Statistics Netherlands

Discussion Paper

Accounting for Hyperparameter Uncertainty in Small Area Estimation Based on State-Space Models

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

2015 | 05

Oksana Bollineni-Balabay
Jan van den Brakel
Franz Palm

Contents

Structural time series models are known as a powerful technique for variance reduction in the framework of small area estimation (SAE) of repeatedly conducted surveys. Such models, however, contain unknown hyperparameters that have to be estimated before the Kalman filter can be launched to estimate state variables of the model. If the uncertainty around these hyperparameters is not taken into account, the newly obtained variance estimates for the state variables of interest become negatively biased, particularly in short time series. The Dutch Labour Force survey (DLFS) is a typical example of SAE in that its sample sizes are too small to produce reliable monthly figures using design-based estimators known from classical sampling theory. Statistics Netherlands currently uses a structural time series model to reduce the variance of the design-based estimates of the unemployed labour force. In order to account for the negative bias in the DLFS model variance estimates, several approximation approaches known in the literature are considered. The results suggest that the best performing approach out of these can correct for a 2-percent negative relative bias of the signal MSE produced by the Kalman filter by offering a positive bias of 2 percent. After accounting for the hyperparameter uncertainty, the standard errors of the model estimates are about 22 percent smaller than the design-based standard errors.

1 Introduction

Monthly figures on the labour force produced by national statistical institutes (NSIs) are important economic indicators. These indicators are generally based on Labour Force Surveys (LFS). Most NSIs apply a rotating panel design in their LFS, but in most cases the sample size is still not large enough to produce sufficiently precise monthly figures about the labour force based on design-based estimators known from classical sampling theory (Särndal et al. (1992), Cochran (1977)). In such cases, statistical modelling can be used to improve the effective sample size of domains by borrowing information from preceding periods or other domains. Such techniques are often referred to as small area estimation (SAE), see Rao (2003). Repeatedly conducted surveys in particular have a potential for improvement within the framework of structural time series (STS) or multilevel time series models.

Common applied SAE procedures are based on multilevel models that are estimated with the empirical best linear unbiased prediction (EBLUP) or the hierarchical Bayesian (HB) approaches. Such models usually contain unknown hyperparameters that have to be estimated, which translates into larger standard errors around the domain predictions. If this uncertainty (here and further in this paper referred to as the hyperparameter uncertainty) is not taken into account, the MSE estimates of the quantities of interest become negatively biased. Within the framework of the EBLUP and HB approaches, it is very common to take the hyperparameter uncertainty into account, see Rao (2003), Ch.6-7, 10.

STS models are not as widely used in SAE as multilevel models. The Kalman filter usually applied to analyse STS models ignores the hyperparameter uncertainty, and therefore produces negatively-biased MSE estimates. For this reason, the apparent gains from the STS technique in terms of reduced variance estimates have to be treated without undue optimism. Applications that give evidence for substantial advantages of STS models over the design-based approach treat the estimated model hyperparameters as known, see, e.g., Krieg and van den Brakel (2012), Van den Brakel and Krieg (2009a), Pfeffermann and Bleuer (1993), Tiller (1992).

At Statistics Netherlands, a multivariate STS model is used to produce official monthly labour force figures. This model was originally proposed by Pfeffermann (1991) and is referred to as the DLFS model in this paper. It uses sample information from preceding time periods and accounts for different aspects of the rotating panel design, such as the so-called rotation group bias (RGB) and autocorrelation in the survey errors. In this way, sufficiently precise monthly estimates for the unemployed labour force are obtained (see Van den Brakel and Krieg (2009a)). However, this application, just like the other aforementioned studies, does not account for the hyperparameter uncertainty in the estimated MSEs.

This paper focuses on the true mean square error (MSE) estimation in the case of the DLFS model. The literature offers several ways to account for the hyperparameter uncertainty in STS models. Among those MSE approximation approaches considered in this paper are the asymptotic approximation developed by Hamilton (1986), as well as parametric and non-parametric bootstrapping approaches developed by Pfeffermann and Tiller (2005) and Rodriguez and Ruiz (2012). These methods are applied to the DLFS model to see whether the hyperparameter uncertainty matters in terms of increased MSEs of the quantities of interest in this real life application.

This paper presents an extended Monte-Carlo simulation study, where the DLFS model acts as the data generation process. The contribution of the paper is three-fold. First of all, it establishes the best approach to the MSE approximation for the DLFS and offers a more realistic evaluation of the variance reduction obtained with the STS model compared to the design-based approach. Secondly, this Monte-Carlo study verifies the claim of Rodriguez and Ruiz (2012) about the superiority of their method over the bootstrap of Pfeffermann and Tiller (2005) in a more complex model. Finally, the Monte-Carlo simulation shows how the model can be checked for possible misspecification.

The paper is structured as follows. Section 2 contains a general description of the DLFS survey design and presents the model currently used by Statistics Netherlands. Section 3 reviews the above-mentioned approaches to the MSE approximation. Some details on the simulation and bootstrap setup specific for the DLFS are given in Section 4. Results of the simulation study are presented in Section 5.

2 The LFS model

The DLFS is based on a stratified two-stage cluster design for households with estimates produced on a monthly, quarterly and annual basis. More details on the sampling design are to be found in Van den Brakel and Krieg (2009a). Quarterly design-based estimates for the unemployed labour force are obtained with the general regression (GREG) estimator (Särndal et al. (1992)). In the case of the DLFS, the GREG estimator is based on a complex model that uses a set of socio-demographic categorical variables, see Van den Brakel and Krieg (2009b). The series considered in the present study are monthly estimates of the total number of the unemployed labour force. Since October 1999, the survey has been conducted as a five-wave rotating panel survey, where a sample of households is drawn every month. Households are approached five times, with a three-month gap between the interviews. The term "wave" in this context means a sample of households that enter the LFS panel at time t and leave the panel in 12 months after five interviews. Let Y_t^{t-j} denote the GREG estimate of the total number of the unemployed labour force in month t. Five such estimates are obtained per month, each of them being respectively based on the sample that entered the survey in month $t-j$,

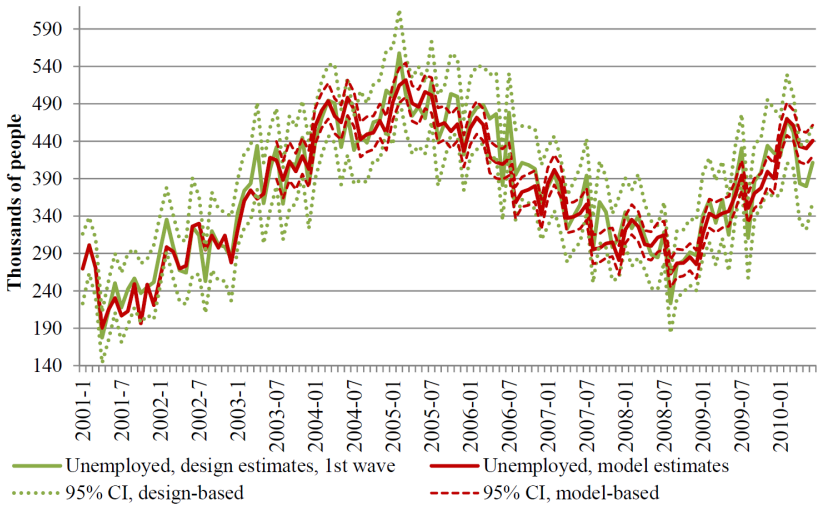

Figure 2.1 Numbers of unemployed in the Netherlands: design- and model-based estimates

$j = \{0, 3, 6, 9, 12\}$. In the middle of 2010, the DLFS was subject to a survey transition that resulted in substantial discontinuities, which required an extension of the model. Therefore, the present study covers the time span from January 2001 until June 2010, in order to avoid unnecessary model complications that are not particularly relevant for this Monte-Carlo study for MSE approximation.

Rotating panel designs are known to have systematic differences between the estimates of their subsequent waves. This phenomenon is usually referred to as the rotation group bias (RGB), see, e.g., Bailar (1975), Kumar and Lee (1983), or Pfeffermann (1991). Common reasons behind the RGB are panel attrition, panel-effects, and differences in questionnaires and modes used in the subsequent waves. In the case of the DLFS, the first wave estimates are assumed to be most reliable, with the subsequent waves systematically underestimating the unemployed labour force numbers, which is reflected by a negative RGB in the level. The seasonal component may be affected too, but in this application the seasonal RGB turned out to be negligible (see Van den Brakel and Krieg (2009a) for details).

Apart from the RGB, another problem with the LFS is small monthly sample sizes. With the first wave net sample size of about 6500 persons on average, the GREG estimates cannot produce sufficiently reliable unemployment figures on a monthly basis. Both problems can be solved with the help of a STS model, which is currently used in the production of official LFS figures. Such a model makes use of the information accumulated over time to produce point-estimates with smaller standard errors.

In a STS model, the series of design estimates is decomposed into unobserved components, whereupon the so-called signal - a more reliable series of point-estimates - can be obtained. The signal is usually extracted with the Kalman filter. The filter removes a great part of the population and sampling noise from the GREG-estimates and produces point- and variance estimates for the signal and its unobserved components. As a result, not only are these signal point-estimates less volatile (see Fig. 2.1), their standard errors are usually substantially lower compared to the design standard errors (in the case of the DLFS, 24 percent smaller). Apart from that, an STS model can extract the RGB pattern from the GREG series.

Let the five-dimensional vector $\boldsymbol{Y_t} = (Y_t^t \; Y_t^{t-3} \; Y_t^{t-6} \; Y_t^{t-9} \; Y_t^{t-12})'$ denote the GREG estimates of the total number of the unemployed labour force for the five DLFS waves observed at time t. Based on Pfeffermann (1991), Van den Brakel and Krieg (2009a) developed the following model

for the five-dimensional vector with GREG estimates :

$$\boldsymbol{Y_t} = \mathbf{1_5}\xi_t + \boldsymbol{\lambda_t} + \boldsymbol{e_t}, \tag{2.1}$$

where $\mathbf{1_5}$ is a five-dimensional column vector of ones, $\boldsymbol{\lambda_t}$ is a vector containing the RGBs, and $\boldsymbol{e_t}$ is a vector with the survey errors. It is assumed that the true population parameter is $\xi_t = L_t + \gamma_t + \varepsilon_t$, which is the sum of a stochastic trend, a stochastic seasonal component, and an irregular component. For the full description of the DLFS model, the reader is referred to Van den Brakel and Krieg (2009b). The description below summarizes its main features.
For the stochastic trend L_t, the so-called smooth-trend model is assumed:

$$\begin{aligned} L_t &= L_{t-1} + R_{t-1}, \\ R_t &= R_{t-1} + \eta_{R,t}, \end{aligned} \tag{2.2}$$

where L_t and R_t represent the level and slope of the true population parameter, respectively, with the slope disturbance term being distributed as: $\eta_{R,t} \overset{iid}{\sim} N(0, \sigma_R^2)$.
In the case of monthly data, the seasonal component γ_t can be trigonometrically decomposed into six stochastic harmonics:

$$\gamma_t = \sum_{l=1}^{6} \gamma_{t,l}, \tag{2.3}$$

where each of these six harmonics follows the process:

$$\begin{aligned} \gamma_{t,l} &= cos(h_l)\gamma_{t-1,l} + sin(h_l)\gamma^*_{t-1,l} + \omega_{t,l}, \\ \gamma^*_{t,l} &= -sin(h_l)\gamma_{t-1,l} + cos(h_l)\gamma^*_{t-1,l} + \omega^*_{t,l}, \end{aligned}$$

with $h_l = \frac{\pi l}{6}$ being the l-th seasonal frequency, $l = \{1, \ldots 6\}$. The zero-expectation stochastic terms $\omega_{t,l}$ and $\omega^*_{t,l}$ are assumed to be normally and independently distributed and to share the same variance within and across all the harmonics, such that:

$$Cov(\omega_{t,l}, \omega_{t',l'}) = Cov(\omega^*_{t,l}, \omega^*_{t',l'}) = \begin{cases} \sigma_\omega^2 & \text{if } l = l' \text{ and } t = t', \\ 0 & \text{if } l \neq l' \text{ or } t \neq t', \end{cases}$$

$Cov(\omega_{t,l}, \omega^*_{t,l}) = 0$ for all l and t.

