible that every domain, except for domain 4, exhibits an increase in variation in 2003 (see Fig. 1.1.1). This can be explained by decreasing sample sizes over time and by the above-mentioned changes in the survey design. The survey design variances are very high in the last decennium of the time series, which results in large fluctuations in the point-estimates. To produce more stable estimates, a multivariate structural time series model is developed in the next section as a form of SAE. This model must also account for the heteroscedasticity in survey errors, as well as for breaks in series levels. The most efficient way of dealing with survey error variance breaks would be using design variance estimates or at least sample sizes. The problem with the DRTS is that neither is available. The remaining option is to treat the sampling variances as piecewise constant.

2.1.1. Horvitz-Thompson estimates of the own account domestic transportation series, 1976(1)-2010(4), 1000 tons.
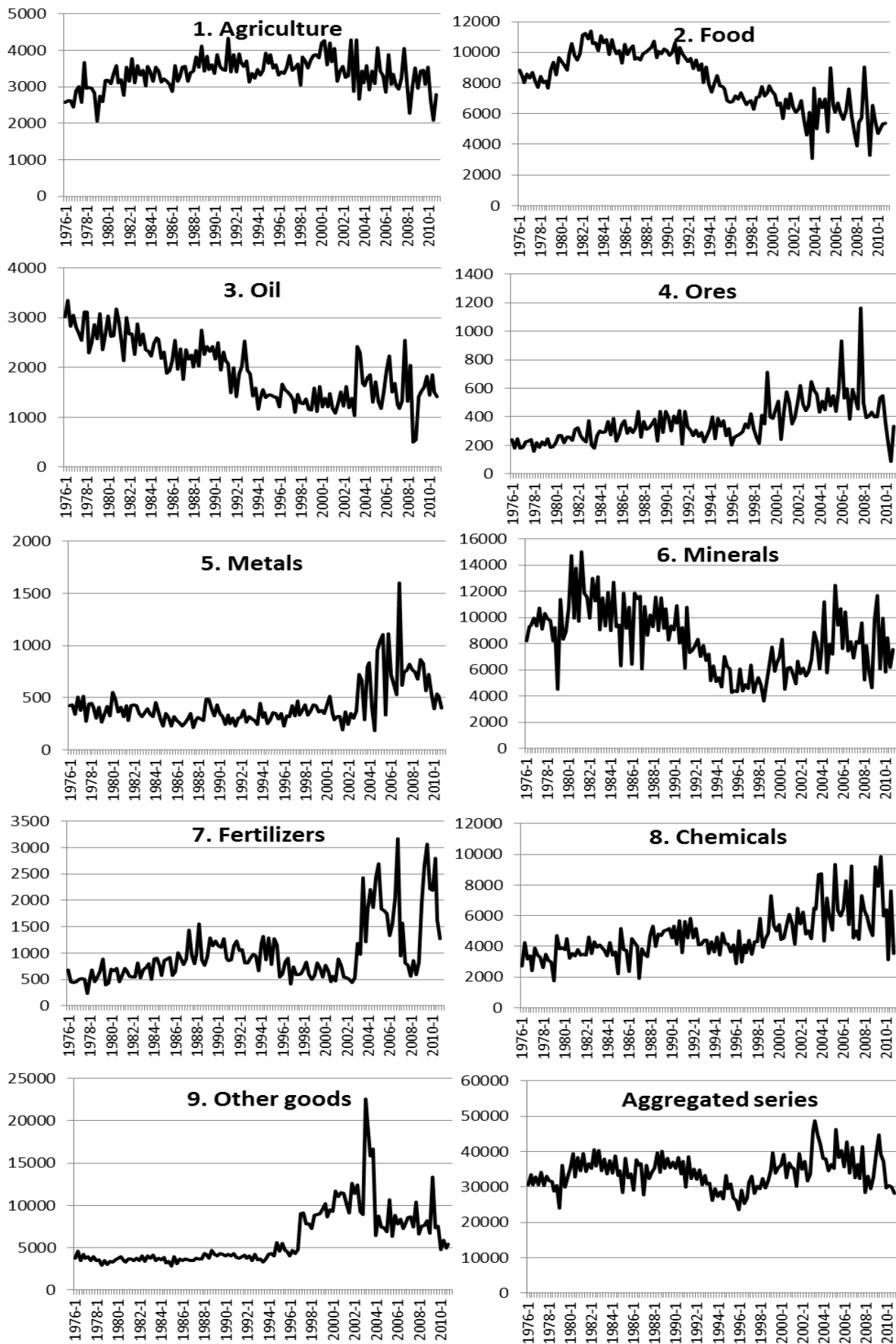
3. Structural time series models and methods employed

In this section, structural time series models are developed for the DRTS series. The core of this technique is decomposing the series into a number of unobserved components, e.g., trend, seasonal, other cyclical and regression components. The variables comprising these unobserved components are referred to as *state variables*. Structural time series models are often expressed in a state-space form and analysed with the Kalman filter (see Durbin and Koopman, 2001). Separate series at the domain level and at the national level can be modelled individually, i.e. in a univariate setting, where the variance reduction comes from borrowing information over time. However, the potential of the state-space modelling technique is exploited to a much larger extent when the domains are jointly modelled in a multivariate setting, which makes it possible to borrow information both over time and domain space. Subsection 3.1 will give the overview of a general model for D domains. The estimation results, together with univariate models as special cases of this D-dimensional model, will be presented in the subsequent section.

3.1 Structural time series model specification

This subsection provides a theoretical set-up for the chosen statistical model designed for D domains. Let $y_{i,t,d}$ denote a target variable observation at sampling unit i at time t in domain d. Generally, survey observations are subject to measurement errors, which gives rise to the following measurement error model: $y_{i,t,d} = u_{i,t,d} + M_{i,d} + w_{i,t,d}$. Here, $u_{i,t,d}$ is the true but not directly observable parameter of interest at the unit level at time t in domain d; $M_{i,d}$ is a systematic measurement error at the unit level, and $w_{i,t,d}$ is a random measurement error. At the domain level, the unknown population parameter of interest at time t is denoted as $u_{t,d}$. In this application, the parameter of interest is the total number of tons transported, so $u_{t,d}$ is defined as the sum of $u_{i,t,d}$ over all the population units belonging to domain d in the hypothetical situation of a complete enumeration. In order to estimate this parameter, a sample is drawn at each time period t.
Let $\hat{Y}_{t,d}$ denote a series of HT estimates for the unknown population parameter. These HT estimates constitute an input series for the time series model and are expressed as $\hat{Y}_{t,d} = \theta_{t,d} + e_{t,d}$, where $e_{t,d}$ is the total effect of a sampling error and random measurement errors $w_{i,t,d}$; and $\theta_{t,d}$ is the parameter of interest obtained when the whole population is surveyed under a particular survey design. In this way, $\theta_{t,d} = u_{t,d} + M_d$ is the sum of the true population parameter $u_{t,d}$ and a measurement bias M_d of the survey design at the domain level. These two terms cannot be separated from each other with the survey data at hand. Durbin and Quenneville (1997) illustrate how the measurement bias can be estimated if exact or more accurate benchmarks are available, e.g., some bias-free auxiliary series or annual data.

Let M_d refer to the design applied at the beginning of the time series. When the survey undergoes some kind of redesign, the measurement error M_d shifts to another level, say M_d'. Assuming that the difference $M_d - M_d' \equiv \beta_d$ is constant over time, it can be modelled with a level intervention using a dummy regressor. This gives rise to the following representation of the observed series: $\hat{Y}_{t,d} = \theta_{t,d} + \delta_{t,d}\beta_d + e_{t,d}$, where $\delta_{t,d}$ is a

dummy regressor that switches from zero to one when a redesign takes place. These dummy regressors are used not only for level shifts, but also for outliers. Parameter $\theta_{t,d}$ can be decomposed into a trend L_t, a seasonal component γ_t, and an irregular term ε_t, so that $\theta_{t,d} = L_{t,d} + \gamma_{t,d} + \varepsilon_{t,d}$. In a cross-sectional survey like the DRTS, it is not possible to separate the irregular term $\varepsilon_{t,d}$ from the sampling error $e_{t,d}$. Therefore, these two are combined into one composite error term $v_{t,d} = e_{t,d} + \varepsilon_{t,d}$. It is assumed that $v_{t,d}$ is dominated by the sampling error $e_{t,d}$ and is normally and independently distributed with a zero-expectation (see Krieg and Van den Brakel (2012) for empirical evidence that $e_{t,d}$ is indeed dominated by the sampling error). Due to changes in the survey design, the sampling error variance often changes, and so does the variance of the term $v_{t,d}$. Structural time series models account for sudden changes in the input series precision by modelling these variance breaks with time dependent variances of the composite error terms.

Consider a domain where redesigns resulted in K_d discontinuities, or level shifts. The aforementioned considerations imply the following model for the domains:
$$\hat{Y}_{t,d} = L_{t,d} + \gamma_{t,d} + \delta_{t,d,1}\beta_{d,1} + ... + \delta_{t,d,K_d}\beta_{d,K_d} + v_{t,d}, \quad d \in \{1, ..., D\}, \text{(3.1)}$$
where $\delta_{t,d,k}$ is equal to one during the *k*-th level intervention, and otherwise to zero. Each of the *D* models above uses sample information from previous periods through the trend and seasonal components in order to improve the precision of the $\theta_{t,d}$-estimates. Modelling the correlation between the trend and seasonal disturbances of different domains can further improve the precision of the $\theta_{t,d}$-estimates with the help of sample information from other domains.

Borrowing strength over both time and space can be implemented within the scope of a multivariate structural time series model constructed by stacking the *D* univariate models as in (3.1). This results in the following vector notation:
$$\hat{\boldsymbol{Y}}_t = \boldsymbol{L}_t + \boldsymbol{\gamma}_t + \boldsymbol{x}_{t,1}\beta_1 + \cdots + \boldsymbol{x}_{t,K}\beta_K + \boldsymbol{v}_t, \text{(3.2)}$$
where all the vectors have dimension *D*. In this notation, K is equal to the sum of K_d over all the domains and stands for the total number of level interventions in the model. Vectors $\boldsymbol{x}_{t,k}$ contain domain-specific dummy regressors $\delta_{t,d,k}$, and otherwise zeros.

In the present application, the so-called *smooth trend model* was chosen to model the trends $L_{t,d}, d \in \{1, ..., D\}$. This model is well-known in the econometric literature for its reasonable flexibility and parsimony (Durbin and Koopman, 2001, Ch. 3, Harvey, 2000). The smooth trend model (or the integrated random walk) can be defined by the following two equations for series *d*:
$$L_{t,d} = L_{t-1,d} + R_{t-1,d},$$
$$R_{t,d} = R_{t-1,d} + \eta_{R,t,d}, \tag{3.3}$$
where the state variables $L_{t,d}$ and $R_{t,d}$ are the level and slope of the *d*-th series, respectively. In the multivariate setting, the slope disturbance terms are assumed to be

$$\text{Cov}(\eta_{R,t,d}, \eta_{R,t',d'}) = \begin{cases} \sigma_{\eta R,d}^2 & \text{if } t = t' \text{and } d = d', \\ \varsigma_{\eta R,d,d'} & \text{if } t = t' \text{and } d \neq d', d \in \{1, ..., D\} \\ 0 & \text{if } t \neq t'. \end{cases}$$

For the seasonal component $\gamma_{t,d}$, a trigonometric model is assumed (Hannan *et al.*, 1970, Koopman *et al.*, 1999a-b, Harvey, 1989). This model is widely applied in econometric time series modelling. If *s* denotes the number of seasons, then the model is defined as:
$$\gamma_{t,d} = \sum_{j=1}^{[s/2]} \gamma_{t,d,j}, \quad d \in \{1, ..., D\}, \text{where } s=4 \text{ for quarterly data, and thus the seasonal}$$
component consists of two harmonics $\gamma_{t,d,j}$. The first harmonic is generated according to the following process:

$$\gamma_{t,d,1} = \cos\frac{\pi}{2}\gamma_{t-1,d,1} + \sin\frac{\pi}{2}\gamma_{t-1,d,1}^* + \omega_{t,d,1},$$

$$\gamma_{t,d,1}^* = -\sin\frac{\pi}{2}\gamma_{t-1,d,1} + \cos\frac{\pi}{2}\gamma_{t-1,d,1}^* + \omega_{t,d,1}^*, \quad d \in \{1, \dots, D\},$$

where disturbances $\omega_{t,d,1}$ and $\omega_{t,d,1}^*$ are uncorrelated. The last harmonic, in this case the second one, is always generated by only one stochastic variable:
$\gamma_{t,d,2} = -\gamma_{t-1,d,2} + \omega_{t,d,2}, \, d \in \{1, \dots, D\}$. The following covariance matrix structure is assumed for the seasonal disturbance terms (except for $\omega_{t,d,2}^*$):

$$\mathrm{Cov}\left(\omega_{t,d,j}, \omega_{t',d',j'}\right) = \mathrm{Cov}\left(\omega_{t,d,j}^*, \omega_{t',d',j'}^*\right)$$
$$= \begin{cases} \sigma_{\omega,d,j}^2 & \text{if } t = t', d = d' \text{and } j = j', \\ \varsigma_{\omega,d,d',j} & \text{if } t = t', d \neq d' \text{and } j = j', \ j \in \{1,2\}, d \in \{1, \dots, D\}, \\ 0 & \text{if } t \neq t' \text{ or } j \neq j'. \end{cases}$$

The last expression states that the variances of the two disturbances $\omega_{t,d,1}, \omega_{t,d,1}^*$ are the same. Apart from that, the model can be simplified by assigning the same variance to the disturbances belonging to different harmonics. The equality of the two harmonics' hyperparameters has been tested in a univariate setting using the likelihood ratio test. It turns out that the null hypothesis for the two variances' equality could not be rejected at any reasonable significance level, suggesting that $\sigma_{\omega,d,1}^2 = \sigma_{\omega,d,2}^2 = \sigma_{\omega,d}^2, d \in \{1, \dots, D\}$. The correlation between the corresponding harmonics of different domains can be modelled in order to borrow information over space. In this application, however, the covariances $\varsigma_{\omega,d,d',j}$ are set to zero, since their modelling results in a very little variance reduction. This is due to the fact that, first of all, the seasonal effect size has turned out to be relatively small compared to the magnitude of the corresponding trends, and, secondly, most of the seasonal component hyperparameters have been estimated to be not significantly different from zero (domains 1, 3, 4, 5, 7, 9).

A distinctive feature of the DRTS is that its sample design has been changed several times in the course of the survey. The consequences thereof have been two-sided. First of all, multiple survey redesigns caused shifts in the series level. To account for it, the model is augmented with K auxiliary variables $\boldsymbol{x}_{t,k}$ (that stand for level breaks and, possibly, outliers in the series) and with K regression coefficients β_k, each of which is meant for a specific domain (hence the omission of the subscript d). Each of the vectors $\boldsymbol{x}_{t,k} = \left(0, \dots 0, \delta_{t,k}, 0, \dots 0, \delta_{t,k}\right)'$, being D-dimensional, contains D-1 zeros and a dummy variable $\delta_{t,k}$. The variable $\delta_{t,k}$ takes a value of one for the time points when the intervention is effective and a zero value for time points before and optionally after the intervention in question. The position of $\delta_{t,k}$ in vector $\boldsymbol{x}_{t,k}$ is defined by the number of the domain the intervention is effective for. A distinctive feature of this application is that level breaks do not generally occur in all the domains (remember the case of domain 9 in Section 2) and therefore are modelled individually for every domain, i.e. each domain's intervention starts at a particular point in time and lasts for a particular number of periods, during which the intensity of the intervention remains constant and domain-specific. The regression coefficient estimate $\hat{\beta}_k$, k=1,…, K, is time-invariant and can be interpreted as an estimated impact of the k-th outlier or discontinuity induced by a survey redesign. The discontinuities in the series mainly occurred due to major survey redesigns described in the previous section. The domains' respective $\delta_{t,k}$ will be introduced as discontinuities and outliers in Subsection 4.1.

The second consequence of the redesigns is that they affected the variance of the HT estimates, which had a discernible effect on the variance of the composite error term $\nu_{t,d}$.

One way to model heteroscedasticity in the $v_{t,d}$-term is using the variance of the HT estimates as prior information in the model by defining $\sigma_{v,t,d}^2 = \hat{\sigma}_{HT,t,d}^2 + \sigma_{\varepsilon,d}^2$. Another approach, proposed by Binder and Dick (1990), as well as by Durbin and Quenneville (1997), suggests that $\sigma_{v,t,d}$ is modelled proportionally to the standard error of the HT estimator. In this application, the variances of the HT estimator are unfortunately not available, and the micro-data for their calculation is available only for the most recent years. It would be possible to model variances $\sigma_{v,t,d}^2$ proportional to the sample sizes, but these are not available either. Therefore, the variance $\sigma_{v,t,d}^2$ of the composite error term is modelled as time-varying by allowing different $\sigma_{v,t,d}^2$-values for a few sub-periods specific for each domain.

It should be noted that the composite error terms $v_{t,d}$ have non-zero covariances because the present domains do not represent independently drawn samples, but rather a sub-division of the target variable into different types of goods transported. The fact that certain sampling units may transport different freight categories and thus appear in different domains causes non-zero covariances between these domains. The time-dependent covariance matrix of the composite error terms is denoted by $\boldsymbol{R_t = E(v_t v_t{'})}$ whose elements are assumed to have a zero-expectation and be normally distributed with the following properties:

$$\text{Cov}(v_{t,d}, v_{t',d'}) = \begin{cases} \sigma_{v,t,d}^2 & \text{if } t = t' \text{and } d = d', \\ \varsigma_{v,t,d,d'} & \text{if } t = t' \text{and } d \neq d', \quad d \in \{1, \dots, D\} \\ 0 & \text{if } t \neq t'. \end{cases} \tag{3.4}$$

Vector $\boldsymbol{v_t}$ is also independent on the disturbance terms of the state variables described above. As it has already been mentioned, these terms' variances are dominated by the sampling error and therefore vary over time. This implies that covariances between the composite error terms of different domains are time-varying. The full covariance matrix could be modelled, but this will result in a large number of covariance hyperparameters for different sub-periods. Therefore, another more parsimonious alternative is considered. It is based on the assumption of time-invariant correlations between the composite error terms of different domains. At the first stage, the composite error terms and their variances are estimated from a model where the covariances between these terms are restricted to zero. Next, these error terms are standardized as $\hat{v}_{t,d}^{St.} = \dfrac{\hat{v}_{t,d}}{\hat{\sigma}_{v,t,d}}$ and the correlations between the $\hat{v}_{t,d}^{St.}$-terms are calculated. This information is used at the next estimation stage in order to approximate the time-varying covariance terms in (3.4) as:

$$\varsigma_{v,t,d,d'} = \hat{\sigma}_{v,t,d} \hat{\sigma}_{v,t,d'} \widehat{\text{Cov}}\left(\frac{\hat{v}_{t,d}}{\hat{\sigma}_{v,t,d}}, \frac{\hat{v}_{t,d'}}{\hat{\sigma}_{v,t,d'}}\right), d \in \{1, \dots, D\}.$$

The obtained model is compared with a model where the covariances between the domains are assumed to be equal to zero. In the present application, it has been found that point-estimates are not affected when the model is augmented with the time-invariant correlations. Only a few domains experience a variance reduction in the state variable estimates, but this effect is negligible. These results are not presented in this paper. For simplicity and with little loss in precision of the state estimates, we proceed by restricting the covariances in question to zero.

To summarize, a time series model for domain estimates has been obtained in order to produce more reliable indicators for the evolution of the parameters of interest. There are two reasons why the time series model estimates can be expected to be more stable and reliable compared to the survey estimates. Firstly, the effective sample size for model

estimates for a particular period and domain is increased by using sample information from the preceding periods and other domains. Secondly, sudden differences in the measurement error level caused by different redesigns are excluded from these time series model estimates if such level changes are explicitly modelled with intervention variables. From this point of view, the focus of interest lies on the trend and signal estimates, the latter being the sum of the trend and seasonal estimates. Depending on whether or not a survey modification or a change in the population of interest (as in domain 9) are viewed as an improvement, level interventions can be considered to be part of the signal estimates. This paper mainly focuses on the analysis of the trend and signal estimates, the signal being defined as the sum of the trend, seasonal, and level interventions.

3.2 Model estimation setting

Structural time series models are generally put in a state-space form and analysed with the Kalman filter, see Harvey (1989), Durbin and Koopman (2001). The unobserved components described in Subsection 3.1 – the level, slope, $(s-1)$ seasonal harmonics per each series, and the β_k-regression components – form the so-called state vector $\boldsymbol{\alpha}_t$ of dimension m:

$$\boldsymbol{\alpha}_t = \left(\boldsymbol{\alpha}_t^L \ \boldsymbol{\alpha}_t^\gamma \ \boldsymbol{\alpha}_t^\beta\right)', \text{ where}$$
$$\boldsymbol{\alpha}_t^L = \left(L_{t,1} \ R_{t,1} \ ... \ L_{t,D} \ R_{t,D}\right),$$
$$\boldsymbol{\alpha}_t^\gamma = \left(\gamma_{t,1,1} \ \gamma_{t,1,1^*} \ \gamma_{t,1,2} \ ... \gamma_{t,D,1} \ \gamma_{t,D,1^*} \ \gamma_{t,D,2}\right),$$
$$\boldsymbol{\alpha}^\beta = (\beta_1 \ ... \ \beta_K).$$

A state-space form comprises two types of equations. The first one is the measurement or signal equation: $\widehat{\boldsymbol{Y}}_t = \boldsymbol{Z}_t \boldsymbol{\alpha}_t + \boldsymbol{v}_t$, where $\widehat{\boldsymbol{Y}}_t = \left(\widehat{Y}_{t,1}, ..., \widehat{Y}_{t,D}\right)'$ denotes the input vector with design estimates. The measurement equation reflects the relation between the observed design $\widehat{\boldsymbol{Y}}_t$ estimates and the vector of m unobserved state variables $\boldsymbol{\alpha}_t$ through a time-dependent design matrix $\boldsymbol{Z}_t$. Furthermore, $\boldsymbol{v}_t$ denotes a vector with composite error terms presented in the previous sub-section. The other equation – the transition equation – describes how each state variable evolves over time: $\boldsymbol{\alpha}_{t+1} = \boldsymbol{T}\boldsymbol{\alpha}_t + \boldsymbol{\eta}_{t+1}$, through a time-invariant design matrix $\boldsymbol{T}$. Vector $\boldsymbol{\eta}_t$ is assumed to be a zero-expectation vector of normally, identically and serially independently distributed state disturbances with the contemporary covariance matrix $\boldsymbol{Q} = E(\boldsymbol{\eta}_t \boldsymbol{\eta}_t')$.
The transition matrix is defined as $\boldsymbol{T} = Blocdiag\left[\boldsymbol{T_L} \ \boldsymbol{T_\gamma} \ \boldsymbol{T_\beta}\right]$ with:

$$'\boldsymbol{T_L} = \boldsymbol{I}_{[D]} \otimes \boldsymbol{T_L},$$
$$\boldsymbol{T_\gamma} = \boldsymbol{I}_{[D]} \otimes \boldsymbol{H},$$
$$\boldsymbol{T_\beta} = \boldsymbol{I}_{[K]},$$

where $\boldsymbol{T_L} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$, $\otimes$ is the Kronecker product which multiplies every element of the first matrix by the second matrix, $\boldsymbol{I}_{[p]}$ is a p-dimensional identity matrix, and $\boldsymbol{H}$ is the design matrix for the seasonal component's harmonics:

$$\boldsymbol{H} = \begin{bmatrix} \cos\frac{2\pi}{s} & \sin\frac{2\pi}{s} & 0 \\ -\sin\frac{2\pi}{s} & \cos\frac{2\pi}{s} & 0 \\ 0 & 0 & -1 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix}. \tag{3.5}$$

The transition matrix T_β for the dummy regression coefficients contains a K-dimensional identity matrix $I_{[K]}$ according to the number of discontinuities and outliers explicitly modelled in the domains. The structure of the T_β-matrix, together with a zero-variance of the corresponding state stochastic terms, means that the discontinuities and outliers evolve as $\beta_k = \beta_{k-1}$, k=1,..., K. The design matrix Z_t for the measurement equation has the following form:

$$Z_t = \begin{bmatrix} I_{[D]} \otimes (1 \quad 0) & I_{[D]} \otimes (1 \quad 0 \quad 1) & Z_{\beta,t} \end{bmatrix},$$

where $Z_{\beta,t}$ consists of K vertical vectors $x_{t,k}$ described in Subsection 3.1:

$$Z_{\beta,t} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \delta_{t,5} \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \delta_{t,1} & \delta_{t,2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \delta_{t,3} & \delta_{t,4} & 0 \end{bmatrix}.$$

The state noise covariance matrix $Q = E(\eta_t \eta_t') = Blockdiag[Q_L \; Q_\gamma \; Q_\beta]$ consists of the following parts:

$$Q_L = \begin{bmatrix} 0 & 0 & 0 & 0 & & 0 & 0 \\ 0 & \sigma_{\eta R,1}^2 & 0 & \varsigma_{\eta R,1,2} & & 0 & \varsigma_{\eta R,1,D} \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & \varsigma_{\eta R,2,1} & 0 & \sigma_{\eta R,2}^2 & & 0 & \varsigma_{\eta R,2,D} \\ & & \vdots & & \ddots & & \vdots \\ 0 & 0 & 0 & 0 & & 0 & 0 \\ 0 & \varsigma_{\eta R,D,1} & 0 & \varsigma_{\eta R,D,2} & \cdots & 0 & \sigma_{\eta R,D}^2 \end{bmatrix}.$$

$$Q_\gamma = Blockdiag[\Gamma_1 \ldots \Gamma_D], \quad \Gamma_d = \begin{bmatrix} \sigma_{\omega,d,1}^2 & 0 & 0 \\ 0 & \sigma_{\omega^*,d,1}^2 & 0 \\ 0 & 0 & \sigma_{\omega,d,2}^2 \end{bmatrix}, \quad d \in \{1,\ldots,D\},$$

$Q_\beta = 0_{[K \times K]}$ since the regression coefficients are modelled as time-independent. The measurement equation error term covariance matrix $R_t = E(v_t v_t')$ is diagonal, if domains are not overlapping:

$R_t = Diag(\sigma_{v,t,1}^2 \; \sigma_{v,t,2}^2 \ldots \sigma_{v,t,D}^2)$. The model proposed in Subsection 3.1 contains non-stationary state variables and time-invariant regression coefficients. The state variables are initialised with a diffuse state vector to which the exact initial Kalman filter is applied as in Koopman (1997).

The estimation of the slope disturbance covariance matrix, say Q_R, that is a part of matrix Q, is carried out through the Cholesky decomposition of the form $Q_R = ADA'$, where A is a lower triangular matrix of orthonormalised eigenvectors with ones on the main diagonal, and D is a diagonal matrix of eigenvalues. The slope (co)variance corresponding to the aggregated series are derived on the basis of equations (3.4) and (3.5). The Cholesky decomposition ensures that the maximum likelihood estimate for the matrix Q_R is positive-(semi)definite.

Another advantage of this decomposition is that by detecting some dependent series through the D-matrix one can implement the concept of cointegration (the so-called common factor model) and thus reduce the number of hyperparameters to be estimated. Detecting and modelling common factors further reduces the signal variance. In this application, a common factor model for trends is implemented by modelling the slope disturbance covariances. The existence of a common trend specification implies that the trends of D domains are driven by a smaller number of underlying stochastic trends. If an eigenvalue of matrix D is equal to zero, the corresponding domain's stochastic part of the trend can be expressed as a linear combination of the other domains' stochastic trends. Insignificant eigenvalues of the slope disturbance covariance matrix are removed step-wise until the number of common trends is identified. Detecting and modelling cointegrated stochastic trends allows to obtain a more parsimonious model and to interpret the relationship between the D domains. The cointegration concept and the relative testing procedures in the context of state-space models are presented, e.g., in chapter 8 of Harvey (1989), Nyblom and Harvey (2001), as well as in Koopman *et al*. (1999).

The Kalman filter assumes that the hyperparameters in matrices Q and R_t, i.e. the state and measurement equation disturbance (co)variances, are known, but it is generally not the case. These hyperparameters are replaced by their maximum-likelihood estimates. The numerical procedure used to solve this nonlinear optimisation problem is the Broyden–Fletcher–Goldfarb–Shanno method (MaxBFGS in the OxMetrics package). The analysis is conducted with OxMetrics.5 (Doornik, 2007) in combination with SsfPack 3.0 package (Koopman *et al*., 1999, 2008). The variances of the Kalman filter estimates reported in this paper do not account for additional uncertainty caused by the replacement of the hyperparameters with their maximum-likelihood estimates. The variance hyperparameters are estimated on the log-scale to avoid negative variance estimates. These estimates are presented in the appendix.

The Kalman filter produces what is called *filtered* estimates, which are the optimal state variable estimates on the basis of information accumulated up to and including period t. These estimates can be improved by various smoothing algorithms, where information is pooled over the entire time span. Smoothed estimates are usually treated as the most realistic ones, since they are based on the entire set of information available. This is true when the focus is on the analysis of unobserved components of the time series model. In the context of this paper, however, structural time series models are used for production purposes in official statistics. In this case, the focus is on filtered, rather than smoothed estimates, since the former ones better reflect what can be obtained with this modelling approach in the real production process, when statisticians have information at their disposal only up to (and including) time t. However, filtered estimates do not fully imitate the real-time production of official estimates, since the hyperparameter maximum-likelihood estimates in this case are still based on the whole length of the time series. Therefore, we will concentrate on the so-called *concurrent* estimates. Concurrent estimates for period t are based on the Kalman filter, with the hyperparameter maximum-likelihood estimates also being based on the information available up to and including period t. In this way, concurrent estimates exactly reflect the real production process outcomes and therefore are more realistic compared to filtered estimates.