Since the estimates of the first wave are assumed to be RGB-free (as motivated in Van den Brakel and Krieg (2009a)), the RGB vector for the five waves can be written in the following form: $\boldsymbol{\lambda_t} = (0\ \lambda_t^{t-3} \lambda_t^{t-6} \lambda_t^{t-9} \lambda_t^{t-12})'$. The RGB variables for the last four waves are time-dependent and are modelled as a random walk process:

$$\lambda_t^{t-j} = \lambda_{t-1}^{t-j} + \eta_{\lambda,t}^{t-j}, j = \{3, 6, 9, 12\}. \tag{2.4}$$

The RGB disturbances are not correlated across different waves and are normally distributed: $\eta_{\lambda,t}^{t-j} \overset{iid}{\sim} (0, \sigma_\lambda^2)$, with equal variances in all the four waves.
The last component in the model is the survey error. Their variance estimates are the GREG design variances available from the micro-data, and so the survey errors can be modelled as $e_t^{t-j} = \tilde{e}_t^{t-j} \sqrt{Var(\hat{Y}_t^{t-j})}$. This formulation makes it possible to account for heterogeneity caused by differences in the survey design and sample sizes over time, see Binder and Dick (1990). Since the samples of different waves, starting from wave 2, are based on the same people as the preceding wave samples, there will be correlation between the survey errors. To account for this correlation, the survey error terms in Van den Brakel and Krieg (2009a) are modelled as an AR(1) process:

$$\begin{aligned} \tilde{e}_t^t &= v_t^t; \\ \tilde{e}_t^{t-j} &= \rho \tilde{e}_{t-3}^{t-j} + v_t^{t-j}, j = \{3, 6, 9, 12\}. \end{aligned} \tag{2.5}$$

The survey error autoregressive parameter ρ is common to all the four waves and is estimated from the input data using a procedure developed by Pfeffermann (1998). In this application $\hat{\rho} = 0.208$, see Van den Brakel and Krieg (2009b). This estimate is used as a prior input into the model, together with the survey design standard errors of the five waves. Since the variance of the product $\sqrt{Var(\hat{Y}_t^{t-j})}\tilde{e}_t^{t-j}$ is meant to be close to the variance estimate of the GREG estimator, $Var(\tilde{e}_t^{t-j}) \approx 1$. For the first wave, $Var(\tilde{e}_t^t) = \sigma^2_{\nu_t^t} \approx 1$, and for the four subsequent waves, with an AR(1) process assumed, $Var(\tilde{e}_t^{t-j}) \approx \sigma^2_{\nu_t^{t-j}}/(1-\rho^2) \approx 1, \quad j = \{3, 6, 9, 12\}$. Five different hyperparameters $\sigma^2_{\nu_t^{t-j}}$ are assigned to the five waves, and their estimates have indeed turned out to be close to unity. It is assumed that ν-disturbance terms are normally and independently distributed: $\nu_t^{t-j} \overset{iid}{\sim} (0, \sigma^2_{\nu_t^{t-j}}), j = \{0, 3, 6, 9, 12\}$.

Linear structural time series models with unobserved components are usually analysed with the Kalman filter after putting them into a state-space form:

$$\boldsymbol{\alpha_{t+1}} = \boldsymbol{T_{t+1}}\boldsymbol{\alpha_t} + \boldsymbol{\eta_t}, \tag{2.6}$$

$$\boldsymbol{Y_t} = \boldsymbol{Z_t}\boldsymbol{\alpha_t}. \tag{2.7}$$

Here, $\boldsymbol{T_t}$ and $\boldsymbol{Z_t}$ are known design matrices of the transition and measurement equations, respectively. The transition equation (2.6) describes how the state vector $\boldsymbol{\alpha_t}$ evolves over time, whereas the measurement equation (2.7) reflects the linear relationship between the observations and the state vector. The autoregressive structure of survey errors in the rotating panel design can be taken into account if the errors $\tilde{e}_t^{t-j}$ are modelled as state variables. They thus disappear from the measurement equation (2.7) and move to the state vector $\boldsymbol{\alpha_t}$. The zero-expectation vector $\boldsymbol{\eta_t}$ contains identically and independently distributed disturbance terms. The design standard errors $\sqrt{Var(\hat{Y}_t^{t-j})}$ become time-varying elements of the matrix $\boldsymbol{Z_t}$ and are from now on denoted as $z_t^{t-j}, j = \{0, 3, 6, 9, 12\}$. More details on the state-space form of the DLFS model are presented in Appendix A.

Collecting the variables mentioned in this section produces the following state vector (with the survey error terms indexed differently for the state-space form):

$\boldsymbol{\alpha_t} = (\boldsymbol{\alpha_t^{\xi}}\ \boldsymbol{\alpha_t^{\lambda}}\ \boldsymbol{\alpha_t^{e}})'$, where

$\boldsymbol{\alpha_t^{\xi}} = (L_t\ R_t\ \gamma_{t,1}\ \gamma^*_{t,1} \ldots \gamma_{t,5}\ \gamma^*_{t,5}\ \gamma_{t,6})$,

$\boldsymbol{\alpha_t^{\lambda}} = (\lambda_t^{t-3}\lambda_t^{t-6}\lambda_t^{t-9}\lambda_t^{t-12})$,

$\boldsymbol{\alpha_t^{e}} = (\tilde{e}_t^t\ \tilde{e}_t^{t-3}\ \tilde{e}_t^{t-6}\ \tilde{e}_t^{t-9}\ \tilde{e}_t^{t-12}\ \tilde{e}_{t-2}^{t-2}\ \tilde{e}_{t-2}^{t-5}\ \tilde{e}_{t-2}^{t-8}\ \tilde{e}_{t-2}^{t-11}\ \tilde{e}_{t-1}^{t-1}\ \tilde{e}_{t-1}^{t-4}\ \tilde{e}_{t-1}^{t-7}\ \tilde{e}_{t-1}^{t-10})$.

All the state variables are initialised with a diffuse prior, except for the five survey error components $\tilde{e}_t^{t-j}, j = \{0, 3, 6, 9, 12\}$. These stationary state variables are initialised with zeros and with the initial variances taken equal to unity.

The disturbance variances, together with the autocorrelation parameter ρ, are collected in a hyperparameter vector called $\boldsymbol{\theta} = (\sigma_R^2\ \sigma_\omega^2\ \sigma_\lambda^2\ \sigma^2_{\nu^1}\ \sigma^2_{\nu^2}\ \sigma^2_{\nu^3}\ \sigma^2_{\nu^4}\ \sigma^2_{\nu^5}\ \rho)$, where the superscripts {1,…5} stand for the numbers of the waves (note that the variance of the last eight state variables is zero, see Appendix A). Hyperparameter ρ is going to be estimated as in Pfeffermann et al. (1998) from the input data, whereafter the disturbance variance hyperparameters are estimated by the quasi-maximum likelihood method, treating $\hat{\rho}$-estimates as given. The vector with the hyperparameter estimates is denoted as $\hat{\boldsymbol{\theta}}$.

Numerical analysis of this paper is conducted with OxMetrics 5 (Doornik (2007)) in combination with SsfPack 3.0 package (Koopman et al. (2008)).

3 Review of MSE Approximation Approaches

State variables in structural time series models are usually extracted by the Kalman filter according to the following recursions:

$$\hat{\boldsymbol{\alpha}}_{t|t} = \hat{\boldsymbol{T}}_t \hat{\boldsymbol{\alpha}}_{t-1|t-1} + \boldsymbol{K}_t \boldsymbol{\epsilon}_t,$$
$$\hat{\boldsymbol{P}}_{t|t} = \hat{\boldsymbol{P}}_{t|t-1} - \hat{\boldsymbol{P}}_{t|t-1} \boldsymbol{Z}_t' \boldsymbol{K}_t',$$
$$\hat{\boldsymbol{P}}_{t|t-1} = \hat{\boldsymbol{T}}_t \hat{\boldsymbol{P}}_{t-1|t-1} \hat{\boldsymbol{T}}_t' + \hat{\boldsymbol{Q}}_t,$$

where $'$ stands for a transpose, $\hat{\boldsymbol{\alpha}}_{t|t}$ and $\hat{\boldsymbol{P}}_{t|t} = \boldsymbol{E}_t[(\hat{\boldsymbol{\alpha}}_{t|t} - \boldsymbol{\alpha}_t)(\hat{\boldsymbol{\alpha}}_{t|t} - \boldsymbol{\alpha}_t)']$ denote the conditional mean of the state vector and its MSE, respectively, extracted by the Kalman filter based on information available up to and including time t. This kind of estimates is usually referred to as *filtered* estimates. Matrix $\hat{\boldsymbol{P}}_{t|t-1}$ is the predicted state covariance matrix with the prediction mean square error estimates on its main diagonal: $\hat{\boldsymbol{P}}_{t|t-1} = \boldsymbol{E}_t[(\hat{\boldsymbol{\alpha}}_{t|t-1} - \boldsymbol{\alpha}_t)(\hat{\boldsymbol{\alpha}}_{t|t-1} - \boldsymbol{\alpha}_t)']$, where the word *predicted* implies that the estimates for period t are based on information up to and including time $t-1$. $\boldsymbol{K}_t$ is the so-called Kalman gain: $\boldsymbol{K}_t = \hat{\boldsymbol{P}}_{t|t-1} \boldsymbol{Z}_t' \boldsymbol{F}_t^{-1}$, $\boldsymbol{\epsilon}_t = \boldsymbol{Y}_t - \boldsymbol{Z}_t \hat{\boldsymbol{\alpha}}_{t|t-1}$ are innovations and $\boldsymbol{F}_t$ is the innovation covariance matrix: $\boldsymbol{F}_t = \boldsymbol{Z}_t \hat{\boldsymbol{P}}_{t|t-1} \boldsymbol{Z}_t'$ (note that the covariance matrix of the measurement equation error terms is lacking here because those error terms have been placed in the state vector). Note that the covariance matrix $\boldsymbol{Q}$ and transition matrix $\boldsymbol{T}$ are time-invariant here, the former being a diagonal matrix with the state disturbance variance hyperparameters, and the latter containing the autoregressive parameter ρ.