3.3 Aggregated series estimation

In the DRTS, the aggregated series design estimates are not sufficiently precise due to the sample size reduction and sample design modifications, particularly in the second half of the time period under consideration. Fig. 2.1.1 shows that the variability of the observed series does indeed increase after 2003. If the direct estimates at the aggregated level were sufficiently precise, they could be directly used as benchmarks for the domain model estimates, which is frequently done in SAE. This alternative way to borrow strength over space also provides a form of built-in robustness against model misspecification, see Pfeffermann and Burck (1990), Pfeffermann and Bleuer (1993), and Pfeffermann and Tiller (2006). Durbin and Quenneville (1997) exploit aggregation over the time dimension. Namely, they benchmark monthly or quarterly survey estimates to exact or substantially more precise annual figures.

However, the high volatility of the DRTS design estimates at the national level suggests that they themselves could benefit from the structural time series approach. There are a few ways in which this can be done. First of all, a univariate model could be developed for the aggregated series. This may seem to be a good option since the signal-to-noise ratio of the aggregated series is higher than that of the underlying domains. Another approach is to derive the aggregated estimates as a linear combination of domain estimates from a D-dimensional multivariate model. Which approach is more efficient in terms of signal variances, is an empirical question.

If the univariate framework offers smaller variance estimates for variables of interest, it may be preferred to model the aggregated series in a univariate setting separately from the underlying domains. This, however, will inevitably result in differences between figures published at the national level and the sum of published domain model estimates. In official statistics, the method of Lagrange multipliers can be considered to adjust the estimates proportionally to their mean squared error, so that the sum over the domain estimates exactly equals the aggregated level estimates. The Lagrange method, however, requires the availability of covariance estimates between the point-estimates of each of the domains and of the aggregated series. Assuming a diagonal covariance matrix structure will produce distorted point- and variance estimates. Namely, the aggregated signal variance has turned out to be heavily underestimated in this application.

On the one hand, it may seem useful to jointly model the domains and aggregated series in one $(D + 1)$-dimensional setting with a restriction that makes the sum over the domain state variables and disturbance terms respectively equal to the corresponding states and disturbances of the aggregated series. In this way, the series with a better signal-to-noise ratio would enter the multivariate model, and the sum of the domain model estimates would be equal to the aggregated series model estimates at each point in time. On the other hand, it can be argued that no additional information would enter the model in this case. This effectively means that a $(D + 1)$-dimensional approach must be identical to a D-dimensional multivariate model. A simulation has been performed (results presented in the appendix) that confirms that the D - and $(D + 1)$-dimensional approaches produce identical outcomes, though only if the known hyperparameter values are used in the Kalman filter. Both models are also applied to the DRTS series, and the outcomes are comparable.

Thus, the choice for obtaining aggregated figures lies between deriving them from a D-dimensional model and modelling the aggregated series univariately. Which approach is going to be chosen by an NSI for official figure production, depends, firstly, on the accuracy of the aggregated figures obtained under both approaches, and, secondly, on whether the equality between the sum over domain estimates and the corresponding estimate for their aggregate should be maintained in publications.

4. Model selection and estimation results

The first part of the present section presents the development of univariate models for the ten series separately. Subsection 4.2 describes a multivariate model based on the nine univariate models for the DRTS domains. The common factor structure of the multivariate model is given in Subsection 4.3. Finally, empirical findings are summarized in Subsection 4.4.

4.1 Univariate models for the nine domains and the national level series

The univariate analysis for each of the nine domains is a special case of the nine-dimensional model presented in (3.2) with no correlations among the slope disturbance terms. The univariate analysis is conducted for several purposes. First of all, it allows detecting outliers and discontinuities in the level and in the variance of the measurement equation term. Secondly, it is of great interest to compare the performance of univariate models with that of multidimensional models in terms of variance reduction in the signals and state variables. As regards the model choice for the trend, the preference has been given to the smooth trend model, as in (3.3). This model is a special case of the local linear trend model, whose level equation also has a stochastic term. The absence of this term makes the trend less volatile. The local linear trend model has also been tested and proved to produce very volatile trends, indicating that the data are overfitted with this model. For these reasons, the smooth trend model is preferred. An additional (third) 'acceleration' component (see Harvey (1989) Ch.6.1.5 for the quadratic trend model) has also been tested and found to have no added value.

The detection of outliers and discontinuities is based on the pattern of the available HT estimates, taking into account that little detailed information on sample redesign and sample sizes is available, especially before 2003. Apart from that, while some domains are visibly affected by a certain intervention, this effect may not be observable in other domains. Therefore, taking into account all the available information on the survey redesigns, as summarised in Section 2, one must be selective when trying to implement those changes in a model. In this application, only three domains need level interventions or outliers.
Table 4.1.1 provides a compact summary of level and variance interventions implemented in the univariate models. The nine domain series serve as the basis for constructing a multivariate nine-dimensional model. The presence of level breaks and outliers can be detected using the auxiliary residual diagnostics as in Harvey and Koopman (1992). Variance breaks for the measurement equation disturbance terms require additional hyperparameters in this application. These breaks were tested with the help of the likelihood ratio test. This test suggests that four variance breaks for the measurement equation disturbances should be modelled for the aggregated series in a univariate setting, i.e. for five different periods of time. Out of these five, the second and forth periods share the same hyperparameter. As for level breaks, domain 9, for instance, has two breaks attributed to changes in the population of interest. Further, another level intervention for the period after 2003(4) has been found to be insignificant, which suggests that the change between the series levels before 1997(1) and after 2003(4) is captured by the stochastic trend. In a univariate setting, the only break

identified as significant in the aggregated series, is the one for the four quarters of 2003, while the multivariate setting suggests that all the level interventions, including the insignificant ones, enter the aggregated series as a result of the summation of the domain estimates.

The selection of univariate models was based on the likelihood ratio test, as well as on three tests on the normality and independence of standardised innovations: the Doornik-Hansen normality test (see Doornik and Hansen, 1994), the Durbin-Watson test of first-order serial correlation, and a two-sided F-test for heteroscedasticity. Table A.1 in the appendix contains maximum likelihood estimates of the hyperparameters as well as the model evaluation measures of the finally selected univariate models for the ten series. Some seasonal hyperparameters have turned out to be close to zero with vast standard errors. Therefore, they were removed from the model, whereupon it was checked for a decrease in the likelihood function maximum.

The results suggest that only domains 2, 6 and 8 need a stochastic term for their seasonal component. It should be noted that several time series fail to satisfy the above-mentioned tests for standardised innovations. These are domain 3 with marginally positively autocorrelated standardised innovations; domains 4, 9 and the aggregated series violating normality, of which domain 9 also exhibits heteroscedasticity. However, given the extreme erratic pattern of the series (see Fig. 2.1.1), the obtained model fits can be viewed as satisfactory.

4.1.1 Level and variance breaks in the series modelled in a univariate setting

Series	Level interventions	Sub-periods for which different variances apply
1. Agriculture	–	1976(1)-2002(4); 2003(1)-2010(4)
2. Food	–	1976(1)-2002(4); 2003(1)-2010(4)
3. Oil	2008(3)-(4)	1976(1)-1993(4); 1994(1)-2002(4); 2003(1)-2010(4)
4. Ores	–	1976(1)-1996(4); 1997(1)-2010(4)
5. Metals	–	1976(1)-2002(4); 2003(1)-2006(4); 2007(1)-2010(4)
6. Minerals	–	1976(1)-1991(4); 1992(1)-2002(4); 2003(1)-2010(4)
7. Fertilizers	2003(1)-2010 (4), 2007(1)-2008(4)	1976(1)-2002(4) 2003(1)-2010(4)
8. Chemicals	–	1976(1)-2002(4); 2003(1)-2010(4)
9. Other goods	1997(1)-2002 (4), 2003(1)-(4)	1976(1)-1996(4); 1997(1)-2002(4); 2003(1)-2010(4)
Aggregated series	2003(1)-(4)	1976(1)-1984(4); 1985(1)-1987(4); 1988(1)-1993(4); 1994(1)-2002(4); 2003(1)-2010(4)

Filtered point-estimates of the signal and its components from the univariate setting are nearly identical to those from the multivariate model discussed in Subsection 4.2. Therefore, they are depicted only once and can be found in Fig. 4.4.1, 4.4.2 and 4.4.3 for the most interesting domains that are representative of the others. If a survey redesign is considered to be an improvement, the trend level during the time period before the intervention can be corrected by the value of the level break estimate. Therefore, the signals are presented with the level shifts included. The level breaks, in turn, are presented separately as an indication of the size of a certain discontinuity. The standard errors of the filtered signals are depicted in Fig. 4.4.4, where peaking standard errors reflect an additional uncertainty brought about by the inclusion of a level break. It takes about one year for the standard errors to decay and stabilize around a new level (see domain 9 and the aggregate series in Fig. 4.4.4).

4.2 Multivariate models for the nine domains

While the implementation of univariate structural time series models makes it possible to borrow strength over time, multivariate models also benefit from information available over space. The present multivariate model is based on the level and variance breaks described for the nine univariate models in Table 4.1.1. Model (3.2) combines the nine series of the domains in one multivariate setting with non-zero covariances among the slope disturbance terms. Modelling the slope disturbance covariances significantly increases the maximum value of the likelihood function. While the point estimates remain almost unchanged, the variance of the signals gets considerably reduced. This is an important result because variance reduction, together with structural break modelling, is one of the most important reasons behind considering a model-based approach in this application. Apart from considerably increased accuracy, yet another improvement brought about by modelling the correlation among the slope disturbances is the absence of the serial correlation that was present in the standardised innovations of the third domain's univariate model (compare Tables A.1 and A.2 in the appendix).

Detecting and modelling common trends further reduces the signal variance. The existence of the common factor specification implies that the trends of D domains are driven by a smaller number of underlying stochastic trends that are usually called common factors. This results in more parsimonious models. The common factor detection is based on an examination of eigenvalues of the slope disturbance covariance matrix. Trends whose eigenvalues are significantly different from zero are called common factors. These drive the development of the whole system. Insignificant eigenvalues of the slope disturbance covariance matrix are removed step-wise until the number of common trends is identified.
Modelling the slope covariances in this study has led to insignificant variances of the seasonal stochastic terms. However, three of them (those of domains 6, 8 and 9), still being close to zero, turned out to be one order of magnitude larger than the others. The question at this point is how to proceed with the common factor model: by either first removing insignificant seasonal disturbances and then removing eigenvalues that are close to zero, or the other way around. These two approaches lead to different models. The former approach leads to a model with six common factors, whereas the latter one results in only five, which turned out to be the minimum possible number of common trends required. After removing all insignificant hyperparameters, the seasonal hyperparameters in both models become considerably larger compared to the case where all eigenvalues are estimated. The first approach yields only two significant seasonal hyperparameters (domains 6 and 8) that are excessively large, whereas the second one results in four significant seasonal disturbance

variances (in domains 2, 3, 6 and 8) whose magnitudes are comparable to those in univariate models. Taking into account the large dimension of the model, the Bayesian Information Criterion (BIC) is chosen for model selection (Ch. 7.4, Durbin and Koopman, 2001), as it imposes a larger penalty on nuisance parameters than the Akaike Information Criterion. It turns out that the model with five common trends and 59 hyperparameters to be estimated, is superior to the model obtained from the first approach based on 63 hyperparameters. This is despite the fact that the maximum value of the likelihood function in the former model is slightly lower than in the model with six common trends (2296.72 against 2298.74). The parameter space reduction that leads to a lower BIC-value can be explained by the fact that the eigenvector corresponding to yet another zero-eigenvalue (of domain 4) does not need to be parameterised. In the model with five common factors, the resulting zero-eigenvalues are those of domains 4, 7, 8, and 9. In this case, moving from the six-factor model to the model with five common factors saves six hyperparameters.

It should be noted that the whole data set is used for detecting common factors among the domain trends. The main reason behind this is that the survey has been frequently redesigned since 2003. This, in combination with adding level breaks, might affect the number of common trends before and after a redesign. In fact, a concurrent estimation of the complete nine-dimensional model, i.e. with all the seasonal hyperparameters estimated, suggests that the number of common trends varied up through the first quarter of 2010, whereafter it became stable until the end of the series, with domains 4, 7, 8 and 9 obtaining nearly zero-eigenvalues.

Not only is the model with five common trends selected as the best one according to the BIC, but it also suggests significantly lower signal variances in domains 4 and 8 compared to the model with six common trends. The variance of the other corresponding domains across the two models is almost the same. This model's diagnostics based on standardised innovations, as well as the maximum-likelihood hyperparameter estimates along with their asymmetric confidence intervals, can be found in the appendix. The asymmetry in the confidence intervals arises from the fact that the hyperparameters are estimated on a log-scale. The Fisher information matrix is used to estimate the asymptotic standard errors of the log-transformed hyperparameters, whereafter the confidence interval bounds are transformed back to the original scale. Point-estimates obtained from the nine-dimensional and univariate models are virtually the same. The variance comparison will be illustrated in Subsection 4.4.

4.3 Interpretation of common trends

The previous subsection demonstrates that there are five stochastic trends that drive the development of all the nine domains. The trend equation could be expressed in the following form:

$$L_{t,d} = L_{0,d} + R_{0,d}t + \sum_{j=1}^{t-1}\sum_{i=1}^{j}\eta_{R,i,d},$$

(4.1)

where $L_{0,d}$ is an intercept and $R_{0,d}$ is the slope coefficient of the deterministic part in the trend (Koopman *et al.,* 1999a, Ch. 6.4.4 and 9.1.4.2). The double partial sum of η's is the stochastic part of the trend. The implication of the common factor model is that, for each of these D trends, this partial sum can be expressed as a linear combination of a smaller number of common factors. The system can also be written in a matrix form:

$$L_t = \Theta L_t^{\dagger} + L_0 + R_0 t,$$

(4.2)

where Θ is a $D \times m$ matrix of factor loadings, m being the number of common factors; vector $L_t^{\dagger}$ contains time-specific estimates for the m common stochastic trends. The first m elements of vectors L_0 and R_0 are zeros, and the remaining entries are those described in the preceding equation.

There are different ways to construct common factors and factor loading matrices. One possibility is to take m orthogonal common factors, where the factor loading matrix $\boldsymbol{\Theta}$ is equal to the lower triangular matrix $\boldsymbol{A}$ from the Cholesky decomposition of the covariance matrix $\boldsymbol{Q}.$ Then, the extracted trend of the first domain will be equal to one of these factors, and the other trends, in turn, to a linear combination of the identified common trends. However, in order to see how the remaining $(D-m)$ trends depend on the extracted trends that correspond to zero-eigenvalues, we can apply a factor rotation that turns the upper part of the $\boldsymbol{\Theta}$ matrix into an m-dimensional identity matrix. This approach results in correlated common factors because they are equal to the trends of the first m domains. After the series are slightly reshuffled (in the following order: 1, 2, 3, 5, 6; 4, 7, 8, 9), i.e. all the trends with zero-eigenvalues are kept as the last ones in the system, the factor loadings in the lower $(D-m)\times m$ part of the modified matrix $\boldsymbol{\Theta}^*$ can be obtained from the eigenvector matrix of the covariance matrix $\boldsymbol{Q}$, as shown in   Ch. 6.4.1 of Koopman *et al.* (1999a). It results in the following factor   loading matrix:

$$\boldsymbol{\Theta}^* = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0.31 & -0.08 & 0.35 & -0.74 & 0.03 \\ 0.40 & 0.29 & 1.92 & -0.16 & -0.34 \\ 2.36 & -0.31 & 3.63 & -1.25 & -0.04 \\ 4.08 & -1.86 & 1.10 & -6.46 & 0.89 \end{bmatrix}.$$

While the first five trends, those of domains 1, 2, 3, 5 and 6, are entirely explained by the stochastic components, the zero-eigenvalue trends 4, 7, 8 and 9 do have a non-stochastic remainder in the form of an intercept and a deterministic time trend, as follows from (4.2) (see Koopman *et al.*, 1999, Ch. 9.1.4 for more details about smooth trends with common slopes). The intercept and slope coefficients can be found in Table 4.3.1. These estimates are the same both when the common trends are correlated and orthogonal.

### 4.3.1   Intercept and slope coefficients for the common trend model

Domain	4	7	8	9
L_0	-977.8	-5381.7	-10510.3	-1252.9
R_0	4.8	20.8	50.9	15.6

Fig. 4.3.1 displays the smoothed correlated common factors that are equal to the extracted smoothed trends of domains 1, 2, 3, 5 and 6. This subsection dwells upon smoothed trends (rather than upon filtered or concurrent ones), since the focus of interest is on the relationship between state variables, i.e. the domain trends in this case. In order to establish this relationship, it is best to use all the available information. The trends in this application are smoothed by the fixed-interval smoother (Durbin and Koopman, 2001, Section 4.3.1). The smoothed trends of the domains with zero-eigenvalues can be found below in Fig. 4.3.2. The two figures, when confronted with each other, along with the matrix $\boldsymbol{\Theta}^*$, give an insight as to how much each of the common factors contributes to the trend of the remaining domains (4, 7, 8 and 9). For instance, the trends of the large-scaled domains 8 (chemicals) and 9 (other goods) exhibit a very similar behaviour (Fig. 4.3.2). They also resemble the trend of domain 6 (minerals) in terms of local extrema. Indeed, domain 9 has a large positive factor loading that

corresponds to the common factor of domain 6. Although the value 0.89 is smaller than the factor loading for the common factor of domain 1 (4.08), the common factor of domain 6 drives the development of the trend of domain 9 to a larger extent, since the magnitude of domain 6 (in terms of tons transported) is much larger than that of domain 1 (check Fig. 2.1.1 for the domain magnitude). Unlike in domain 9, the trend of domain 8 has a small negative factor loading on common factor 6. Still, trends 8 and 9 behave in a similar way. This is confirmed by the similarity of their factor loadings on the rest of the common factors. Taking into consideration the large scale of domains 6, 8 and 9 in terms of tons transported, the similarities between them seem quite plausible, since these domains are very likely to reflect overall developments in the economy.

As for the zero-eigenvalue trend of domain 4 (ores), it is correlated to a large degree with the trend of domain 5 (metals), which is confirmed by the largest factor loading for this common trend in an absolute value (-0.74). This gives empirical evidence to the fact that two separate stochastic factors are redundant to explain the variation of these two similar categories of goods. The negative sign in the factor loading could be explained by a substitution effect between these domains.

### 4.3.1 Smoothed common correlated factors, 1000 tons.
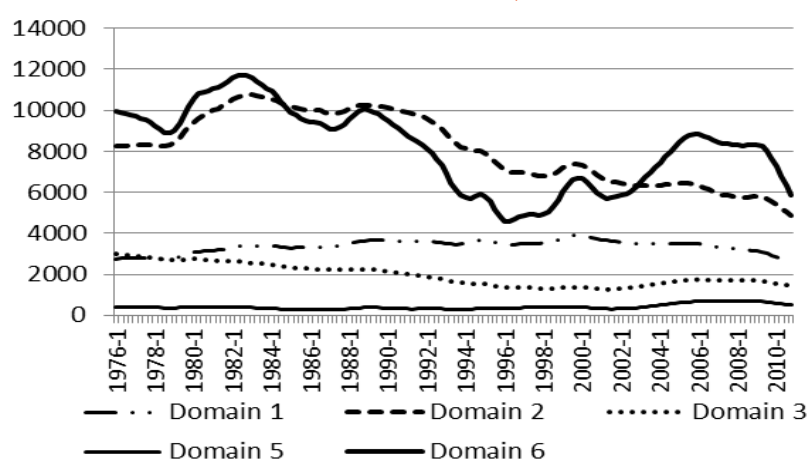
4.3.2 Smoothed zero-eigenvalue trends as a combination of common factors, excluding the intercept and deterministic trend, 1000 tons.
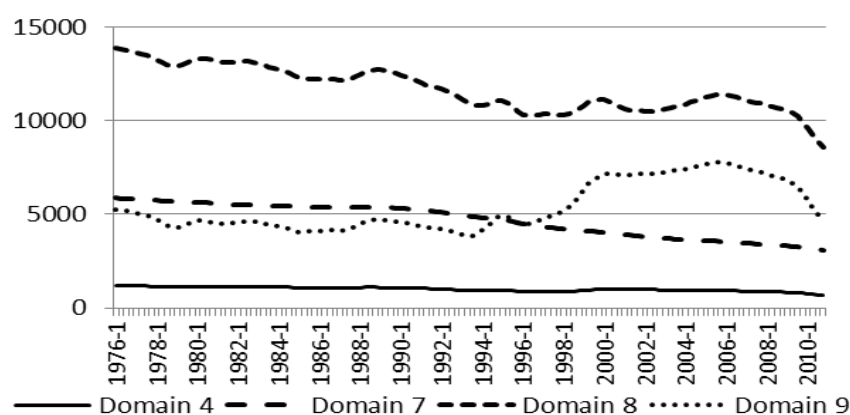
Further, the trends of domains 4 and 9 are strikingly similar, as Fig. 4.3.3. shows. Here, the trend of domain 4 is magnified with a factor of 15, which allows to superimpose it on the graph of trend 9. This similarity is also supported by the factor loadings of these two trends: positive and negative factor loadings correspond to each other, and so do their largest values (e.g., -0.74 and -6.46).

4.3.3 Smoothed trends of domain 9 (black line) and domain 4, scaled by a factor of 15 (grey line)
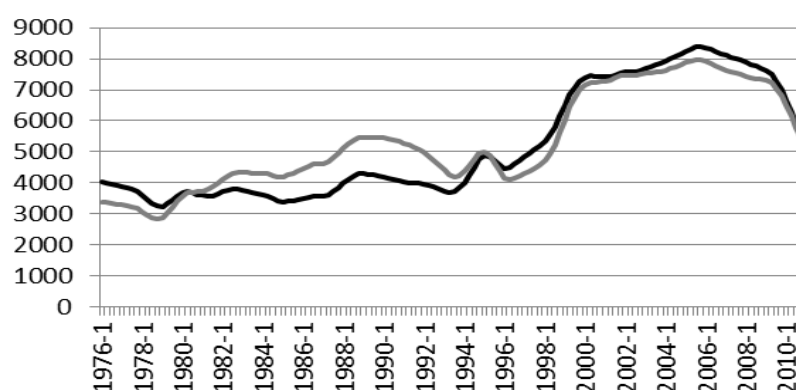
4.4 Estimation results

This subsection compares the estimation results obtained with the univariate and nine-dimensional models. For the aggregated series, estimates based on the univariate model are compared with estimates derived from the nine-dimensional model, as described in Subsection 3.3. The graphs below (Fig. 4.4.1, 4.4.2, 4.4.3) show filtered and concurrent signal estimates of the national level series and the three most interesting domains, as well as their underlying trends and some of the level interventions employed in the model. Point-estimates in the three graphs are obtained from the nine-dimensional model, but they are basically the same for the univariate models. Moving backwards in calculating concurrent estimates, one comes to a point in time when the model needs to be vastly re-specified, i.e. certain level and variance breaks have to be removed, and common factors possibly re-identified. Therefore, concurrent estimates are calculated starting from 2007(1). In fact, the estimates for the periods before 2007(1), depicted in one line together with these concurrent estimates, constitute filtered estimates obtained by means of hyperparameters based on the information available up to and including 2006(4). The whole series is going to be referred to as 'concurrent estimates' further in this paper. The difference between the filtered and concurrent point-estimates of the three domains becomes quite salient as early as in late 90's/early 2000's, especially in domain 5. This might be explained by the frequent changes introduced into the survey during this period. In domain 5, for instance, an increased variation due to the redesign of 2003 is accounted for by the variance break in the measurement equation error term. However, the gradually rising level of point-estimates, caused by this redesign, cannot be remedied by modelling the aforementioned variance break. Allowing for a variance break in the trend is likely to result in a model that better describes the visible pattern of the series. However, this is not inappropriate because such a break would imply a real change in the population. Therefore, with a time-constant trend hyperparameter, the concurrent estimates continue following the same pattern as they did before 2003, until sufficient observations show up in the more volatile post-redesign period. In the last couple of years, the concurrent estimates come back to the path of the filtered ones.

Domain 9 features two level intervention variables illustrated in Fig. 4.4.3. The concurrent estimates of the first level intervention (1997–2002) start deviating from the filtered ones from 1998 and converge to the level of the latter estimates only after 2008. There are visible differences between the concurrent and filtered point-estimates of the signal in the same period of time. The trend estimates are presented in Fig. 4.4.2. The trend of domain 4 closely resembles the one of domain 9 and therefore is not presented here.

4.4.1 Dutch own-account road transportation, 1000 tons: Horvitz-Thompson (dashed line) and the nine-dimensional model-based filtered (black solid line) and concurrent (from 2007(1), grey solid line) estimates of the signal.
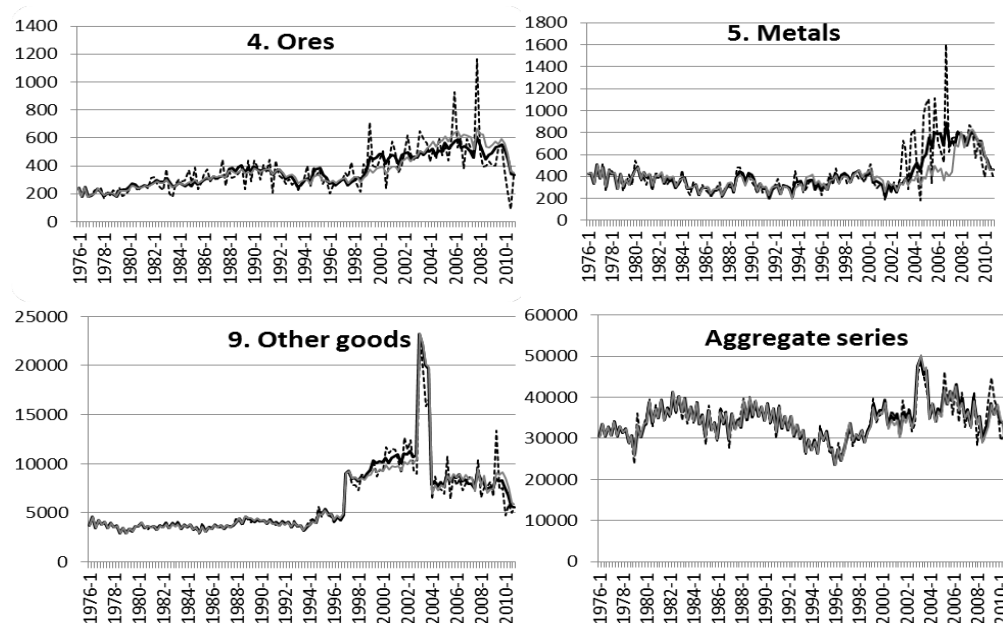

Fig. 4.3.4 illustrates a reduction in the signal variance estimates when one moves from the univariate models to the multivariate one. The standard errors of the filtered signals are depicted in black, and those of the concurrent estimates in grey. As mentioned in subsection 3.2, concurrent estimates give a more realistic picture of what can be obtained with this estimation approach in a production environment, compared with filtered estimates, since under the former approach both the hyperparameter and state variable estimates will be based on the information available in real-time conditions. The outliers modelled for 2007–2008 in domains 3 and 7 are not included in concurrent estimation for a better real-time imitation, since observations at the end of a series are unlikely to be identified as outliers.

Apart from that, the variance break in domain 5 is included only in 2008(3), when the change in the variance becomes sufficiently pronounced. As in the case with the point-estimates, there may be substantial differences between the concurrent and filtered standard error estimates. This is mostly the case for the small domains 4 and 5. Again, by the end of the time period under consideration the two estimates nearly coincide. The variance of the filtered and concurrent signal estimates of the aggregated series coming from the nine-dimensional model for the domains are calculated as the sum of domain variances and the corresponding covariances.