The mean square error extracted by the Kalman filter conditionally on the information up to and including time t is:

$$\boldsymbol{P}_{t|t} = \boldsymbol{E}_t[(\hat{\boldsymbol{\alpha}}_{t|t}(\boldsymbol{\theta}) - \boldsymbol{\alpha}_t)(\hat{\boldsymbol{\alpha}}_{t|t}(\boldsymbol{\theta}) - \boldsymbol{\alpha}_t)'], \tag{3.1}$$

where the expectation is taken with respect to the distribution of the state vector at time t, provided this expectation exists. In practice, however, the true hyperparameter vector is replaced by its estimates $\hat{\boldsymbol{\theta}}$ in the Kalman filter recursions. Then, the MSE in (3.1) is no longer the true MSE and is called "naive" as it does not incorporate the uncertainty around the $\hat{\boldsymbol{\theta}}$-estimates. The true MSE then becomes:

$$\boldsymbol{MSE}_{t|t} = \boldsymbol{E}_t[(\hat{\boldsymbol{\alpha}}_{t|t}(\hat{\boldsymbol{\theta}}) - \boldsymbol{\alpha}_t)(\hat{\boldsymbol{\alpha}}_{t|t}(\hat{\boldsymbol{\theta}}) - \boldsymbol{\alpha}_t)'], \tag{3.2}$$

which is larger than the MSE in (3.1) and can be decomposed as the sum of the filter uncertainty and parameter uncertainty:

$$\boldsymbol{MSE}_{t|t} = \boldsymbol{E}_t[(\hat{\boldsymbol{\alpha}}_{t|t}(\boldsymbol{\theta}) - \boldsymbol{\alpha}_t)(\hat{\boldsymbol{\alpha}}_{t|t}(\boldsymbol{\theta}) - \boldsymbol{\alpha}_t)'] + \boldsymbol{E}_t[(\hat{\boldsymbol{\alpha}}_{t|t}(\hat{\boldsymbol{\theta}}) - \hat{\boldsymbol{\alpha}}_{t|t}(\boldsymbol{\theta}))(\hat{\boldsymbol{\alpha}}_{t|t}(\hat{\boldsymbol{\theta}}) - \hat{\boldsymbol{\alpha}}_{t|t}(\boldsymbol{\theta}))']. \tag{3.3}$$

The first term, the filter uncertainty, is approximated by the naive MSE-estimates $\hat{\boldsymbol{P}}_{t|t}$ delivered by the Kalman filter. Approximation of the second term, the parameter uncertainty, requires some additional effort. The literature on the true mean square error approximation proposes two main approaches: asymptotic approximation and bootstrapping. Bootstrapping can be performed in a parametric or non-parametric way. A few general details have to be mentioned about these methods in the context of STS models and of the DLFS model specifically.

If a STS model contains non-stationary components, as is the case with the DLFS model, a special procedure is required for bootstrap samples to be drawn conditionally on the given dataset $\{\boldsymbol{Y}_1, ..., \boldsymbol{Y}_T\}$. The simulation smoother algorithm developed by Durbin and Koopman (2002) can be exploited to generate conditional draws for the trend, seasonal and the RGB state variables. At the first step of this algorithm, the state variables and the data series are

generated unconditionally on the original data set, either parametrically or non-parametrically. The generated series $\{\boldsymbol{Y_1^{b,\dagger}}, ..., \boldsymbol{Y_T^{b,\dagger}}\}$ will surely diverge from the dataset they have been bootstrapped from. They must therefore undergo some kind of correction. The state variables are corrected by the magnitude of the smoothed mean of their counterparts extracted from a "correction" dataset, which constitutes differences between the original data $\{\boldsymbol{Y_1}, ..., \boldsymbol{Y_T}\}$ and the unconditional bootstrap dataset $\{\boldsymbol{Y_1^{b,\dagger}}, ..., \boldsymbol{Y_T^{b,\dagger}}\}$, as described in Koopman, Harvey, Doornik, and Shephard (2008), Ch.8.4.2. The survey errors generated as described in either parametric or non-parametric unconditional state recursion, do not need any adjustments as they constitute (autocorrelated) noise.

For the parametric bootstrap, the necessary disturbances for state recursions (2.6) and (2.7) are drawn from their joint conditional multivariate normal density $\boldsymbol{\eta_t} \overset{iid}{\sim} MN(\mathbf{0}, \boldsymbol{\Omega})$.

Non-parametric bootstrap has an advantage of not depending on any particular assumption about this joint distribution. In this case, a bootstrap sample of standardised innovations $\{\boldsymbol{\epsilon_{r+1}^{b,St}}, ..., \boldsymbol{\epsilon_T^{b,St}}\}$ is obtained by sampling with replacement from $\{\boldsymbol{\epsilon_{r+1}^{St}}(\hat{\boldsymbol{\theta}}), ..., \boldsymbol{\epsilon_T^{St}}(\hat{\boldsymbol{\theta}})\}$ where $\boldsymbol{\epsilon_t^{St}} = \boldsymbol{F_t^{-1/2}}(\hat{\boldsymbol{\theta}})\boldsymbol{\epsilon_t}(\hat{\boldsymbol{\theta}})$ are standardized innovations from the Kalman filter recursions based on the original ML estimates of the hyperparameters. For the DLFS model, a burn-in period r of 30 time-points is chosen. This choice is motivated by the number of state variables that are initialised with a diffuse prior, which is 25; additional 5 time points are skipped to have a burn-in period of two and a half years, and also in order to be on a safe-side. A bootstrap observation set $\{\boldsymbol{Y_1^b}, ..., \boldsymbol{Y_T^b}\}$ is then constructed by running the so-called innovation form of the Kalman filter:

$$\hat{\boldsymbol{\alpha}}_{\boldsymbol{t|t}}^{\boldsymbol{b}} = \hat{\boldsymbol{\alpha}}_{\boldsymbol{t|t-1}}^{\boldsymbol{b}} + \boldsymbol{K_t}(\hat{\boldsymbol{\theta}})\boldsymbol{F_t^{1/2}}(\hat{\boldsymbol{\theta}})\boldsymbol{\epsilon_t^{b,St}} \tag{3.4}$$

$$\hat{\boldsymbol{Y}}_{\boldsymbol{t}}^{\boldsymbol{b}} = \boldsymbol{Z_t}\hat{\boldsymbol{\alpha}}_{\boldsymbol{t|t-1}}^{\boldsymbol{b}} + \boldsymbol{F_t^{1/2}}(\hat{\boldsymbol{\theta}})\boldsymbol{\epsilon_t^{b,St}}, t = d+1, ..., T; \tag{3.5}$$

Note, that the univariate version of a multivariate Kalman filter, suggested by Koopman and Durbin (2000), is computationally more efficient. This version is implemented in SsfPack package that is used for this work. In this case, each series $Y_{t,d}$ in a multivariate model is treated as a univariate one. Then, for each of these series, the Kalman gain $\boldsymbol{K_{t,d}}$ is a vector, whereas $F_{t,d}$ is a scalar.

The following sections contain a brief presentation of the asymptotic approach, as well as the recent Rodriguez and Ruiz (2012) bootstrap approaches (hereafter referred to as the RR bootstrap) and Pfeffermann and Tiller (2005) (hereafter the PT bootstrap) bootstrap approaches.

3.1 Rodriguez and Ruiz Bootstrapping Approach

Rodriguez and Ruiz (2012) developed their bootstrap method for the mean square error approximation conditional on the data. Bootstrapping can be done both parametrically and non-parametrically, following the steps below:

1. Estimate the model and obtain all the hyperparameter estimates $\hat{\boldsymbol{\theta}}$.
2. Generate a bootstrap sample $\{\boldsymbol{Y_1^b}, ..., \boldsymbol{Y_T^b}\}$ using $\hat{\boldsymbol{\theta}}$, either parametrically or non-parametrically, as described before the beginning of this sub-section. If the model is non-stationary, the overall pattern of the bootstrap sample has to be brought in accordance with the pattern of the original sample, as described in the third and fourth paragraphs of the introduction to Section 3.
3. The bootstrap dataset $\{\boldsymbol{Y_1^b}, ..., \boldsymbol{Y_T^b}\}$ is used to obtain both the survey error autocorrelation parameter $\hat{\rho}^b$ estimates and bootstrap ML estimates $\hat{\boldsymbol{\theta}}_{\boldsymbol{ML}}^{\boldsymbol{b}}$. Thereafter, the Kalman filter is launched again using the original series $\{\boldsymbol{Y_1}, ..., \boldsymbol{Y_T}\}$ and the newly-estimated $\hat{\boldsymbol{\theta}}^{\boldsymbol{b}}$, which produces $\hat{\boldsymbol{\alpha}}_{\boldsymbol{t|t}}(\hat{\boldsymbol{\theta}}^{\boldsymbol{b}})$ and $\hat{\boldsymbol{P}}_{\boldsymbol{t|t}}(\hat{\boldsymbol{\theta}}^{\boldsymbol{b}})$.

4. Then, steps 2-3 are repeated B times.
5. Having obtained B bootstrap replicates, the true conditional MSE can be approximated in the following way:

$$\widehat{\boldsymbol{MSE}}_{t|t}^{RR} = \frac{1}{B}\sum_{b=1}^{B}\hat{\boldsymbol{P}}_{t|t}(\hat{\boldsymbol{\theta}}^{b}) + \frac{1}{B}\sum_{b=1}^{B}[\hat{\boldsymbol{\alpha}}_{t|t}(\hat{\boldsymbol{\theta}}^{b}) - \bar{\boldsymbol{\alpha}}_{t|t}][\hat{\boldsymbol{\alpha}}_{t|t}(\hat{\boldsymbol{\theta}}^{b}) - \bar{\boldsymbol{\alpha}}_{t|t}]', \tag{3.6}$$

where $\bar{\boldsymbol{\alpha}}_{t|t} = \frac{1}{B}\sum_{b=1}^{B}\hat{\boldsymbol{\alpha}}_{t|t}(\hat{\boldsymbol{\theta}}^{b})$.
Equation (3.6) is applied both for the parametric and non-parametric bootstrap MSE-estimators (denoted further as MSE^{RR1} and MSE^{RR2}, respectively).