4.4.2 **Dutch own-account road transportation, 1000 tons: the nine-dimensional model-based filtered estimates of the trend (thick black line) and their 95%-confidence interval bands (thin black lines); concurrent estimates and 95%-confidence interval bands from 2007(1) in grey (thick and thin line, respectively).**
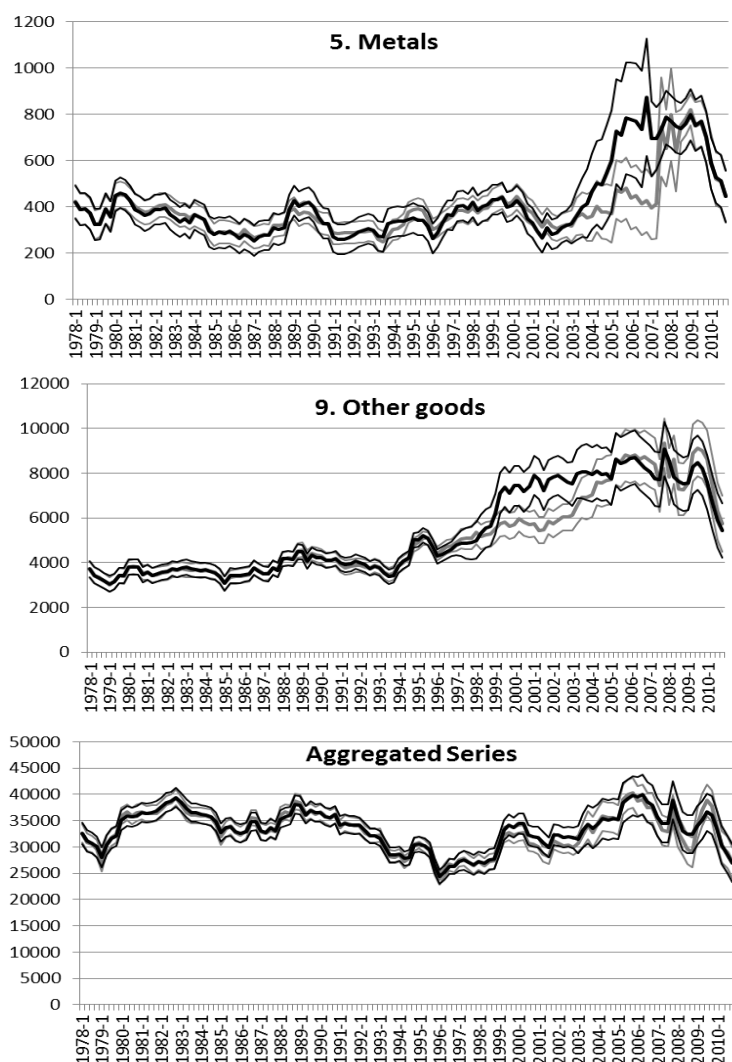
4.4.3 **Filtered (in black) and concurrent (from 2007(1) to the end, in grey) nine-dimensional model-based estimates and their 95%-confidence interval bands for discontinuities in domain 9 of the Dutch own-account road transportation, 1000 tons. Left: level shift 1997(1)-2002(4); right: level shift 2003(1)-(4)**
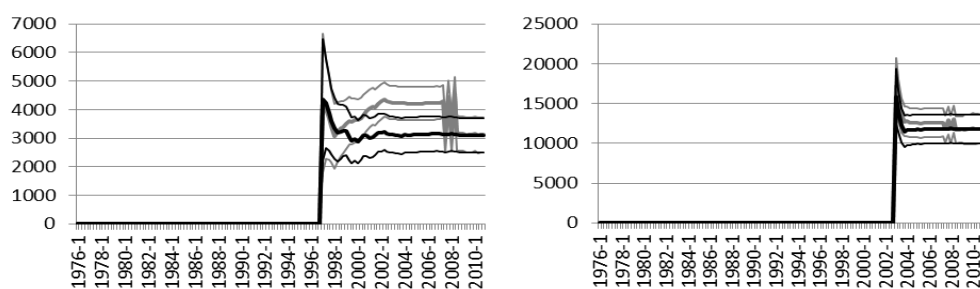
4.4.4 Hyperparameter estimates for the standard errors (1000 tons) of the measurement equation error term from the nine-dimensional model (black dashed line) and the standard errors of the filtered signal estimates obtained from different state-space models: univariate model (black dotted line), nine-dimensional model (thin black line); standard errors of the concurrent esti mates from 2007(1) are depicted in the same style but in grey.
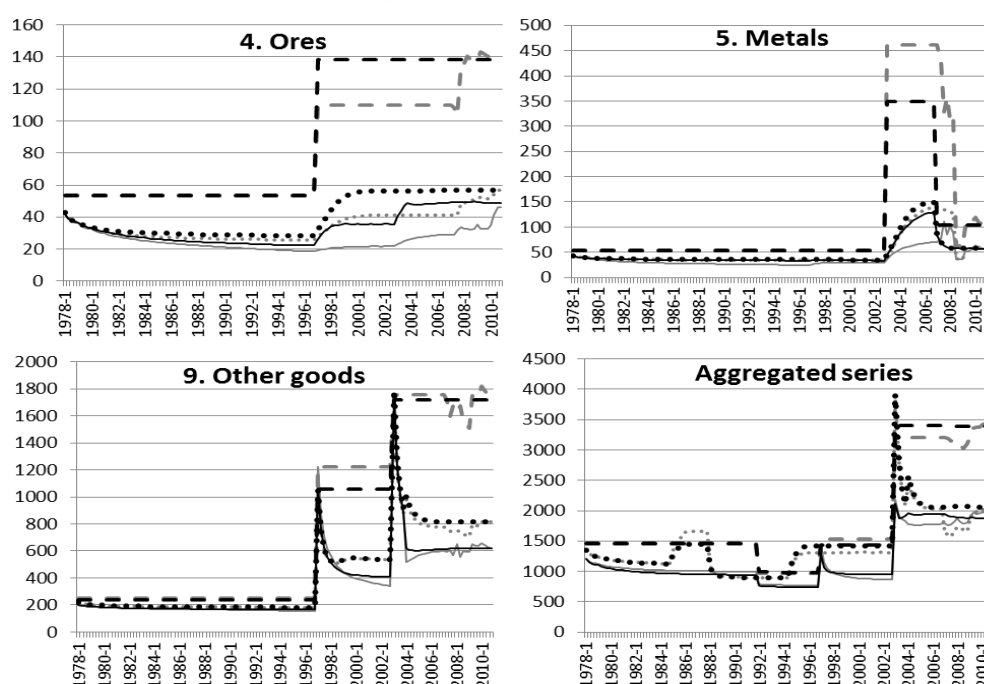
As can be seen, the multivariate nine-dimensional model as well as the estimates for the aggregated series derived therefrom, outperform the univariate models, especially in the last part of the time span. The peaks in the filtered estimate standard errors are caused by intervention variables that model discontinuities in the level of domain 9 and thus also appear in the aggregated series.

The analysis of an increased precision obtained with this modelling approach would be incomplete without the direct estimate variances. As was mentioned in Subsection 3.1, these variance estimates are not available for this survey. The micro-data are available for the last few years, but in this period the complexity of the applied sampling design obstructs a straightforward approximation of the HT estimate variances. Vehicles for an annual sample are drawn according to a stratified sampling design. At the next stage, vehicles belonging to the same owner are clustered and randomly assigned to one of the four quarters. This indirectly implies a two-stage sampling design. An additional complication is that vehicles belonging to the same owner were initially drawn from different strata. This led to a situation where certain companies occurred simultaneously in more than one strata.

Design variances have been approximated, first of all, by assuming a two-stage stratified sampling design, with the companies as the primary sampling units (PSU) and the vehicles as the secondary sampling units, and, secondly, by collapsing the strata in a way that the new stratification scheme is based only on the economy branches the PSUs belong to, without differentiating between the vehicles' characteristics. The design variance has been eventually approximated only for 2008(1) using a variance estimator that assumes a stratified sampling of PSU's with replacement. This estimator is commonly used for complex two- or multiple-stage sample designs, see Särndal *et al*. (1992), Ch. 4.6 (equation (4.6.2)). Since the PSUs are assigned randomly to a certain week in a quarter, the variation over time is still ignored under

this approach. This problem caused the variances of the HT estimates to be underestimated. This might be one of the reasons why some of the design-estimate standard errors in Table 4.4.1 do not exceed the model-based estimates to the expected degree. The design variance estimate of domain 2 in 2008(1) has turned out to be even smaller than that of the uni- and multidimensional model. This could be explained not only by the above-mentioned problems, but also by the fact that the variance estimates themselves are subject to uncertainty. Moreover, the measurement equation error term variance $(\sigma_{v,t,d}^2)$ comprises the average value of the sampling variances over each specific period where this hyperparameter $(\sigma_{v,t,d}^2)$ is assumed to be constant. So obtaining such a low value of the design variance estimate is quite plausible (see, for instance, van den Brakel and Krieg (2009) where the variability of the general regression estimator over time is demonstrated). However, if it is assumed that the measurement equation error term is dominated by the sampling error, then the model variance estimates of this error term could be used as a proxy for the variance estimates of the HT estimator. This design variance approximation approach is advocated by an empirical finding of Krieg and Van den Brakel (2012), where the standard deviation of the measurement equation error term is defined as the product of a hyperparameter and of the design standard error of the input series. The obtained maximum likelihood estimates of this hyperparameter are nearly equal to one for all the domains considered in their work. Fig. 4.4.4 shows that the model-based approach offers a considerable variance reduction as compared to the HT-estimator. This reduction is particularly salient in the most recent period when effective sample sizes get smaller. Not only do small-scale domains, such as domains 4 and 5 of this application, benefit from this modelling approach by experiencing up to 70 percent reduction in their standard errors, but also large domains like domain 9, as well as the national level series, become about twice as

4.4.1 Design standard errors of the Horvitz-Thompson estimates, standard errors of the concurrent model-based signal estimates, and standard deviation maximum-likelihood estimates of the measurement equation composite error term from the nine-dimensional model; 1000 tons, 2008(1).

Domains	Univariate model	Nine-dimensional model	Horvitz-Thompson estimator	SD of the composite error term
1. Agriculture	156	162	529	405
2. Food	648	520	481	1 382
3. Oil	158	125	451	447
4. Ores	57	49	163	136
5. Metals	58	58	259	311[2]
6. Minerals	938	875	1 063	1 388
7. Fertilizers[1]	204	196	226	681
8. Chemicals	765	488	1 105	1 586
9. Other goods	816	618	723	1 625
Aggregated series	2 065	1 892	2 361	3 152

[1] Concurrent estimates of this domain in 2008(1) did not include the outliers modelled for 2007(1)–2008(4).
[2] In the course of a concurrent estimation of this domain, the variance break due in 2008(1) was not included, as the change in the variance had not yet been sufficiently pronounced.

precise in terms of standard errors in the latter part of the time period (after 1997). The filtered and concurrent estimates that have been considered so far, illustrate what can be obtained with this model-based approach when the sample information in period t becomes available. An issue with this estimation procedure in production is that these estimates can be improved if new information becomes available after period t. Depending on the size of the adjustments, it might be necessary for an NSI to consider a revision strategy. Besides the

variance of the filtered signal, it is interesting to analyse the variance of revisions $\left(\hat{l}_{t|t+k}^{d} - \hat{l}_{t|t}^{d}\right)$, where the smoothed estimate $\hat{l}_{t|t+k}^{d}$ denotes a revised signal estimate of domain d for period t using the information available at time $(t+k)$, whereas $\hat{l}_{t|t}^{d}$ denotes a filtered estimate at time t. Revisions may be quite significant as it is usually difficult to produce a good estimate at the end of the series. This issue is discussed in Orphanides and Van Norden (2002) or Planas *et al.* (2013), and is closely related to the problem of revision strategy for seasonally adjusted figures. Many NSIs continuously revise their official releases of seasonally adjusted series in order to improve the seasonal effect estimates, as new information becomes available. For a revision strategy, it is important to choose the best revision horizon. Large revisions indicate that a certain revision strategy might be required. When revisions are small, it may be more convenient to leave the initially published figures unchanged.

To illustrate the size of revisions in the DRTS, the revised signal estimates at four different horizons (one-, two-, four-, and eight-quarters) are plotted together with 95-percent confidence intervals of the filtered estimates in Fig. 4.4.5. These revised signal estimates are calculated starting from 1988(3), namely, for the last 89, 88, 86 and 82 quarters of the sample for one-, two-, four-, and eight-quarter revisions, respectively. The Kalman filter is run conditionally on the hyperparameter set estimated on the basis of the complete sample. The revisions at all the above-mentioned horizons remain within the confidence interval bands of the filtered estimates.

4.4.5. The filtered estimates' confidence interval bands (dotted lines) and revised signal estimates in between: after 1 quarter (black solid line), 4 quarters (black dashed line), and 8 quarters (grey solid line)
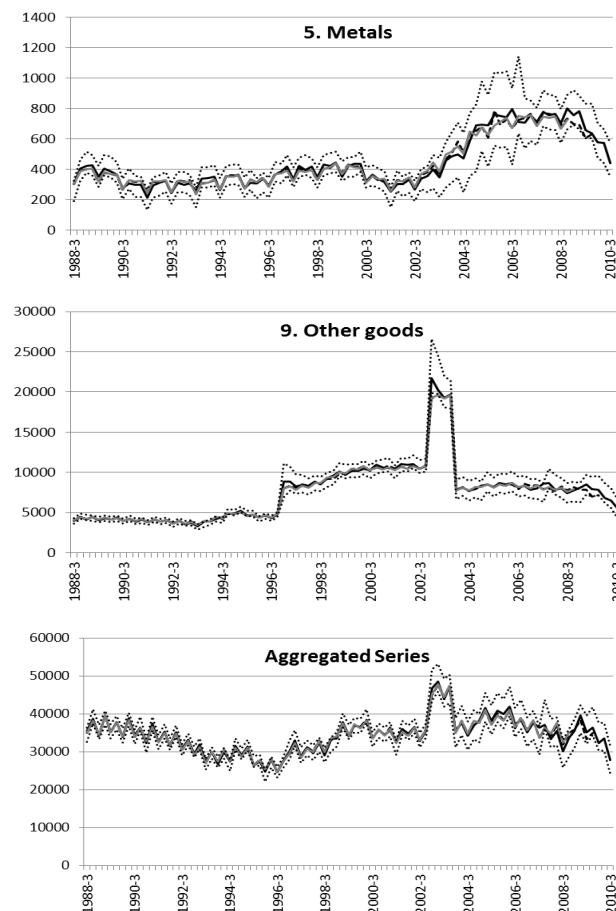
Table 4.4.2 presents the sample mean of absolute revisions (*AR*) in absolute values:

$$MAR_k = \frac{100\%}{89-k} \sum_{t=51}^{140-k} \sqrt{|\hat{l}_{t|t+k}^d - \hat{l}_{t|t}^d|}$$. Sample means of relative revisions (*RR*) in

absolute values are defined as: $MRR_k = \frac{100\%}{89-k} \sum_{t=51}^{140-k} \sqrt{|\hat{l}_{t|t+k}^d - \hat{l}_{t|t}^d|/\hat{l}_{t|t}^d}$ and

presented in Table 4.4.3.

4.4.2. Sample mean and standard deviation of the signals' absolute revisions after *k* quarters and the average standard error of the filtered signals based on 1988(3)–2010(4), 1000 ton.

	MAR_k, 1000 tons				$SD(AR_k)$, 1000 tons				$\frac{1}{89}\sum_{t=51}^{140}\sqrt{\widehat{MSE}(\hat{l}_{t|t}^d)}$
	k=1	k=2	k=4	k=8	k=1	k=2	k=4	k=8	
1. Agriculture	46	60	72	79	63	81	89	95	124
2. Food	122	151	162	188	158	199	209	257	327
3. Oil	26	35	53	60	34	45	67	72	100
4. Ores	15	20	20	18	21	27	28	22	33
5. Metals	23	26	31	30	34	37	41	39	50
6. Minerals	197	244	299	362	266	333	406	491	601
7. Fertilizers	43	60	79	84	75	89	120	150	130
8. Chemicals	123	161	171	163	176	220	238	207	335
9. Other goods	219	313	334	297	360	552	615	591	437
Aggregated series	630	831	857	855	862	1081	1121	1101	1255

4.4.3. Relative revisions of the signals after k quarters in percentage: sample mean and standard deviation over time.

	MRR_k, %				$SD(RR_k)$, %			
	k=1	k=2	k=4	k=8	k=1	k=2	k=4	k=8
1. Agriculture	1.37	1.78	2.09	2.27	1.92	2.47	2.68	2.77
2. Food	1.83	2.31	2.43	2.85	2.51	3.31	3.49	4.50
3. Oil	1.81	2.38	3.45	3.91	2.81	3.48	4.61	4.97
4. Ores	3.61	4.58	4.87	4.54	4.84	5.98	6.19	5.54
5. Metals	4.94	5.93	7.07	7.23	6.48	7.45	8.62	8.83
6. Minerals	2.88	3.53	4.24	5.14	3.90	4.88	5.62	6.82
7. Fertilizers	4.02	5.43	7.22	7.95	6.65	7.62	10.26	13.01
8. Chemicals	2.25	2.89	3.09	3.13	3.06	3.69	3.98	3.86
9. Other goods	2.69	3.62	3.79	3.34	3.73	4.80	5.12	4.59
Aggregated series	0.02	0.02	0.02	0.02	0.02	0.03	0.03	0.03

Based on the sample, the MRR mostly remain under 5 percent for all the revision horizons. The small domains 5 and 7 are exceptions with revisions occasionally exceeding 7 percent depending on the revision horizon. For the choice of the most appropriate revision horizon, it is important to note that MAR and MRR clearly increase with the revision horizon, but at a decreasing rate. Namely, the increment in the mean revisions of the domains is the highest when one moves from no revision to one- and two-quarter horizons. After the second quarter, little is changed by subsequent revisions. This suggests that two-quarter revisions are worth considering. As for the aggregated series, the MRR is as small as 0.02 percent for all the revision horizons, indicating that the initially estimated aggregated series is quite reliable. However, if the estimates are subject to revision, so will the aggregated series estimates be. Assuming that the difference between the filtered signal estimates $\hat{l}_{t|t}^d$ and smoothed estimates $\hat{l}_{t|t+k}^d$ is stationary and independently distributed, the ARs' sample standard deviations can act as a proxy the volatility measure of absolute revisions. These sample

estimates are also given in Table 4.4.2. The ARs' sample standard deviations can also be compared to filtered signal standard errors.

Although the uncertainty of the filtered signals has been found to be different in several sub-periods, an average of these standard errors (over 1988(3)-2010(4)) is presented in the last column of Table 4.4.2 for indicative purposes. As can be seen, the standard deviation of absolute revisions never exceeds the standard error of the filtered signal estimates, with an exception of domain 9.

The sample standard deviations of the RRs are presented as well in the last part of Table 4.4.3 to provide an indication of how far the revisions can reach in relative terms. The standard deviations of the RRs and ARs show that the eight-quarter revisions are more volatile in all the domains, compared to revisions at the other three horizons, just as expected. As for cross-sectional differences, the signal estimates of domain 4, 5 and 7 are the least stable, with the RRs' standard deviations reaching sometimes 7 percent, whereas the aggregated series filtered estimates do not seem to be much affected by revisions. The three above-mentioned domains are the smallest ones and feature a highly volatile pattern of design estimates.

5. Discussion

This paper presents an application of univariate and multivariate structural time series models to the domestic own-account segment of the Dutch Road Transportation Survey (DRTS). Two problems are solved simultaneously with this approach. The first problem is frequent survey redesigns that have led to several level shifts in the direct estimates of this survey. So-called discontinuities hamper the comparability of the published figures over time. Secondly, several survey modifications have reduced the effective sample size, which resulted in variance breaks in the published figures. In addition to that, the gradually increasing variances made series too imprecise and excessively volatile.

The DRTS, being a cross-sectional survey and at the same time featuring these problems, can be improved by developing a multivariate time series model that accounts for level and variance breaks and improves the series precision by borrowing strength over time and space. The role of information accumulated in the past is especially important if sample sizes have been shrinking in the course of time. Apart from that, the survey can benefit from a multivariate setting that makes it possible to borrow information over domain space. In this way, domains can be estimated more precisely by linking them to related domains. Our findings suggest that a multivariate model for the domains outperforms the univariate setting. Not only is the multivariate model able to estimate the survey discontinuities, but it also features significantly lower signal variances for both the domains and aggregated series, compared to the design-based approach. Namely, the design estimator standard errors can be reduced by 40 to 60 percent in large domains and/or in the aggregated series in the period following the most recent major survey redesign. When it comes to small domains, the variance reduction may be three- or even four-fold.

Estimates at the national level can be dealt with in several ways. The aggregated level design estimates can be used as benchmarks for domain model estimates according to Pfeffermann and Tiller (2006), if the former estimates are sufficiently precise. In the DRTS, however, it is

not the case. A decent quality could be reached if the aggregated series estimates are derived from a multivariate model developed for domains. In the case of the DRTS, this approach resulted in a lower signal variance of the aggregated series compared to that obtained when the univariate model was applied to the aggregated series itself. This gives evidence to the fact that the multivariate model for the domains is well specified, and that accounting for survey modifications at a lower aggregation (domain) level produces better outcomes for the aggregated series.

The analysis of revisions shows that concurrent estimates can be improved with the information that becomes available in the next two quarters. The model estimates are, however, not largely affected by revisions. In extreme cases of small volatile domains with level breaks, the standard deviation of relative revisions may exceed 10 percent, whereas that of the aggregated series remains as low as 0.03 percent even at the eight-quarter revision horizon. In the present case, a two-quarter horizon may be considered, since little is corrected at longer revision horizons.

An additional advantage of the structural time series modelling approach is that it offers a breakdown of the signal into the trend, seasonal and intervention components. Seasonally adjusted series are therefore obtained as a by-product. The technique presented here can be applied to any small area estimation problem, where the survey features a subdivision into several domains. Furthermore, if series under consideration suffer from small sample sizes and have an erratic pattern similar to that of the DRTS series, switching from a traditional design-based approach to the state-space modelling technique in the production of official statistics would most likely pay off the efforts exerted. Further improvements, specific to this application, can be derived with the help of information from other segments covered by this survey: hire-and-reward and international transportation.

The model-based approach will be even more attractive if the model is augmented with design variances. But if the design variance estimate are not available, average sampling variance values have to be estimated as hyperparameters for several time-periods, and the method described in this paper will be limited in its application to the data from the past. This is due to the fact that, once a break in the variance occurs, a certain number of observations are needed to estimate another hyperparameter for the new sub-period, while the figures have to be produced and published on a continuous basis. When a sufficient number of observations become available after the break, the model has to be adjusted, which might require a revision of the published figures. This can be avoided if design variances are produced along with the point-estimates.

References

Binder, D.A. and Dick, J.P. (1990). A method for the analysis of seasonal ARIMA models. *Survey Methodology*, 16, 239–253.

Doornik, J.A. (2007). *An Object-Oriented Matrix Programming Language Ox 5.* Timberlake Consultants Press, London.

Doornik, J. A., and Hansen, H. (2008). An Omnibus Test for Univariate and Multivariate Normality. *Oxford Bulletin of Economics and Statistics*, 70, 927–939.

Durbin, J. and Koopman, S. J. (2001). *Time Series Analysis by State-space Methods*. Oxford University Press, Oxford. MR1856951

Durbin, J., and Quenneville, B. (1997). Benchmarking by state-space models. *International Statistical Review*, 65, 23–48.

Hannan, E. J., Terrell, R. D. and Tuckwell, N. (1970). The seasonal adjustment of economic time series. *International Economic Review*, 11, 24–52.

Harvey, A. C. (1989). Forecasting, Structural Time Series Models and the Kalman Filter. Cambridge Univ. Press, Cambridge.

Harvey, A. C. (2000). Trend Analysis. University of Cambridge, Faculty of Economics and Politics, Manuscript.

Harvey, A. C., and Chung, C.H. (2000). Estimating the underlying change in unemployment in the UK. *Journal of the Royal Statistical Society, Series A*, 163, 303–339.

Harvey, A.C., and Durbin, J. (1986). The effects of seat belt legislation on British road casualties: a case study in structural time series modelling. *Journal of the Royal Statistical Society, Series A*, 149, 187–227.

Harvey, A. C., and Koopman, S. J. (1992). Diagnostic checking of unobserved-components time series models. *Journal of Business & Economic Statistics*, 10, 377–389.

Horvitz, D.G. and Thompson, D.L. (1952). A generalisation of sampling without replacement from finite-universe. *Journal of the American Statistical Association*, 47, 663–685.

Koopman, S.J. (1997). Exact Initial Kalman Filtering and Smoothing for Nonstationary Time Series Models. *Journal of the American Statistical Association*, 92, 1630–1638.

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

Koopman, S.J., Shephard, N. and Doornik, J.A. (1999b). Statistical algorithms for models in sate space form using SsfPack 2.2. *Econometrics Journal*, 2, 113–66.

Koopman, S.J., Shephard, N. and Doornik, J.A. (2008). *SsfPack 3.0: Statistical Algorithms for Models in State-space Form*. London Timberlake Consultants Press, London.

Krieg, S. and Van den Brakel, J.A. (2012). Estimation of the Monthly Unemployment rate for Six Domains through Structural Time Series Modelling with Cointegrated Trends. *Computational Statistics and Data Analysis,* 56, 2918–2933.

Narain, R. (1951). On sampling without replacement with varying probabilities. *Journal of the Indian Society of Agricultural Statistics*, 3, 169–174.

Nyblom, J. and Harvey, A.C. (2001). Testing against smooth stochastic trends. *Journal of Applied Econometrics*, 16, 415–29.

Orphanides, A., and Van Norden, S. (2002). The unreliability of output-gap estimates in real time. *Review of Economics and Statistics*, 84, 569–583.

Pfeffermann, D. and Bleuer, S.R. (1993). Robust Joint Modelling of Canadian Labor Force Series of Small Areas. *Survey Methodology*, 19, 149–164.

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

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

Planas, C., Roeger, W. and Rossi, A. (2013). The information content of capacity utilization for detrending total factor productivity. *Journal of Economic Dynamics and Control*, 37, 577–590.

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

Tiller, R.B. (1992). Time series modelling of sample survey data from the US current population survey. *Journal of Official Statistics*, 8, 149–166.

Van den Brakel, J.A. and Krieg, S. (2009). Estimation of the monthly unemployment rate through structural time series modelling in a rotating panel design. *Survey Methodology*, 35, 177–190.

Van den Brakel, J.A. 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.

List of Acronyms

AR – absolute revisions

DRTS – Dutch Road Transportation Survey

HR – Hire-and-Reward

HT – Horvitz-Thompson

MAR – mean absolute revisions

MRR – mean relative revisions

OA – Own-Account

NSI – national statistical institute

RR – relative revisions

SAE – small area estimation

A. Maximum-likelihood estimates of the hyperparameters and model diagnostics

A.1. Univariate models: hyperparameter estimates (1000 tons) and model diagnostics based on standardised innovations

Series	1	2	3	4	5	6	7	8	9	Total
$\sigma_{R,d}$	3 (1; 11)	41 (26; 64)	6 (2; 15)	2 (1; 4)	7 (4; 11)	83 (39; 174)	4 (2; 7)	26 (13; 50)	48 (25; 92)	201 (121; 332)
$\sigma_{\omega,d}$	-	15 (4; 59)	-	-	-	66 (33; 133)	-	26 (9; 71)	-	-
$\sigma_{\nu,t,d}$ *)	281 (243; 325); 436 (339; 562)	406 (347; 475); 1370 (1060; 1770)	244 (203; 294); 164 (126; 213); 387 (297; 504)	54 (46; 64); 140 (115; 170)	54 (47; 63); 357 (247; 514); 102 (68.9; 151)	1230 (989; 1530); 720 (536; 966); 1650 (1240; 2180)	155 (135; 178); 541 (417; 701)	553 (472; 647); 1880 (1450; 2430)	267 (219; 325); 1050 (774; 1430); 1790 (1370; 2330)	1st: 1690 (1290; 2220); 2nd and 4th: 2410 (1920; 3040); 3rd: 1310 (944; 1820); 5th: 3910 (2990; 5110)
N	0.970 [0.616]	0.634 [0.728]	0.596 [0.742]	9.974 [0.007] †	2.057 [0.358]	2.578 [0.276]	1.048 [0.592]	3.229 [0.199]	47.399 [0.000] †	8.292 [0.016]*
Skewness	-0.202	-0.161	0.151	0.451	0.221	0.303	0.203	0.153	2.076	0.614
Excess kurtosis	-0.114	-0.170	-0.036	1.379	0.323	-0.193	0.028	0.542	10.743	0.503
$\sigma_{St.in.}$	0.982	1.009	1.014	1.010	0.997	1.015	1.016	1.016	1.273	1.047
DW	1.731	1.940	1.639*	2.166	2.024	1.706	1.868	1.846	2.098	1.725
H	1.287	0.932	1.220	1.656	1.005	0.908	1.323	1.276	3.867†	0.987

*Significant at the 5% level. †Significant at the 1% level.

Asymmetric confidence intervals in ( ). The asymmetry arises from the fact that the hyperparameters are estimated on a log scale. The Fisher information matrix is applied for estimating the asymptotic standard errors of the log-transformed hyperparameters, whereafter the confidence interval bounds are transformed back to the normal scale.

N: Doornik-Hansen normality test (chi-square test); p-values in [ ];

$\sigma_{St.in.}$: standard deviation of standardized innovations;

DW: Durbin-Watson test statistic for serial correlation;

H: F-test for heteroscedasticity

*) Multiple rows in this field stand for the separate variance values in different sub-periods according to Table 4.1. The dates defining the sub-periods can be found in the same table.

A.2. Nine-dimensional model: hyperparameter estimates (1000 tons) and model diagnostics based on standardised innovations

Series	1	2	3	4	5	6	7	8	9
Zero-restrictions on eigenvalues	-	-	-	Yes	-	-	Yes	Yes	Yes
$\sigma_{R,d}$	24 (12; 48)	75 (17; 247)	11 (3; 53)	8 (2; 27)	10 (3; 41)	152 (20; 549)	8 (5; 50)	75 (13; 273)	89 (23; 309)
$\sigma_{\omega,d}$		17 (6; 54)	4 (0; 125)			70 (36; 135)		26 (10; 70)	
$\sigma_{v,t,d}$ *)	256 (221; 296); 430 (331; 558)	387 (330; 455); 1350 (1044; 1740)	236 (196; 285); 161 (124; 208); 382 (295; 493)	54 (46; 63); 138 (114; 168)	54 (46; 62); 348 (243; 500); 103 (70; 153)	1220 (998; 1500); 582 (435; 779); 1660 (1240; 2210)	152 (131; 176); 537 (405; 711);	538 (462; 627); 1830 (1420; 2360)	239 (199; 288); 1060 (784; 1430); 1720 (1330; 2230)
N	1.121 [0.571]	0.512 [0.774]	1.210 [0.546]	18.439 [0.000] †	0.660 [0.719]	3.162 [0.206]	1.461 [0.482]	1.685 [0.431]	58.133 [0.000] †
Skewness	-0.105	-0.060	0.223	0.736	0.117	0.338	0.248	0.231	2.379
Excess kurtosis	0.205	0.076	0.019	2.569	0.072	-0.161	-0.079	0.207	12.530
$\sigma_{St.in.}$	0.968	1.055	1.012	1.049	1.000	1.039	1.019	1.046	1.336
DW	2.078	1.952	1.746	2.129	2.051	1.798	1.932	1.955	2.242
H	1.289	0.997	1.262	1.161	1.091	0.903	1.258	1.268	3.481†

*Significant at the 5% level. †Significant at the 1% level.

Asymmetric confidence intervals in ( ). The asymmetry arises from the fact that the hyperparameters are estimated on a log scale. The Fisher information matrix is applied for estimating the asymptotic standard errors of the log-transformed hyperparameters, whereafter the confidence interval bounds are transformed back to the normal scale.

N: Doornik-Hansen normality test (chi-square test); p-values in [ ];

$\sigma_{St.in.}$: standard deviation of standardized innovations;

DW: Durbin-Watson test statistic for serial correlation;

H: F-test for heteroscedasticity

*) Multiple rows in this field stand for the separate variance values in different sub-periods according to Table 4.1. The dates defining the sub-periods can be found in the same table.