3.2 Pfeffermann and Tiller Bootstrapping Approach

The bootstrap developed by Pfeffermann and Tiller (2005) differs from the one described in the previous subsection in that expectations of the squared error loss elements in (3.3) are taken unconditionally on the data, whereas the Rodriguez and Ruiz (2012) approach conditions on the original dataset $\{\boldsymbol{Y_1}, ..., \boldsymbol{Y_T}\}$. Further, unlike Rodriguez and Ruiz (2012), Pfeffermann and Tiller (2005) drop the terms that are of order $O\left(\frac{1}{T^2}\right)$, which is theoretically proven in Pfeffermann and Tiller (2005), Appendix C. Using results in Hall and Martin (1988), they show that the true MSE, being

$$\boldsymbol{MSE}_{t|t} = \hat{\boldsymbol{P}}_{t|t}(\boldsymbol{\theta}) + \boldsymbol{E}[(\hat{\boldsymbol{\alpha}}_{t|t}(\hat{\boldsymbol{\theta}}) - \hat{\boldsymbol{\alpha}}_{t|t}(\boldsymbol{\theta}))(\hat{\boldsymbol{\alpha}}_{t|t}(\hat{\boldsymbol{\theta}}) - \hat{\boldsymbol{\alpha}}_{t|t}(\boldsymbol{\theta}))'], \tag{3.7}$$

can be approximated by its bootstrap analogues as follows:

$$\hat{\boldsymbol{P}}_{t|t}(\boldsymbol{\theta}) = 2\hat{\boldsymbol{P}}_{t|t}(\hat{\boldsymbol{\theta}}) - \frac{1}{B}\sum_{b=1}^{B}\hat{\boldsymbol{P}}_{t|t}(\hat{\boldsymbol{\theta}}^{b}) + \boldsymbol{O}\left(\frac{1}{\boldsymbol{T^2}}\right), \tag{3.8}$$

$$\boldsymbol{E}[(\hat{\boldsymbol{\alpha}}_{t|t}(\hat{\boldsymbol{\theta}}) - \hat{\boldsymbol{\alpha}}_{t|t}(\boldsymbol{\theta}))(\hat{\boldsymbol{\alpha}}_{t|t}(\hat{\boldsymbol{\theta}}) - \hat{\boldsymbol{\alpha}}_{t|t}(\boldsymbol{\theta}))'] = \frac{1}{B}\sum_{b=1}^{B}[\hat{\boldsymbol{\alpha}}_{t|t}^{b}(\hat{\boldsymbol{\theta}}^{b}) - \hat{\boldsymbol{\alpha}}_{t|t}^{b}(\hat{\boldsymbol{\theta}})][\hat{\boldsymbol{\alpha}}_{t|t}^{b}(\hat{\boldsymbol{\theta}}^{b}) - \hat{\boldsymbol{\alpha}}_{t|t}^{b}(\hat{\boldsymbol{\theta}})]' + \boldsymbol{O}\left(\frac{1}{\boldsymbol{T^2}}\right). \tag{3.9}$$

Equations (3.8) and (3.9) correspond to the first and the second terms of equation (3.3), respectively. The resulting MSE-estimator below has a bias of correct order $O\left(\frac{1}{T^2}\right)$:

$$\widehat{\boldsymbol{MSE}}_{t|t}^{PT} = 2\hat{\boldsymbol{P}}_{t|t}(\hat{\boldsymbol{\theta}}) - \frac{1}{B}\sum_{b=1}^{B}\hat{\boldsymbol{P}}_{t|t}(\hat{\boldsymbol{\theta}}^{b}) + \frac{1}{B}\sum_{b=1}^{B}[\hat{\boldsymbol{\alpha}}_{t|t}^{b}(\hat{\boldsymbol{\theta}}^{b}) - \hat{\boldsymbol{\alpha}}_{t|t}^{b}(\hat{\boldsymbol{\theta}})][\hat{\boldsymbol{\alpha}}_{t|t}^{b}(\hat{\boldsymbol{\theta}}^{b}) - \hat{\boldsymbol{\alpha}}_{t|t}^{b}(\hat{\boldsymbol{\theta}})]', \tag{3.10}$$

Equation (3.10) is applied both for the parametric and non-parametric bootstrap MSE-estimators (denoted further as MSE^{PT1} and MSE^{PT2}, respectively). MSE-calculation in (3.10) requires three Kalman filter runs for every bootstrap series. One run is needed to get MSE-matrix $\hat{\boldsymbol{P}}_{t|t}(\hat{\boldsymbol{\theta}}^{b})$ based on the original data set $\{\boldsymbol{Y_1}, ..., \boldsymbol{Y_T}\}$ and bootstrap parameters $\hat{\boldsymbol{\theta}}^{b}$. In the second run, $\hat{\boldsymbol{\alpha}}_{t|t}^{b}(\hat{\boldsymbol{\theta}}^{b})$ is estimated from the bootstrap data set $\{\boldsymbol{Y_1^b}, ..., \boldsymbol{Y_T^b}\}$ and the same bootstrap parameters $\hat{\boldsymbol{\theta}}^{b}$. The third Kalman filter run is needed to produce the state estimates $\hat{\boldsymbol{\alpha}}_{t|t}^{b}(\hat{\boldsymbol{\theta}})$ based on data set $\{\boldsymbol{Y_1^b}, ..., \boldsymbol{Y_T^b}\}$ and $\hat{\boldsymbol{\theta}}$-estimates that were obtained from the original dataset.
The bootstrap procedure is summarized below:
1. Estimate the model using the original dataset and obtain the hyperparameter vector estimates $\hat{\boldsymbol{\theta}}$. Apart from that, save the "naive" MSE estimates $\hat{\boldsymbol{P}}_{t|t}(\hat{\boldsymbol{\theta}})$ for future use in (3.10).

2. Using either the parametric or non-parametric method, generate a bootstrap sample $\{\boldsymbol{Y_1^b}, ..., \boldsymbol{Y_T^b}\}$ conditional on the observed sample using the original ML estimates $\hat{\boldsymbol{\theta}}$, as described in Section 3.1.
3. Estimate bootstrap hyperparameter estimates $\hat{\boldsymbol{\theta}}^{\boldsymbol{b}}$ from the newly generated bootstrap dataset. Run the Kalman filter for three times to get $\hat{\boldsymbol{\alpha}}_{\boldsymbol{t|t}}^{\boldsymbol{b}}(\hat{\boldsymbol{\theta}}^{\boldsymbol{b}})$, $\hat{\boldsymbol{P}}_{\boldsymbol{t|t}}(\hat{\boldsymbol{\theta}}^{\boldsymbol{b}})$, $\hat{\boldsymbol{\alpha}}_{\boldsymbol{t|t}}^{\boldsymbol{b}}(\hat{\boldsymbol{\theta}})$, as described under (3.10).
4. Repeat steps 2-3 B times.
5. Approximate the true conditional MSE using (3.10).
Pfeffermann and Tiller (2005) note that, in the case of the parametric bootstrap, the third Kalman filter run can be avoided because the true state vector is generated (and thus known) for every bootstrap series. Thus, the state estimates $\hat{\boldsymbol{\alpha}}_{\boldsymbol{t|t}}^{\boldsymbol{b}}(\hat{\boldsymbol{\theta}})$ in (3.10) can be replaced by the true vector $\boldsymbol{\alpha_t^b}$ to obtain the following MSE estimator:

$$\widehat{\boldsymbol{MSE}}_{\boldsymbol{t|t}}^{\boldsymbol{PT1}} = \hat{\boldsymbol{P}}_{\boldsymbol{t|t}}(\hat{\boldsymbol{\theta}}) - \frac{\boldsymbol{1}}{\boldsymbol{B}}\sum_{\boldsymbol{b=1}}^{\boldsymbol{B}} \hat{\boldsymbol{P}}_{\boldsymbol{t|t}}(\hat{\boldsymbol{\theta}}^{\boldsymbol{b}}) + \frac{\boldsymbol{1}}{\boldsymbol{B}}\sum_{\boldsymbol{b=1}}^{\boldsymbol{B}} [\hat{\boldsymbol{\alpha}}_{\boldsymbol{t|t}}^{\boldsymbol{b}}(\hat{\boldsymbol{\theta}}^{\boldsymbol{b}}) - \boldsymbol{\alpha_t^b}][\hat{\boldsymbol{\alpha}}_{\boldsymbol{t|t}}^{\boldsymbol{b}}(\hat{\boldsymbol{\theta}}^{\boldsymbol{b}}) - \boldsymbol{\alpha_t^b}]'. \qquad (3.11)$$

In this formulation, there is only one $\hat{\boldsymbol{P}}_{\boldsymbol{t|t}}(\hat{\boldsymbol{\theta}})$ in the right-hand side of (3.11). This is due to the fact that the new term $E_B[\hat{\alpha}_{t|t}^b(\hat{\theta}^b) - \alpha_t^b]^2$, corresponding to the last term in the right-hand side of (3.11), can itself be decomposed, in the same fashion as in (3.7), into the parameter uncertainty $E_B[\hat{\alpha}_{t|t}^b(\hat{\theta}^b) - \hat{\alpha}_{t|t}^b(\hat{\theta})]^2$ and the filter uncertainty term $P_{t|t}^b(\hat{\theta}) = E[\hat{\alpha}_{t|t}^b(\hat{\theta}) - \alpha_t^b]^2$, $\hat{\theta}$ being the true parameter vector the bootstrap state variable α_t^b is generated with. However, the bootstrap average term $\frac{1}{B}\sum_{b=1}^{B}[\hat{\alpha}_{t|t}^b(\hat{\theta}) - \alpha_t^b]^2$ replacing $\hat{P}_{t|t}(\hat{\theta})$ may need much more bootstrap iterations to converge. Further, one should be aware of the fact that this simplified method may result in an additional bias if the normality assumption about the model error terms is violated. Then, the term $E_B[\hat{\alpha}_{t|t}^b(\hat{\theta}^b) - \alpha_t^b]^2$ will also contain a non-zero expectation of the cross-terms: $E\{[\hat{\alpha}_{t|t}^b(\hat{\theta}) - \alpha_t^b][\hat{\alpha}_{t|t}^b(\hat{\theta}^b) - \hat{\alpha}_{t|t}^b(\hat{\theta})]\}$. In this application, the influence of non-zero cross-term bootstrap averages has turned out to be of a negligible importance, but the bootstrap average $\frac{\boldsymbol{1}}{\boldsymbol{B}}\sum_{\boldsymbol{b=1}}^{\boldsymbol{B}}[\hat{\boldsymbol{\alpha}}_{\boldsymbol{t|t}}^{\boldsymbol{b}}(\hat{\boldsymbol{\theta}}) - \boldsymbol{\alpha_t^b}][\hat{\boldsymbol{\alpha}}_{\boldsymbol{t|t}}^{\boldsymbol{b}}(\hat{\boldsymbol{\theta}}) - \boldsymbol{\alpha_t^b}]'$ exhibited large deviations from the term it was meant to replace. This may be explained by the fact that the true Kalman filter MSE in (3.1) can be obtained from simulated series if the distribution of the state-vector is sufficiently dispersed. When bootstrapping non-stationary models, however, the bootstrap series are forced to follow the pattern of the underlying original series, as it has been mentioned in the description of the simulation smoother algorithm. Therefore, the replacement of $\hat{\boldsymbol{P}}_{\boldsymbol{t|t}}(\hat{\boldsymbol{\theta}})$ with the term $\frac{\boldsymbol{1}}{\boldsymbol{B}}\sum_{\boldsymbol{b=1}}^{\boldsymbol{B}}[\hat{\boldsymbol{\alpha}}_{\boldsymbol{t|t}}^{\boldsymbol{b}}(\hat{\boldsymbol{\theta}}) - \boldsymbol{\alpha_t^b}][\hat{\boldsymbol{\alpha}}_{\boldsymbol{t|t}}^{\boldsymbol{b}}(\hat{\boldsymbol{\theta}}) - \boldsymbol{\alpha_t^b}]'$ in (3.11) is not equivalent. For this reason, both parametric (denoted as PT1) and non-parametric (PT2) bootstraps in this application rely on the estimator in (3.10).