A.3. Ten-dimensional model including the aggregated series: hyperparameter estimates (1000 tons) and model diagnostics based on standardised innovations

Series	1	2	3	4	5	6	7	8	9	Total***)
Zero-restrictions on eigenvalues	-	-	-	Yes	-	-	Yes	Yes	Yes	-
$\sigma_{R,d}$	47 (23; 94)	168 (43; 507)	33 (3; 125)	15 (3; 49)	15 (3; 55)	403 (79; 1330)	22 (4; 117)	153 (35; 501)	155 (37; 502)	989 (856; 3212)
$\sigma_{\omega,d}$		21 (7; 61)				397 (306; 514)		34 (9; 124)		399 (306; 532)
$\sigma_{v,t,d}$ **)	258 (224; 298); 406 (312; 528)	373 (317; 439); 1340 (1050; 1710)	232 (196; 275); 157 (121; 202); 389 (300; 505)	56 (47; 65); 140 (114; 172)	66 (57; 77); 328 (231; 465); 99 (66; 148)	276 (65; 1170); 5 (0; 6.3·10^{18}*)); 1130 (740; 1720)	148 (128; 171); 502 (387; 651)	480 (403; 571); 1710 (1350; 2180)	224 (185; 271); 1350 (1020; 1780); 1510 (1160; 1980)	804 (642; 1471); 755 (638; 6.31·10^{18}*)); 735 (619; 6.31·10^{18}*)); 1525 (1183; 6.31·10^{18}*)); 2999 (2282; 3970); 2982 (2271; 3945)
N	0.494 [0.781]	1.333 [0.514]	1.874 [0.392]	16.242 [0.000]†	1.852 [0.396]	2.372 [0.306]	1.237 [0.539]	5.927 [0.052]	79.592 [0.000]†	12.351 [0.002]†
Skewness	-0.012	-0.156	0.284	0.851	0.051	0.308	0.188	0.454	3,026	0.787
Excess kurtosis	0.086	0.217	0.039	2.421	0.355	0.225	0.150	0.827	19.756	1.123
$\sigma_{St.in.}$	0.991	1.102	1.043	1.041	0.986	1.058	1.038	1.099	1.474	1.230
DW	2.171	2.113	1.818	2.158	1.714	2.180	2.048	2.267	2.277	2.401*
H	1.126	0.850	1.044	0.939	1.552	0.554	1.314	1.170	3.516†	0.853

*Significant at the 5% level. †Significant at the 1% level.

Asymmetric confidence intervals in brackets. The asymmetry arises from the fact that the hyperparameters are estimated on a log scale. The Fisher information matrix is applied for estimating the asymptotic standard errors of the log-transformed hyperparameters, whereafter the confidence interval bounds are transformed back to the normal scale.

N: Doornik-Hansen normality test (chi-square test); p-values in [ ];

$\sigma_{St.in.}$: standard deviation of standardized innovations;

DW: Durbin-Watson test statistic for serial correlation;

H: F-test for heteroscedasticity

*) The large upper margins are due to a failure to invert the Hessian when the corresponding hyperparameter is estimated close to zero.

**) Multiple rows in this field stand for the separate variance values in different sub-periods according to Table 4.1 in the main file. The dates defining the sub-periods can be found in the same table. There are six different variance values for the national level series due to an overlap between sub-periods that stand for distinctive variances of different domains.

***) Hyperparameters of the aggregated series and their confidence intervals are not estimated but derived from domain hyperparameters.

A.4. Ten-dimensional model including the aggregated series, where the Kalman filter uses the hyperparameter estimates from the nine-dimensional model: model diagnostics based on standardised innovations

Series	1	2	3	4	5	6	7	8	9	Total
N	1.697 [0.428]	0.480 [0.787]	1.591 [0.451]	16.202 [0.000]†	0.731 [0.694]	3.566 [0.168]	1.846 [0.397]	2.830 [0.243]	47.093 [0.000]†	13.044 [0.002]†
Skewness	-0.074	-0.095	0.261	0.690	0.118	0.379	0.279	0.351	2.064	0.791
Excess kurtosis	0.326	0.036	0.005	2.265	0.095	-0.029	-0.065	0.073	10.630	0.973
$\sigma_{St.in.}$	0.973	1.095	1.025	1.084	1.018	1.103	1.042	1.096	1.404	1.358
DW	2.115	1.972	1.769	2.049	2.044	1.948	1.832	1.917	2.186	2.109
H	1.254	0.951	1.203	1.142	1.094	0.896	1.474	1.227	3.136†	1.026

*Significant at the 5% level. †Significant at the 1% level.

N: Doornik-Hansen normality test (chi-square test); p-values in [ ];

$\sigma_{St.in.}$: standard deviation of standardized innovations;

DW: Durbin-Watson test statistic for serial correlation;

H: F-test for heteroscedasticity

B. Simulation study for a (D+1)-dimensional model

1. Introduction

This appendix contains some issues not covered in the body of the main article. The major issue is whether or not an aggregated series of interest should be modelled jointly with underlying D domains in one multivariate setting. It is mentioned in the article that such a model benchmarks domain model estimates to the aggregated series model estimates. However, since no additional information enters the model, a D- and $(D+1)$-dimensional settings are expected to have identical outcomes. A small simulation study shows that it is so when the Kalman filter operates under true hyperparameter values. As soon as they are replaced by their maximum-likelihood estimates, differences emerge both between point- and variance estimates produced by the two models.

The state-space representation of the $(D+1)$-dimensional model is given in Subsection 2.1. The simulation results are presented in the subsequent subsection. Outcomes of the two models applied to the Dutch Road Transportation Survey (DRTS) are compared in Subsection 2.2.

2. Multivariate model for D domains extended with an aggregated series

2.1 State-space representation

This subsection presents a $(D+1)$-dimensional model applicable to the DRTS series. The components are described in the main body of the article in Subsection 3.1. The input vector in this case is extended to $\widehat{Y}_t = \left(\widehat{Y}_{t,1} \dots \widehat{Y}_{t,D}\ \widehat{Y}_{t,\text{Tot}}\right)'$, where $\widehat{Y}_{t,\text{Tot}} = \sum_{d=1}^{D} \widehat{Y}_{t,d}$ are the aggregated series design estimates. The trend and seasonal components of the state vector are extended similarly to the $\widehat{Y}_t$-vector, whereas the vector with level interventions remains unchanged. Then the state vector $\boldsymbol{\alpha}_t$ has the following structure:

$$\boldsymbol{\alpha}_t = \left(\boldsymbol{\alpha}_t^L\ \boldsymbol{\alpha}_t^\gamma\ \boldsymbol{\alpha}^\beta\right)',$$

$$\boldsymbol{\alpha}_t^L = \left(L_{t,1}\ R_{t,1}\ \dots\ L_{t,D}\ R_{t,D}\ L_{t,\text{Tot}}\ R_{t,\text{Tot}}\right),$$

$$\boldsymbol{\alpha}_t^\gamma = \left(\mathbf{1}'_{[D+1]} \otimes \boldsymbol{\gamma}_{t,d}\right), \text{ where}$$

$$\gamma_{t,d} = \left(\gamma_{t,1,d} \; \gamma_{t,1,d}^* \; \cdots \gamma_{t,(s/2-1),d} \; \gamma_{t,(s/2-1),d}^* \; \gamma_{t,s/2,d} \right), \; d \in \{1, \ldots, D+1\},$$

$$\alpha^\beta = (\beta_1 \ldots \beta_K).$$

$\mathbf{1}'_{[p]}$ is a p-dimensional horizontal vector of ones, $\otimes$ is the Kronecker product which multiplies every element of the first matrix by the second matrix, and s is the number of seasons. The inclusion of an aggregated series which is exactly equal to the sum over the domain estimates implies that each state variable of the aggregated series is restricted to the sum of the corresponding state variables of all the domains. These restrictions are incorporated in the design matrix T of the transition equation $\alpha_{t+1} = T\alpha_t + \eta_{t+1}$. . The transition matrix is defined as $T = Blockdiag[T_L \; T_\gamma \; T_\beta]$ with

$$T_L = \begin{bmatrix} I_{[D]} \otimes T_L & \mathbf{0}_{[2D \times 2]} \\ \mathbf{1}'_{[D]} \otimes T_L & \mathbf{0}_{[2 \times 2]} \end{bmatrix}, \; T_\gamma = \begin{bmatrix} I_{[D]} \otimes H & \mathbf{0}_{[(s-1)D \times (s-1)]} \\ \mathbf{1}'_{[D]} \otimes H & \mathbf{0}_{[(s-1) \times (s-1)]} \end{bmatrix}, \; T_\beta = I_{[K]},$$

where $T_L = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$, , and $I_{[p]}$ is a p-dimensional identity matrix, and $\mathbf{0}_{[m \times n]}$ denotes an $m \times n$ matrix. In this application, the number of seasons is $s=4$, which implies three

seasonal stochastic variables per domain and thus a three-dimensional design matrix H:

$$H = \begin{bmatrix} \cos\frac{2\pi}{s} & \sin\frac{2\pi}{s} & 0 \\ -\sin\frac{2\pi}{s} & \cos\frac{2\pi}{s} & 0 \\ 0 & 0 & -1 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix}.$$

The design matrix Z_t for the measurement equation $\widehat{Y}_t = Z_t \alpha_t + \nu_t$ has the following form:

$$Z_t = \begin{bmatrix} I_{[D+1]} \otimes (1 \quad 0) & I_{[D+1]} \otimes (1 \quad 0 \quad 1) & Z_{\beta,t} \end{bmatrix},$$

where $Z_{\beta,t}$ consists of K vertical vectors $x_{t,k}$ extended with one element in order to include each of the level interventions contained in vector α_t^β into the aggregated signal. As described in Subsection 4.1, five level interventions are modelled for the DRTS, which results in:

$$Z_{\beta,t} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \delta_{t,5} \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \delta_{t,1} & \delta_{t,2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \delta_{t,3} & \delta_{t,4} & 0 \\ \delta_{t,1} & \delta_{t,2} & \delta_{t,3} & \delta_{t,4} & \delta_{t,5} \end{bmatrix}.$$

This implies that the magnitude of breaks and outliers in the aggregated series is equal to the cumulative magnitude of breaks and outliers in the domains at each point in time. Further, proper restrictions have to be imposed on the covariance matrix of state disturbances $Q = E(\eta_t \eta_t') = Blockdiag[Q_L \; Q_\gamma \; Q_\beta]$. Elements of Q are defined below. These restrictions imply that the covariance between the slope disturbances of the

aggregated series and those of the d-th domain must be equal to the sum of covariances between the stochastic terms of the domain in question and the other series:

$$\text{Cov}\big(\eta_{R,t,d}, \eta_{R,t,\text{Tot}}\big) = \sum_{d'=1}^{D} \text{Cov}\big(\eta_{R,t,d}, \eta_{R,t,d'}\big), \ d \in \{1, \dots, D\}. \tag{2.1}$$

For the slope disturbance variance of the aggregated series, the following will hold:

$$\sigma_{R,D+1}^2 = \sum_{d=1}^{D} \sigma_{R,d}^2 + 2 \sum_{d'=1}^{D} \sum_{d=d'+1}^{D} \varsigma_{\eta_R,d,d'}. \tag{2.2}$$

These restrictions translate into the following matrix:

$$\boldsymbol{Q_L} = \boldsymbol{G Q_R G'}, \text{ where}$$

$$\boldsymbol{G} = \begin{bmatrix} \boldsymbol{I}_{[D]} \otimes \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \\ \boldsymbol{1'}_{[D]} \otimes \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \end{bmatrix}'$$

and $\boldsymbol{Q_R}$ is a $[2D \times 2D]$ covariance matrix of the D domains' slope disturbance terms:

$$\boldsymbol{Q_R} = \begin{bmatrix} 0 & 0 & 0 & 0 & & 0 & 0 \\ 0 & \sigma_{\eta_R,1}^2 & 0 & \varsigma_{\eta_R,1,2} & & 0 & \varsigma_{\eta_R,1,D} \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & \varsigma_{\eta_R,2,1} & 0 & \sigma_{\eta_R,2}^2 & & 0 & \varsigma_{\eta_R,2,D} \\ & \vdots & & & \ddots & & \vdots \\ 0 & 0 & 0 & 0 & & 0 & 0 \\ 0 & \varsigma_{\eta_R,D,1} & 0 & \varsigma_{\eta_R,D,2} & \cdots & 0 & \sigma_{\eta_R,D}^2 \end{bmatrix}.$$

If covariances between the seasonal disturbances are restricted to zero, restrictions on the covariance matrix of harmonic disturbances are reduced to a much simpler form than those for the slope disturbances:

$$\text{Cov}\big(\omega_{t,d,j}, \omega_{t,D+1,j}\big) = \text{Cov}\big(\omega_{t,d,j}^*, \omega_{t,D+1,j}^*\big) = \sigma_{\omega,d}^2, \ j \in \{1, \dots, [s/2]\}, d \in \{1, \dots, D\},$$

$$\sigma_{\omega,D+1}^2 = \sum_{d=1}^{D} \sigma_{\omega,d}^2.$$

These restrictions have the following state-space representation:

$$= \begin{bmatrix} \boldsymbol{\Gamma_1} & \boldsymbol{O} & \boldsymbol{O} & & \boldsymbol{\Gamma_1} \\ \boldsymbol{O} & \boldsymbol{\Gamma_2} & \boldsymbol{O} & \cdots & \boldsymbol{\Gamma_2} \\ \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{\Gamma_3} & & \boldsymbol{\Gamma_3} \\ & \vdots & & \ddots & \vdots \\ \boldsymbol{\Gamma_1} & \boldsymbol{\Gamma_2} & \boldsymbol{\Gamma_3} & \cdots & \sum_{d=1}^{D} \boldsymbol{\Gamma_d} \end{bmatrix}, \boldsymbol{\Gamma_d} = Diag\big[\sigma_{\omega,d,1}^2 \ \sigma_{\omega^*,d,1}^2 \ \sigma_{\omega,d,2}^2\big], d \in \{1, \dots,$$

where $\sigma_{\omega,d,j}^2$, $\sigma_{\omega^*,d,j}^2$, are the disturbance variances corresponding to the j-th seasonal harmonic of domain d, and $\boldsymbol{O}$ is a 3×3 zero-matrix.

The level interventions are modelled as time-invariant, so their covariance matrix is a $K \times K$ zero-matrix: $\boldsymbol{Q_\beta} = \boldsymbol{O}_{[K \times K]}$.

Assuming that the covariances between the domain survey errors are equal to zero, the restrictions on the covariance matrix $\boldsymbol{R_t} = E\big(\boldsymbol{v_t v_t'}\big)$ of the measurement equation disturbances will have a simplified form:

$$\varsigma_{v,t,d,\text{Tot}} = \sigma_{v,t,d}^2, \ d \ \epsilon \ \{1, \ldots, D\}, \tag{2.3}$$

$$\sigma_{v,t,\text{Tot}}^2 = \sum_{d=1}^{D} \sigma_{v,t,d}^2. \tag{2.4}$$

Matrix $\boldsymbol{R_t}$ has the following state-space form:

$$,\boldsymbol{R_t} = \begin{bmatrix} \sigma_{v,t,1}^2 & 0 & 0 & & \sigma_{v,t,1}^2 \\ 0 & \sigma_{v,t,2}^2 & 0 & \cdots & \sigma_{v,t,2}^2 \\ 0 & 0 & \sigma_{v,t,3}^2 & & \sigma_{v,t,3}^2 \\ & \vdots & & \ddots & \vdots \\ \sigma_{v,t,1}^2 & \sigma_{v,t,2}^2 & \sigma_{v,t,3}^2 & \cdots & \sum_{d=1}^{D} \sigma_{v,t,d}^2 \end{bmatrix},$$

where $\sigma_{v,t,d}^2$ stands for the measurement equation error term variance of domain d at time $t, d \ \epsilon \ \{1, \ldots, D\}$.