3.3 Asymptotic Approximation

An asymptotic approximation (AA) to the true MSE in equation (3.3) was developed by Hamilton (1986) and can be expressed as an expectation over the hyperparameter joint asymptotic distribution $\pi_a(\boldsymbol{\theta}|\boldsymbol{Y})$ conditional on the given dataset $\boldsymbol{Y} \equiv \{\boldsymbol{Y_1}, ..., \boldsymbol{Y_T}\}$. In this application, a part of the hyperparameter vector that is estimated with the ML-method (denoted as $\boldsymbol{\theta_{ML}}$), depends on the value of the autoregressive parameter ρ. Therefore, the hyperparameter joint distribution has the following form: $\pi(\boldsymbol{\theta}|\boldsymbol{Y}) = \pi(\rho|\boldsymbol{Y})\pi(\boldsymbol{\theta_{ML}}|\rho, \boldsymbol{Y})$. The MSE is approximated as follows:

$$\boldsymbol{MSE_{t|t}} = \boldsymbol{E_{\pi(\theta|Y)}}[\hat{\boldsymbol{P}}_{\boldsymbol{t|t}}(\boldsymbol{\theta})] + \boldsymbol{E_{\pi(\theta|Y)}}\left\{\boldsymbol{E_t}[(\hat{\boldsymbol{\alpha}}_{\boldsymbol{t|t}}(\boldsymbol{\theta}) - \hat{\boldsymbol{\alpha}}_{\boldsymbol{t|t}}(\hat{\boldsymbol{\theta}}))(\hat{\boldsymbol{\alpha}}_{\boldsymbol{t|t}}(\boldsymbol{\theta}) - \hat{\boldsymbol{\alpha}}_{\boldsymbol{t|t}}(\hat{\boldsymbol{\theta}}))']\right\}. \quad (3.12)$$

Random ρ-realisations are drawn from their asymptotic distribution $N(\hat{\rho}, 1/\sqrt{T})$, where T is the series length. After a value is drawn from this distribution, the other hyperparameters are re-estimated to obtain $\hat{\boldsymbol{\theta}}_{\boldsymbol{ML}}|\rho$ and the information matrix $\hat{\boldsymbol{I}}(\hat{\boldsymbol{\theta}}_{\boldsymbol{ML}}|\rho_{\pi(\rho|\boldsymbol{Y})})$. Finally, a $\boldsymbol{\theta}_{\boldsymbol{ML}}$-draw is made from distribution $\sqrt{T}(\boldsymbol{\theta}_{\boldsymbol{ML}} - \hat{\boldsymbol{\theta}}_{\boldsymbol{ML}}|\rho_{\pi(\rho|\boldsymbol{Y})}) \sim MN(\boldsymbol{0}, T\hat{\boldsymbol{I}}^{-\boldsymbol{1}}(\hat{\boldsymbol{\theta}}_{\boldsymbol{ML}}|\rho_{\pi(\rho|\boldsymbol{Y})}))$. The Kalman filter is run again using ρ- and $\boldsymbol{\theta}_{\boldsymbol{ML}}$-realisations to obtain the state estimates $\hat{\boldsymbol{\alpha}}_{\boldsymbol{t|t}}(\boldsymbol{\theta}_{\boldsymbol{\pi(\theta|Y)}})$ and their MSEs $\hat{\boldsymbol{P}}_{\boldsymbol{t|t}}(\boldsymbol{\theta}_{\boldsymbol{\pi(\theta|Y)}})$. The procedure is repeated until B $\boldsymbol{\theta}_{\boldsymbol{\pi(\theta|Y)}}$-draws are obtained, whereafter (3.12) is approximated by averaging the necessary quantities over B iterations. If all the hyperparameters of the model are estimated within the ML-procedure, B draws can be made directly from $\sqrt{T}(\boldsymbol{\theta}_{\boldsymbol{\pi(\theta|Y)}} - \hat{\boldsymbol{\theta}}) \sim MN(\boldsymbol{0}, T\hat{\boldsymbol{I}}^{-\boldsymbol{1}}(\hat{\boldsymbol{\theta}}))$.

The first term in (3.12) can be approximated by the expected value of the Kalman filter variance $\hat{\boldsymbol{P}}_{\boldsymbol{t|t}}$ across B realizations of the hyperparameter vector, and the second term by the sampling variance of the state vector expectation. By virtue of the continuous mapping theorem, the sample average $\bar{\hat{\boldsymbol{\alpha}}}_{\boldsymbol{t|t}} = \frac{\boldsymbol{1}}{\boldsymbol{B}}\sum_{\boldsymbol{b=1}}^{\boldsymbol{B}} \hat{\boldsymbol{\alpha}}_{\boldsymbol{t|t}}(\boldsymbol{\theta}^{\boldsymbol{b}}_{\boldsymbol{\pi(\theta|Y)}})$ can replace the state vector estimates $\hat{\boldsymbol{\alpha}}_{\boldsymbol{t|t}}(\hat{\boldsymbol{\theta}})$ in (3.12) that are based on the initially estimated hyperparameter vector. An asymptotic approximation to the true MSE could therefore be obtained in the following way:

$$\hat{\boldsymbol{MSE}}^{\boldsymbol{AA}}_{\boldsymbol{t|t}} = \frac{\boldsymbol{1}}{\boldsymbol{B}}\sum_{\boldsymbol{b=1}}^{\boldsymbol{B}} \hat{\boldsymbol{P}}_{\boldsymbol{t|t}}(\boldsymbol{\theta}^{\boldsymbol{b}}_{\boldsymbol{\pi(\theta|Y)}}) + \frac{\boldsymbol{1}}{\boldsymbol{B}}\sum_{\boldsymbol{b=1}}^{\boldsymbol{B}} [\hat{\boldsymbol{\alpha}}_{\boldsymbol{t|t}}(\boldsymbol{\theta}^{\boldsymbol{b}}_{\boldsymbol{\pi(\theta|Y)}}) - \bar{\hat{\boldsymbol{\alpha}}}_{\boldsymbol{t|t}}][\hat{\boldsymbol{\alpha}}_{\boldsymbol{t|t}}(\boldsymbol{\theta}^{\boldsymbol{b}}_{\boldsymbol{\pi(\theta|Y)}}) - \bar{\hat{\boldsymbol{\alpha}}}_{\boldsymbol{t|t}}]'. \quad (3.13)$$

Obviously, this MSE-estimator is entirely based on the asymptotic normality assumption about the hyperparameter vector estimator. Apart from that, this approach usually produces significant biases if the series is not of a sufficient length, in which case the asymptotic distribution would fail to approximate the finite (usually skewed) distribution of maximum-likelihood estimates.

Another problem with the asymptotic approach can appear if some of the hyperparameters are estimated to be close to zero. This can happen to the initial model estimates or during the procedure itself, e.g., due to certain extreme ρ-draws. In these cases, the asymptotic variance of such hyperparameters will be too large, which will inflate the MSE-estimates of the signal and its unobserved components. It may as well lead to a failure in inverting the information matrix for the hyperparameter vector.

4 The DLFS-specific Simulation and Bootstrap Setup for True MSE Approximation

The central data generating process of the present simulation study is the DLFS model. The performance of the five MSE-approximation methods is examined on series of the original length from the DLFS survey (114 monthly time points from 2001(1) until 2010(6)), as well as on shorter series of length 80 months and longer ones of length 200. For each of these series lengths, a Monte-Carlo experiment is set up where multiple series (1000) are simulated on the basis of the LFS model used by Statistics Netherlands. MSEs for each of these series are approximated based on $B = 300$ bootstrap series; for asymptotic approximation, however, at

least $B = 500$ draws turned out to be needed. This number has been found sufficient for the approximated MSEs to converge. MSEs from the five approximation methods averaged over the 1000 simulations are compared to MSE-averages produced by the "naive" Kalman filter. However, for the latter MSE estimates to converge to a certain average value, far more than 1000 simulations turned out to be needed (namely, at least 10000).

The hyperparameter maximum-likelihood estimates of the DLFS model, as well as the Yule-Walker estimate of the survey error autoregressive parameter ρ, are used to generate artificial data series $\hat{\boldsymbol{Y}}_{\boldsymbol{t}}^{\boldsymbol{s}}, s = 1, \ldots S$. Since the system is non-stationary, generating series $\boldsymbol{\alpha}_{\boldsymbol{t}}^{\boldsymbol{s}}$ unconditionally on the true data is going to result in negative or implausibly large numbers of the unemployed. In order to avoid an excessively large number of series with negative values, the unconditional recursion of the state variables is started with their smoothed estimates at one of the highest points the original process has ever reached within the available sample. Further, the first r time points are discarded in order to prevent that the series start at the same time-point (r being chosen equal to 30). With an assumption that unemployment in the Netherlands will not exceed 15 percent, the simulation data set is restricted to positive series below the upper bound of 1 mln of unemployed (this value comprised about 15 percent of the Dutch labour force in 2010). Keeping the artificial series below the upper bound is also done in order not to extrapolate outside of the original data range when simulating the design standard errors z_t^{t-j}, which are the time-dependent elements of the design-matrix $\boldsymbol{Z}_{\boldsymbol{t}}$.

Since design variance estimates depend on the number of unemployed estimated by the GREG-estimator for the wave in question, a model is needed to generate z_t^{t-j}-values that would depend on the corresponding simulated numbers of the unemployed. The well-known formula for the variance of the population total of a binary response variable is helpful to derive the necessary generation process for the z_t^{t-j}-terms (see, e.g., Särndal et al. (1992), Section 3.3):

$$Var(\hat{Y}_t) = N_t^2 \left(1 - \frac{n_t}{N_t}\right) \frac{p_t(1-p_t)}{n_t}, \tag{4.1}$$

where N_t is the population size in period t, n_t is the sample size, and p_t is the sample estimate of the unemployment rate $\frac{\hat{Y}_t}{N_t}$. The design variance can then be approximated by the following expression, after taking logs on both sides of (4.1), rearranging the terms and assuming that unemployment p_t does not reach high values, which allows to neglect the term $ln(1-p_t)$:

$$ln(Var(\hat{Y}_t)) = \alpha ln\left(\frac{n_t}{N_t}\right) + \beta ln(\hat{Y}_t). \tag{4.2}$$

This study simulates numbers of unemployed conditional on the same population and sample sizes as those observed in reality between 2001(1) and 2010(6). This allows to avoid simulating sample sizes in a study for longer series ($T = 200$). Instead, the information carried by the term $ln\left(\frac{n_t}{N_t}\right)$ can be represented by an intercept for the variances of the first wave. Starting from wave 2, each design variance, denoted as $(z_t^{t-j})^2, j = \{3, 6, 9, 12\}$, is highly dependent on the design variance of the preceding wave, since both are based on nearly the same group of people and sample size. The sample size decreases by approximately 10 percent in each subsequent wave due to panel attrition. Further, the signal $l_t^{t-j} = L_t + \gamma_t + \lambda_t^{t-j}$ can act as a proxy for $\hat{Y}_t^{t-j}$ in (4.2), since both the signal and the direct estimate contain information on the level of unemployed and the design effect (in this case, the RGB). The design standard errors z_t^{t-j} can be derived from the following equations for log-variances:

$$\begin{aligned} &ln[(z_t^{t-j})^2] = c + \beta^j ln(l_t^{t-j}) + \epsilon_t^{t-j}, j = 0; \\ &ln[(z_t^{t-j})^2] = \rho^j ln[(z_{t-3}^{t-j})^2] + \beta^j ln(l_t^{t-j}) + \epsilon_t^{t-j}, \epsilon_t^{t-j} \sim N(0, (\sigma_\epsilon^j)^2), j = \{3, 6, 9, 12\}. \end{aligned} \tag{4.3}$$

The regression coefficients in (4.3) are time-invariant. The superscripts are used to denote the wave these coefficients belong to. The coefficient estimates are presented in Table 1. They are obtained using the original design variances for the five panel waves and the extracted filtered signal estimates coming from the DLFS model. For series longer than the original series, the design standard errors $\boldsymbol{z_t^s}$ have to be simulated as well according to the same process.

Table 1 Regression estimates for the design standard error process

	$j=0$	$j=3$	$j=6$	$j=9$	$j=12$
$\hat{c}$	12.219	-	-	-	-
$\hat{\beta}^j$	0.630	0.244	0.354	0.414	0.413
$\hat{\rho}^j$	-	0.859	0.786	0.749	0.751
$\hat{\sigma}_\epsilon^j$	0.202	0.265	0.228	0.225	0.267

Equations (2.6)-(2.7) and the estimated parameter vector $\hat{\boldsymbol{\theta}}$ are used to generate S=1000 series of artificial data. State disturbances (remember survey errors are also modelled as state variables) are randomly drawn from their joint normal distribution $N(\mathbf{0}, \boldsymbol{\Omega}(\hat{\boldsymbol{\theta}}))$, and series are generated unconditionally on the true data. Within each simulation, first the trend, seasonal and RGB components are simulated and summed up to comprise the wave-signals $l_{t,s}^{t-j}, j = \{0, 3, 6, 9, 12\}$. These are used to generate the design standard errors $z_{t,s}^{t-j}$ according to the process described in (4.3). As soon as an artificial data set is generated, ρ is re-estimated and saved as $\hat{\rho}_s$, whereafter the hyperparameter quasi-ML estimates are obtained. These are stored in $\hat{\boldsymbol{\theta}}_{\boldsymbol{s}}$ and used by the Kalman filter to produce the state estimates $\hat{\boldsymbol{\alpha}}_{\boldsymbol{t,s}}$. Both $\hat{\rho}_s$ and $\hat{\boldsymbol{\theta}}_{\boldsymbol{s}}$ are used to generate bootstrap samples. Note that the same set of design standard errors $\boldsymbol{z_{t,s}}$ is used to both generate and estimate bootstrap series within simulation s.
In order to obtain the true MSEs, the DLFS model is simulated a large number of times $M = 50000$, each of these replications being restricted to the same limits as before, i.e. between zero and 1 mln unemployed. The true MSE is calculated in the following way using the true state vector $\boldsymbol{\alpha_{m,t}}$ values known for every simulation m:

$$\boldsymbol{MSE_t^{true}} = \frac{\mathbf{1}}{\boldsymbol{M}} \sum_{\boldsymbol{m=1}}^{\boldsymbol{M}} [(\hat{\boldsymbol{\alpha}}_{\boldsymbol{m,t}}(\hat{\boldsymbol{\theta}}_{\boldsymbol{m}}) - \boldsymbol{\alpha_{m,t}})(\hat{\boldsymbol{\alpha}}_{\boldsymbol{m,t}}(\hat{\boldsymbol{\theta}}_{\boldsymbol{m}}) - \boldsymbol{\alpha_{m,t}})']. \tag{4.4}$$

The true MSE of the signal is calculated in the same way by using the true wave-signal values $\boldsymbol{l_{m,t}}$.

5 MSE Approximation Approaches with Application to the DLFS

The focus of this simulation study is the true MSEs of the trend and of the population signal. The latter consists of the trend and seasonal components and is therefore equal to the signal of the first wave. This study considers four models that differ in terms of the number of hyperparameters to be estimated with the ML method. The most complete model - Model 1 - is the one currently in use at Statistics Netherlands, but with the white noise component ε_t removed from the true population parameter ξ_t. This component has turned out to have an insignificant variance in the case of the DLFS. Moreover, an attempt to estimate its variance

disturbed estimation of other marginally significant hyperparameters. In order to avoid this instability, this irregular component ε_t can be removed from the model. This formulation implies that the population parameter ξ_t does not suffer from any unusual irregularities that cannot be picked up by the stochastic structure of the trend and seasonal components. This assumption can be advocated by a relative rigidity of labour markets. Any alterations of unemployment levels are usually of a gradual character and therefore must be largely incorporated into the stochastic trend movements. The other three models are special cases of Model 1, thus with ε_t component removed (see Table 2).

Table 2 Hyperparameters estimated in the four versions of the DLFS model; the disturbance variances estimated on a log-scale

Models	Description	Parameters estimated
M1	complete model	$\rho, \sigma_R^2, \sigma_\gamma^2, \sigma_\lambda^2, \sigma_{\nu_1}^2, \sigma_{\nu_2}^2, \sigma_{\nu_3}^2, \sigma_{\nu_4}^2, \sigma_{\nu_5}^2$
M2	seasonal time-independent	$\rho, \sigma_R^2, \sigma_\lambda^2, \sigma_{\nu_1}^2, \sigma_{\nu_2}^2, \sigma_{\nu_3}^2, \sigma_{\nu_4}^2, \sigma_{\nu_5}^2$
M3	RGB time-independent	$\rho, \sigma_R^2, \sigma_\gamma^2, \sigma_{\nu_1}^2, \sigma_{\nu_2}^2, \sigma_{\nu_3}^2, \sigma_{\nu_4}^2, \sigma_{\nu_5}^2$
M4	seasonal, RGB fixed	$\rho, \sigma_R^2, \sigma_{\nu_1}^2, \sigma_{\nu_2}^2, \sigma_{\nu_3}^2, \sigma_{\nu_4}^2, \sigma_{\nu_5}^2$

Note that the disturbance variances are estimated on a log-scale in order to avoid negative estimates. The rationale behind studying the other three models becomes clear after inspecting the hyperparameter distribution of Model 1 after a large number of replications. The simulation has shown that the stochastic term variances of the seasonal and, in particular, RGB components are often estimated to be close to zero. This causes bi-modality in the distribution of these variance estimates with a significant mass concentrated around zero. Apart from that, an attempt to estimate both of the hyperparameters, as in Model 1, seems to bring about certain instability to the hyperparameter estimates, such that even normality in $ln(\hat{\sigma}_{\nu_3}^2)$, $ln(\hat{\sigma}_{\nu_4}^2)$ and $ln(\hat{\sigma}_{\nu_5}^2)$ (where indices 3, 4, and 5 stand for the numbers of the waves) is severely violated with extreme outliers and/or a huge kurtosis (see Fig. B.1 in the appendix, where the x-axis is extended due to the outliers). Making the seasonal component time-invariant, as in Model 2, hardly changes the situation for the slope and RGB hyperparameters. Instead, it may even be seen as suboptimal due to more extreme outliers and excess kurtosis in the distribution of all the five survey error hyperparameters (Fig. B.2). By contrast, under both models where the RGB-component is fixed over time (Models 3 and 4), all hyperparameters corresponding to the survey error component have turned out to be normally distributed, see Fig. B.3 and Fig. B.4. Under Model 3, distributions are still skewed for the slope and seasonal components (skewness of -0.88 and -0.72, and excess kurtosis of 2.56 and 1.61, respectively). Fixing σ_γ^2 to zero under Model 4 results in only a marginal improvement: the distribution of $ln(\hat{\sigma}_R^2)$ is negatively skewed (-0.81) with an excess kurtosis of 1.76.

This simulation evidence suggests that the preference in modelling the DLFS series may be given to the more parsimonious Model 3, where only the RGB disturbance variance is set equal to zero. This hyperparameter is however retained for production purposes at Statistics Netherlands to secure the model robustness against sudden changes in the underlying process. The distribution of the survey error autoregressive parameter ρ is hardly affected by model reformulations and ranges between 0 and 40 percent. The simulation procedure described in the previous section and the analysis of bootstrap methods that follows is performed separately for all the four models.

The performance of the Kalman filter at estimated parameter values, as well as of the five approximation methods mentioned in Section 3 is evaluated with the help of the MSE relative bias. First, the approximated MSEs from (3.13), (3.6), (3.10), and (3.11) are averaged over 1000 simulations, and the Kalman filter MSE estimates over 10000 simulations, as mentioned at the beginning of Section 4. These averaged MSE estimates for Model 3 (except for AA for the reason that will become clear soon) are depicted in Fig. 5.1, 5.2 and 5.3 for three different series

lengths, skipping the first $d = 30$ time points of the sample. This is the time needed for the diffuse part of the state covariance matrix to decay (see Koopman (1997) for initialisation of non-stationary state variables). The percentage relative bias is calculated as $RB_t^b = \left(\frac{\overline{MSE}_{t|t}^b}{MSE_{t|t}^{true}} - 1 \right) \cdot 100\%$, where b defines a particular approximation method. The core variables of interest for users of the DLFS are the signal and trend. The percentage relative MSE biases averaged over time (skipping the first $d = 30$ time points) for the signal, the trend and seasonal components are presented in Tables 3, 4, and 5.

The AA-method turned out to be inapplicable to the models with marginally significant hyperparameters. When some of the hyperparameters are estimated close to zero, the matrix $\boldsymbol{I^{-1}}(\hat{\boldsymbol{\theta}}_{\boldsymbol{ML}}|\rho_{\pi(\rho|\boldsymbol{Y})})$ is numerically either singular, leading to a failure in the procedure, or nearly singular. In the latter case, the asymptotic variance becomes excessively large and thus not reliable. Taking this into account, the AA-method could only be considered for Model 4. As expected, the method performs poorly in short series, with positive biases of about 15 percent. The performance for $T = 114$ and $T = 200$ is comparable to that of the PT1-bootstrap, but significantly worse than the PT2 performance.

The simulation results for $T = 80$ suggest that, when averaged over time (starting from $t = 30$), the relative bias of the signal MSE obtained with the Kalman filter ranges between -3.2 and -2.1 percent for the four models considered in the paper. This bias tends to decrease as the series length increases. The KF-biases are quite small for the case of $T = 200$, such that none of the approximation methods offers a smaller bias in absolute terms. One could still apply the best approximation method with positive biases in order to get a range of values containing the true MSE.