2.2 Simulation

This simulation is done in order to check whether and how outcomes from a $(D + 1)$-dimensional model with appropriate restrictions imposed, differ from those obtained from a D-dimensional one. The model for this simulation is built similarly to the DRTS model, although in a slightly simplified way. It contains only three domains and their aggregate. Within this simulation, one set of series is generated for 200 time-points using the SsfPack function *SsfRecursion* (Koopman *et al.*, 1999, Ch.4.1), with hyperparameters of a magnitude similar to that in the DRTS. The trend model is the same smooth-trend model, but the seasonal component is removed for simplicity. Domains 2 and 3 feature level breaks that also enter the aggregated series. The level break in domain 2 constitutes 20 percent of the trend level at the time point preceding the break. This level intervention lasts for only 50 time points, which makes it similar to a change in the population of interest in domain 9 of the DRTS. Domain 3 experiences a 70 percent level increase that lasts until the end of the series. Restrictions given in equations (2.1)–(2.4) obviously render the Kalman filter recursions singular. The issue here is, however, not the singularity (since the univariate approach to multivariate Kalman filtering, which is implemented in SsfPack package, does allow for singularity in the system matrices (see Koopman and Durbin, 2002), but rather the fact that no additional information enters the model. Therefore, it is expected that this approach is identical to a D-dimensional multivariate model, from which the aggregated series estimates can also be derived.

A comparison of the two models has been conducted twice, using two different sets of hyperparameter values. First, the Kalman filter in the D- and $(D + 1)$-dimensional models was run under the hyperparameters that had been used to generate the series. Under these true hyperparameter values, the point- and variance-estimates of the state variables turned out to be exactly the same. The four generated series are presented in Fig. B.1, along with their filtered signal point-estimates. The variance-estimates are not presented because they are equal to each other.

In the other comparison of the two models, maximum-likelihood estimates are used instead of the hyperparameter true values in the Kalman filter recursions. In most cases, evaluation of the likelihood function in *SsfPack* fails due to the singularity problem, so the set of the

hyperparameter maximum-likelihood estimates for running the Kalman filter under the $(D + 1)$-dimensional model would have to be taken from the D-dimensional model, which is also done in this simulation. Filtered signals of the two models based on the estimated hyperparameters are almost identical to those based on the true hyperparameter values. The signal standard errors of the D- and $(D + 1)$-dimensional models are presented in Fig. B.2. The figure shows that, as soon as the hyperparameters used by the Kalman filter deviate from their true values the series were generated with, the signal variances of the two models deviate from each other. In this simulation, the model is quite simple, and the pattern of the input series seems to be much more stable compared to the series in the DRTS, so the hyperparameters are estimated quite accurately (see Table B.1 below). Therefore, the differences between the signal variances of the two models are minute (see Fig. B.2). These differences are so small that they can only be seen if zoomed in (note the scale of the y-axis). For this reason, only the last twenty time points are presented here. The reader should not be misled by the seemingly steep decline in the standard errors. This decrease is very subtle and is part of the wavy behaviour inherent in these standard errors.

B.1. Simulated series (grey line) and their filtered signal estimates (black line) based on the true hyperparameter values in the Kalman filter recursions.
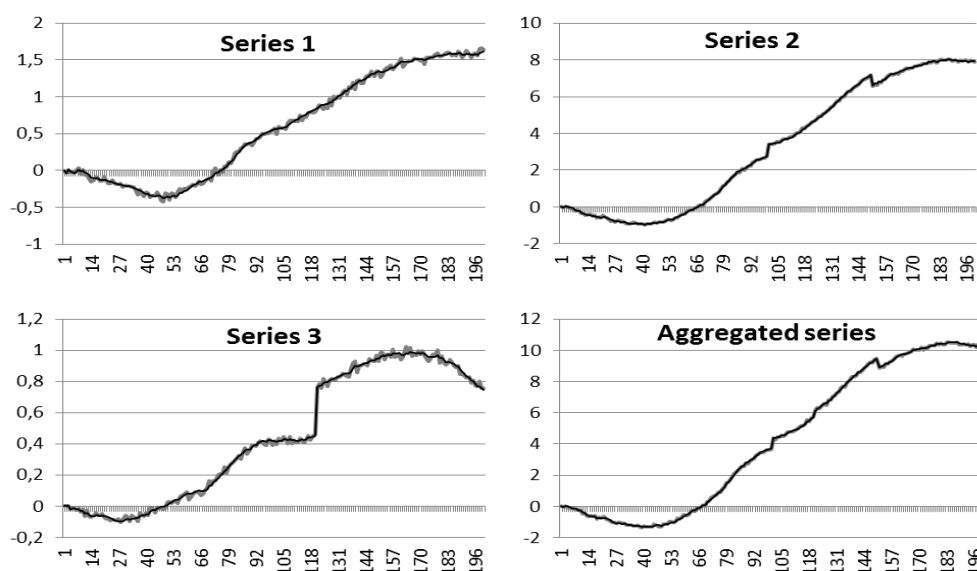

In order to see which model variance estimates are better, one can plot them together with the true variances, the latter being obtained from a model that uses the true hyperparameter values for the Kalman filtering (not presented here due to the scale problem). In the case of this simulation, there is no consistent pattern in which one or the other model over- or underestimates the true signal variances.

B.1. True hyperparameter values and their maximum-likelihood estimates.

	$\ln(\sigma^2_{\eta_R,1})$	$\ln(\sigma^2_{\eta_R,2})$	$\ln(\sigma^2_{\eta_R,3})$	$\varsigma_{\eta_R,1,2}$	$\varsigma_{\eta_R,1,3}$	$\varsigma_{\eta_R,2,3}$	$\ln(\sigma^2_{v,1})$	$\ln(\sigma^2_{v,2})$	$\ln(\sigma^2_{v,3})$
True	-12.00	-11.00	-15.00	3.00	0.30	0.20	-7.00	-6.50	-8.00
Estimated	-12.30	-11.10	-15.30	3.24	0.29	0.24	-6.82	-6.34	-8.01
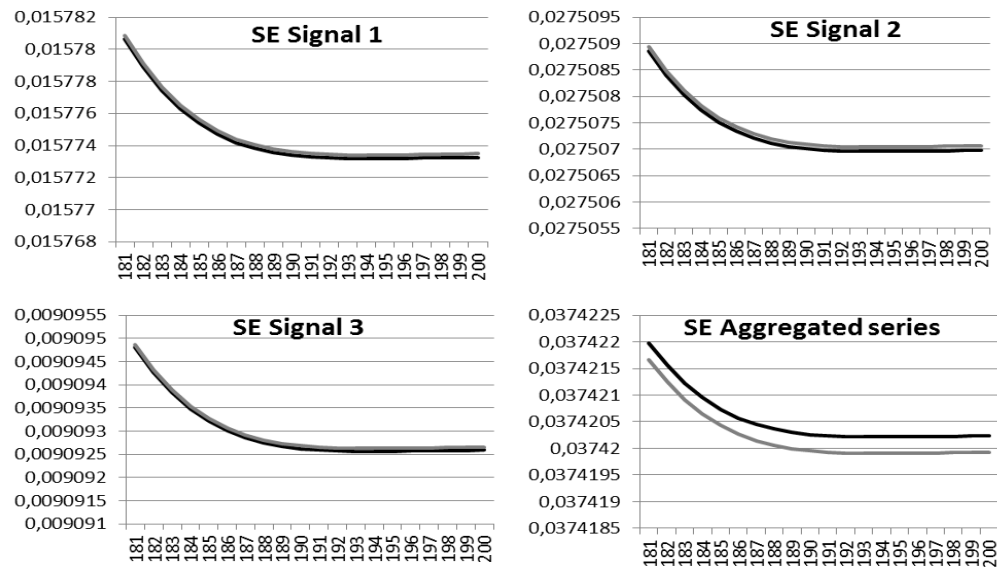
The simulation suggest that, at least in the case when the true hyperparameters are known, the D- and $(D+1)$-dimensional models produce identical outcomes. Therefore, the more parsimonious D-dimensional model is preferable.

2.3 Application to the DRTS

This subsection illustrates results from the $(D+1)$-dimensional model applied to the DRTS series. In particular, the signal variance estimates are compared to those from the D-dimensional model presented in the main body of this article. The maximum-likelihood estimates for the hyperparameters of the ten-dimensional model are given in Appendix A. Since no additional information enters the model when the aggregated series is included into the model, the hyperparameter maximum-likelihood estimates are expected to be equal or at least close to those in the nine-dimensional setting. This is indeed the fact for the measurement equation error term variances and their asymmetric confidence intervals (compare Tables A.3 and A.4 in Appendix A). An exception in this application is a nearly zero estimate of the mid-period measurement equation error term variance of domain 6 in the ten-dimensional setting. It turns out that the numerical optimization procedure in this model persistently ends up in a local optimum. Since the two models – with and without the aggregated series – are asymptotically equivalent, the hyperparameter maximum-likelihood estimates obtained from the nine-dimensional model (in Table A.3) can be treated as given in the ten-dimensional setting. Whether one follows this approach or re-estimates the hyperparameters in the ten-dimensional setting, results in very similar point and variance estimates of the signals (except for the mid-periods in domains 6 and 9). The aim of this section, however, is to illustrate differences that emerge in the Kalman filter procedure carried out under the same set of hyperparameter estimates in both model settings. Therefore, the results presented below are obtained from the two models on the basis of hyperparameters estimated under the nine-dimensional model.

As has been mentioned in the previous subsection, the pattern of the DRTS series is much more erratic than that of the series simulated. Apart from that, the model structure is more

complex in that the model dimension is larger (with 9 domains), and in that a seasonal component enters the model. These problems could have been responsible for visible differences between the point-estimates of the two model settings (in domains 4, 5 and 9, see Fig. B.3). The differences between the signal standard errors are also quite considerable, as shown in Fig. B.4. Moreover, the signal standard errors from the ten-dimensional model are consistently lower than those from the nine-dimensional one. However, without true hyperparameter values at hand, it would be groundless to affirm that the ten-dimensional model estimates are better. In view of the simulation results, the more parsimonious D-dimensional model may be considered, from which the aggregated series estimates can be derived.

B.3. DRTS design estimates (dashed line) and filtered signal point-estimates from the nine-dimensional model (grey line) and from the ten-dimensional model that uses hyperparameter estimates from the nine-dimensional model (black line), 1000 ton.
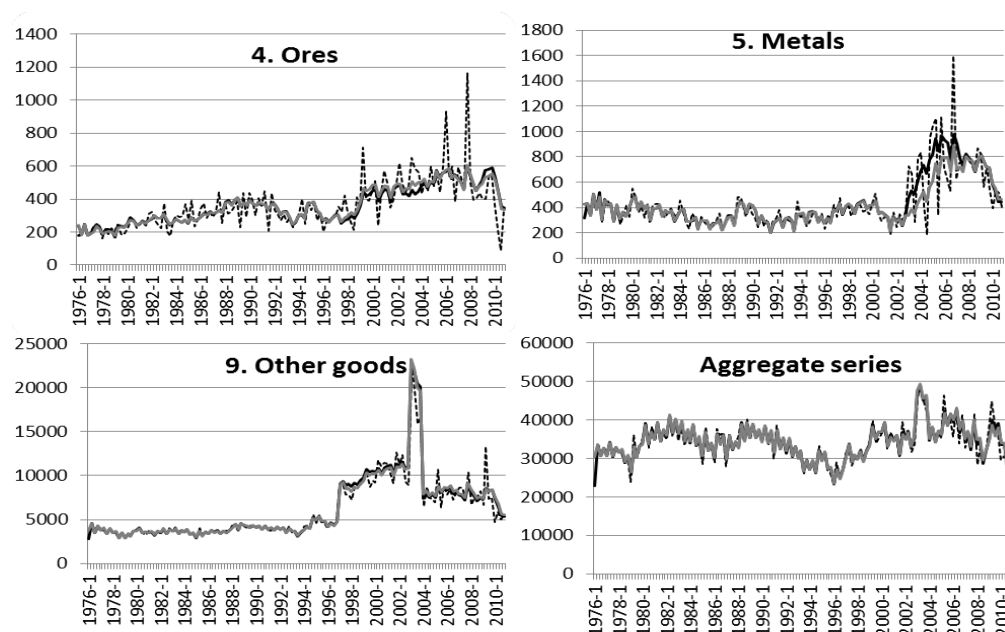
B.4. Standard errors of the filtered signal estimates obtained from the DRTS models: the nine-dimensional model (grey line), and ten-dimensional model that uses hyperparameter estimates from the nine-dimensional model (black line), 1000 ton.
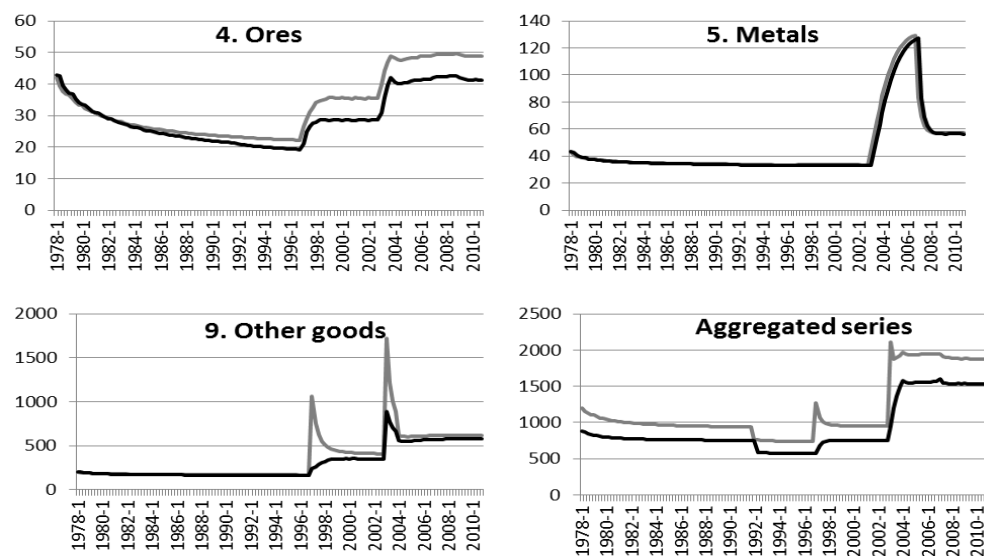
Explanation of symbols

.	Data not available
*	Provisional figure
**	Revised provisional figure (but not definite)
x	Publication prohibited (confidential figure)
–	Nil
–	(Between two figures) inclusive
0 (0.0)	Less than half of unit concerned
empty cell	Not applicable
2013–2014	2013 to 2014 inclusive
2013/2014	Average for 2013 to 2014 inclusive
2013/'14	Crop year, financial year, school year, etc., beginning in 2013 and ending in 2014
2011/'12–2013/'14	Crop year, financial year, etc., 2011/'12 to 2013/'14 inclusive

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