What one immediately sees is negative biases for the RR-bootstrap and positive ones for the PT-method. Against the claim of Rodriguez and Ruiz (2012) that their approach has better finite sample properties compared to the approach of Pfeffermann and Tiller (2005), the case of the DLFS suggests that the RR-estimates, both parametric and non-parametric ones, are even more negatively biased than the uncorrected KF-estimates across all the models and series lengths (except for RR2 in Model 1, $T = 80$ and $T = 114$). The PT-methods never produce negative biases. While the PT-bootstrap is proven to have satisfactory asymptotic properties in Pfeffermann and Tiller (2005), Rodriguez and Ruiz (2012) illustrate the superiority of their method in small samples based on a simple model (a random walk plus noise). The present simulation study reveals that the RR-method may not behave well in more complex applications.

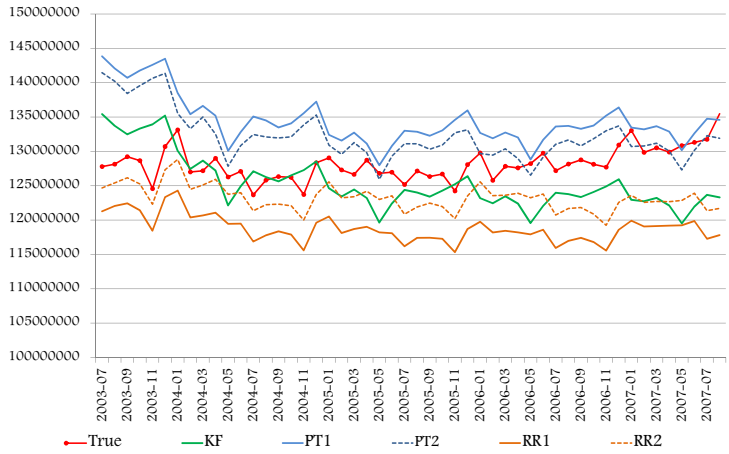

Figure 5.1 Signal MSE comparison for Model 3, T=80 months

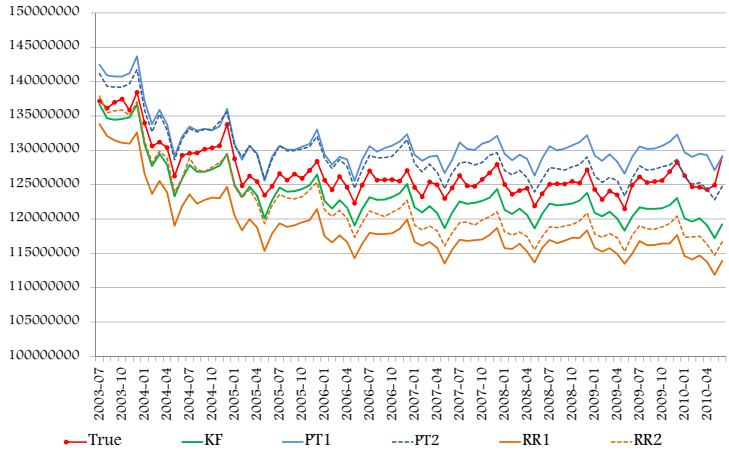

Figure 5.2 Signal MSE comparison for Model 3, T=114 months

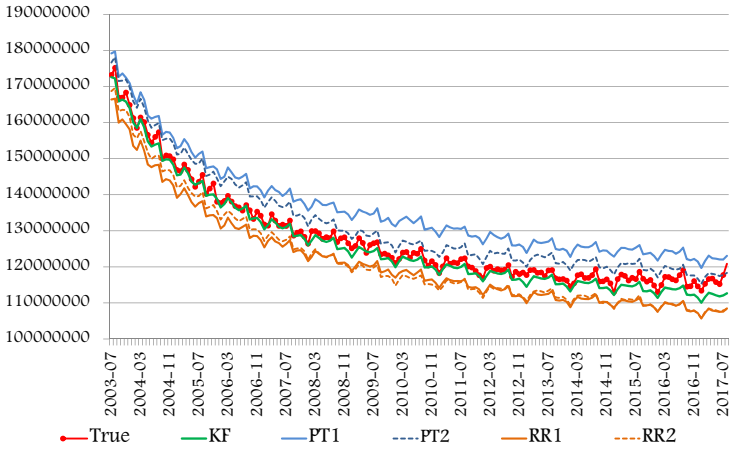

Figure 5.3 Signal MSE comparison for Model 3, T=200 months

For series of lengths $T = 114$ and $T = 80$, positive biases produced by the PT2-bootstrap may slightly exceed the KF-biases in absolute values for models with insignificant hyperparameters (Models 1 and 2). In the more stable models (Models 3 and 4), the positive biases are smaller than the KF negative biases in absolute values.

Table 3 MSE relative bias averaged over time (d=30) in the DLFS model, percent, T=80

	Signal				Trend				Seasonal			
Models	M1	M2	M3	M4	M1	M2	M3	M4	M1	M2	M3	M4
KF	-3.0	-3.2	-2.1	-2.2	-3.5	-3.8	-2.5	-2.5	8.8	2.5	2.9	2.4
AA	N/A	N/A	N/A	14.9	N/A	N/A	N/A	15.0	N/A	N/A	N/A	14.9
PT1	8.6	6.7	4.9	6.2	10.6	8.9	7.1	8.4	20.8	10.7	10.3	11.1
PT2	**4.8**	**3.7**	**1.4**	**2.1**	4.8	4.9	2.1	2.3	17.3	8.2	6.9	7.1
RR1	-7.2	-9.0	-7.3	-7.2	-9.6	-11.2	-9.6	-9.5	-3.8	-9.0	-6.7	-6.6
RR2	6.7	-3.5	-3.9	-3.7	5.3	-4.1	-4.6	-4.3	18.6	-4.7	-4.1	-4.8

Table 4 MSE relative bias averaged over time (d=30) in the DLFS model, percent, T=114

	Signal				Trend				Seasonal			
Models	M1	M2	M3	M4	M1	M2	M3	M4	M1	M2	M3	M4
KF	-2.1	-2.6	-2.4	-2.2	-2.3	-2.7	-2.4	-2.3	2.5	-3.2	-3.1	-2.6
AA	N/A	N/A	N/A	5.2	N/A	N/A	N/A	4.1	N/A	N/A	N/A	12.5
PT1	8.1	5.7	3.3	5.5	10.0	7.9	5.2	7.6	4.9	1.4	1.4	0.3
PT2	**2.2**	**3.2**	**1.9**	**1.5**	3.3	4.3	3.1	2.8	1.2	-2.0	1.0	0.6
RR1	-8.3	-7.8	-6.4	-6.5	-10.7	-9.9	-8.7	-8.9	-3.1	-7.2	-5.5	-5.6
RR2	-1.1	-6.0	-3.9	-3.5	-3.0	-7.6	-5.5	-5.0	7.3	-5.9	-3.2	-3.0

Table 5 MSE relative bias averaged over time (d=30) in the DLFS model, percent, T=200

	Signal				Trend				Seasonal			
Models	M1	M2	M3	M4	M1	M2	M3	M4	M1	M2	M3	M4
KF	-1.3	-1.6	-1.3	-1.3	-1.7	-1.8	-1.6	-1.6	3.8	-1.7	-1.6	-1.6
AA	N/A	N/A	N/A	5.9	N/A	N/A	N/A	5.6	N/A	N/A	N/A	5.6
PT1	**6.3**	6.2	6.3	5.5	7.5	7.7	7.8	7.1	10.8	2.6	3.0	3.0
PT2	6.8	**4.0**	**3.0**	**2.3**	7.6	4.9	4.2	3.6	12.5	2.1	1.3	0.6
RR1	-8.0	-8.0	-4.9	-5.9	-10.0	-9.9	-6.8	-7.1	-1.1	-5.3	-3.8	-3.9
RR2	-5.1	-5.6	-4.5	-5.0	-7.0	-7.4	-6.0	-6.4	3.6	-3.1	-3.3	-3.9

The signal MSE of Model 3, which could be considered a better option for production of official DLFS figures, is best approximated by the PT2 approach, with relative biases of 1.4 and 1.9 percent for $T = 80$ and $T = 114$, respectively. The PT2-bootstrap also seems to be the best method for $T = 200$, but, as has been said, the negative KF biases are already quite small for series of this length.

Note that for both the PT- and RR-bootstraps, the absolute values of relative biases are smaller in the case of the non-parametric approaches, compared to their parametric counterparts. The superiority of the non-parametric approach over the parametric one can be explained by the disturbed normality of the error distribution in the models.

In order to see if the STS model-based approach still offers some reduction in the design variance estimates after correcting for hyperparameter uncertainty, percentage reductions in the standard errors of the GREG estimates are presented in Table 6. These reductions come from applying the DLFS model to the GREG estimates without correcting for hyperparameter uncertainty (KF), as well as after this uncertainty has been taken into account with the help of the five different methods. Note that the RGB and seasonal hyperparameter estimates obtained

from the original DLFS data set are quite small. Therefore, there are no noticeable differences between the point-estimates of the four models. However, these hyperparameter estimates are not small enough to cause the problems mentioned before with regard to the AA-approach, so the results for this approximation method are reported in Table 6 as well. It may first seem that the AA-method accounts for the parameter uncertainty best of all other methods. However, keeping in mind that the AA may on average have some severe positive biases, especially in the case of small hyperparameters, one should feel more secure with the PT2 non-parametric approach that offers about 22 percent reduction in the estimated GREG standard errors. This means that the model-based approach even after accounting for parameter uncertainty offers a significant variance reduction compared with the traditional design-based approach.

Table 6 Reductions in the GREG SE estimates of the DLFS, averaged over time (d=30), and percentage increase in the model SE after approximation in ( )

	Model 1	Model 2	Model 3	Model 4
KF	-24.1	-24.1	-24.5	-24.5
AA	-18.8 (6.9)	-19.0 (6.7)	-19.1 (7.1)	-19.5 (6.6)
PT1	-20.1 (5.2)	-20.1 (5.2)	-21.1 (4.6)	-21.2 (4.4)
PT2	-22.9 (1.6)	-21.2 (3.8)	-22.2 (3.1)	-22.5 (2.6)
RR1	-26.5 (-3.2)	-26.6 (-3.4)	-26.5 (-2.7)	-26.5 (-2.7)
RR2	-24.0 (-0.1)	-25.4 (-1.8)	-25.6 (-1.4)	-25.7 (-1.6)

6 Conclusion

Most applications of small area estimation procedures in the literature are based on multilevel models. In the framework of this approach, it is common practice to account for the hyperparameter uncertainty in estimated MSEs. The literature on structural time series models applied in the context of SAE is still rather limited, with most applications ignoring the hyperparameter uncertainty when computing the MSEs of small area predictions. This renders MSE estimates negatively biased when series are not long enough, which may be a serious issue when it comes to such important economic indicators as unemployment.

The literature offers several procedures to correct for the negative bias in the MSE estimates produced by STS models. The present work aimed at establishing the best approximation approach to the true MSE of a small area estimation approach applied to the DLFS for official production of estimated numbers of the unemployed in the Netherlands.

A simulation study conducted for this purpose reveals that the asymptotic approximation is not applicable to cases with hyperparameters close to zero due to failures when inverting the information matrix of the hyperparameter estimates. The simulation results suggest that the non-parametric bootstraps, being free of normality assumptions about the error distribution, perform better than their parametric counterparts in both Pfeffermann and Tiller (2005) and Rodriguez and Ruiz (2012) methods. A more important finding, however, is that the Pfeffermann and Tiller (2005) bootstrap approaches with their positive biases consistently outperform the respective approaches of Rodriguez and Ruiz (2012), where the biases are generally negative and larger than those of the Kalman filter in absolute terms. This is contrary to the claim of Rodriguez and Ruiz (2012) about the superiority of their method in short time series. Apparently, their findings are purely heuristic and are based on a simple model simulation (random walk plus noise), while Pfeffermann and Tiller (2005) prove that their bootstrapping approach produces MSE estimates with a bias of a correct order.

Another result of this simulation study has revealed that it might be worth considering a more restricted version of the DLFS model, with the variance of the RGB component and of the

population parameter noise set equal to zero. For this model, the relative bias of the signal MSE produced by the Kalman filter can be reduced from about -2.4 to 1.9 percent with the non-parametric Pfeffermann and Tiller (2005) bootstrap approach. Even with this slightly positive bias, the standard errors of the GREG estimates are reduced by about 22 percent. The computation time for bootstrapping the original DLFS model according to the non-parametric Pfeffermann and Tiller (2005) procedure is about 1 hour for 300 bootstrap iterations. For the DLFS application, the bias in the Kalman filter MSE estimates is relatively small, therefore it may be deemed sufficient to rely on these naive MSE estimates for publication purposes.

Acknowledgements

The authors thank Harm-Jan Boonstra and Frank Pijpers (Statistics Netherlands) for thorough reading and valuable comments on a draft of this paper.

References

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

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

Cochran, W. (1977). *Sampling techniques*. New York, Wiley and Sons.

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

Durbin, J. and Koopman, S. (2002). A simple and efficient simulation smoother for state space time series analysis. *Biometrika* 89, 603–615.

Hall, P. and Martin, M. (1988). On bootstrap resampling and iteration. *Biometrika* 75, 661–671.

Hamilton, J. (1986). A standard error for the estimated state vector of a state-space model. *Journal of Econometrics* 33, 387–397.

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

Koopman, S., Harvey, A., Doornik, J., and Shephard, A. (2008). *STAMP: Structural Time Series Analyser, Modeller and Predictor*. Timberlake Consultants Press, London.

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

Koopman, S. J. and Durbin, J. (2000). Fast filtering and smoothing for multivariate state space models. *Journal of Time Series Analysis* 21, 281–296.

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

Kumar, S. and Lee, H. (1983). Evaluation of composite estimation for the Canadian labour force survey. *Survey Methodology* 9, 178–201.

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

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

Pfeffermann, D., Feder, M., and Signorelli, D. (1998). Estimation of autocorrelations of survey errors with application to trend estimation in small areas. *Journal of Business and Economic Statistics* 16, 339–348.

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

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

Rodriguez, A. and Ruiz, E. (2012). Bootstrap prediction mean squared errors of unobserved states based on the Kalman filter with estimated parameters. *Computational Statistics and Data Analysis* 56, 62–74.

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

Tiller, R. (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. and Krieg, S. (2009a). Estimation of the monthly unemployment rate through structural time series modelling in a rotating panel design. *Survey Methodology* 16, 177–190.

Van den Brakel, J. and Krieg, S. (2009b). Structural time series modelling of the monthly unemployment rate in a rotating panel design. *Discussion paper (09031)*.

Appendices

A Some Details on the State-Space Form of the DLFS Model

The measurement equation design matrix is a composite of a four matrices:
$\boldsymbol{Z_t} = (\boldsymbol{Z^{\xi}}\ \boldsymbol{Z^{\lambda}}\ \boldsymbol{Z_t^e}\ \boldsymbol{O_{5\times 8}})$, where $\boldsymbol{O_{5\times 8}}$ denotes a null-matrix of a dimension specified in the subscript,
$\boldsymbol{Z_t^{\xi}} = [\mathbf{1_5} \otimes (1\ 0\ 1\ 0\ 1\ 0\ 1\ 0\ 1\ 0\ 1\ 0\ 1)]$ selects the level L_t and the six seasonal harmonics $\gamma_{t,1}, \ldots \gamma_{t,6}$ for each of the five waves,

$\boldsymbol{Z_t^\lambda} = \begin{pmatrix} \mathbf{0_4'} \\ \boldsymbol{I_4} \end{pmatrix}$ selects the four RGB components for the corresponding waves, and $\boldsymbol{Z_t^e} = Diag(z_t^t\ z_t^{t-3} z_t^{t-6} z_t^{t-9} z_t^{t-12})$ multiplies the five survey error components $\tilde{e}_t^{t-j}$ with the design standard errors. Vectors $\mathbf{1_5}$ and $\mathbf{0_4'}$ denote a vertical vector of ones and a horizontal vector of zeros, respectively, of a dimension specified in the subscript.

The transition matrix $\boldsymbol{T}$ is specified below:

$$\boldsymbol{T} = Blockdiag(\boldsymbol{T^L}\ \boldsymbol{T^\gamma}\ \boldsymbol{T^\lambda}\ \boldsymbol{T^e}),$$

where $\boldsymbol{T^L} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \boldsymbol{T^\gamma} = Blockdiag(\boldsymbol{C_1} \ldots \boldsymbol{C_5}\ -1),$

$$\boldsymbol{C_l} = \begin{pmatrix} cos(\frac{l\pi}{6}) & sin(\frac{l\pi}{6}) \\ -sin(\frac{l\pi}{6}) & cos(\frac{l\pi}{6}) \end{pmatrix}, l = \{1, \ldots, 5\},$$

$$\boldsymbol{T^\lambda} = \boldsymbol{I_4},$$

$$\boldsymbol{T^e} = \begin{pmatrix} \mathbf{0_4'} & 0 & \mathbf{0_4'} & \mathbf{0_4'} \\ \boldsymbol{O_{4\times4}} & \mathbf{0_4} & \rho\boldsymbol{I_4} & \boldsymbol{O_{4\times4}} \\ \boldsymbol{O_{4\times4}} & \mathbf{0_4} & \boldsymbol{O_{4\times4}} & \boldsymbol{I_4} \\ \boldsymbol{I_4} & \mathbf{0_4} & \boldsymbol{O_{4\times4}} & \boldsymbol{O_{4\times4}} \end{pmatrix}.$$

Vector $\boldsymbol{\eta_t}$ contains stochastic terms of the state vector $\boldsymbol{\alpha_t}$:

$\boldsymbol{\eta_t} =$

$(0\, \eta_{R,t}\, \omega_{t,1}\, \omega_{t,1}^*\, \omega_{t,2}\, \omega_{t,2}^* \ldots \omega_{t,5}\, \omega_{t,5}^*\, \omega_{t,6}\, \eta_{\lambda,t}^{t-3} \eta_{\lambda,t}^{t-6} \eta_{\lambda,t}^{t-9} \eta_{\lambda,t}^{t-12}\, \nu_t^t \nu_t^{t-3} \nu_t^{t-6} \nu_t^{t-9} \nu_t^{t-12}\, 0\,0\,0\,0\,0\,0\,0\,0)'$.

The covariance matrix of the state stochastic terms is diagonal:

$$\boldsymbol{\Omega} = Blockdiag(0\ \sigma_R^2\ [\sigma_\omega^2 \mathbf{1_{11}'}]\ \boldsymbol{\Omega^\lambda}\ \boldsymbol{\Omega^e}),$$

$$\boldsymbol{\Omega^\lambda} = \sigma_\lambda^2 \boldsymbol{I_4},$$

$$\boldsymbol{\Omega^e} = Diag(\sigma_{\nu^1}^2\ \sigma_{\nu^2}^2\ \sigma_{\nu^3}^2\ \sigma_{\nu^4}^2\ \sigma_{\nu^5}^2\ \mathbf{0_8'}).$$

In the case of the LFS, all the hyperparameters estimated with the ML-method are contained in the $\boldsymbol{\Omega}$-matrix, whereas the hyperparameter ρ in the transition matrix $\boldsymbol{T}$.

B Hyperparameter Distributions under Four Versions of the DLFS Model

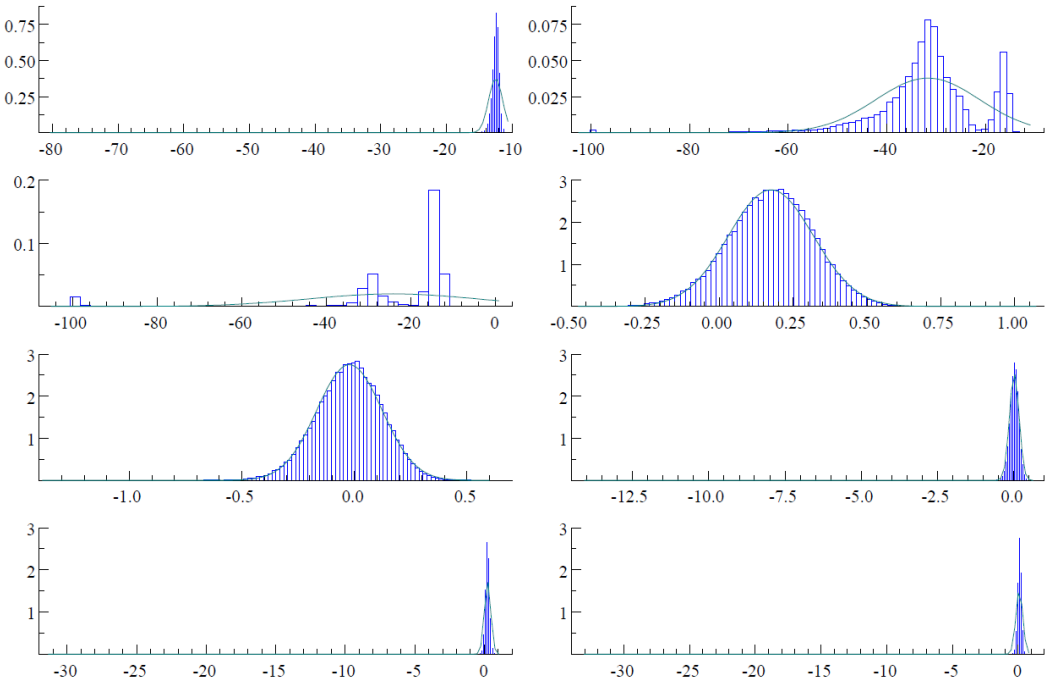

Figure B.1 Parameter distribution under the complete DLFS model (Model 1), left to right: $ln(\hat{\sigma}_R^2), ln(\hat{\sigma}_\gamma^2), ln(\hat{\sigma}_\lambda^2), ln(\hat{\sigma}_{v_1}^2), ln(\hat{\sigma}_{v_2}^2), ln(\hat{\sigma}_{v_3}^2), ln(\hat{\sigma}_{v_4}^2), ln(\hat{\sigma}_{v_5}^2)$**; the normal density with the same mean and variance superimposed; 50000 simulations, T=114.**

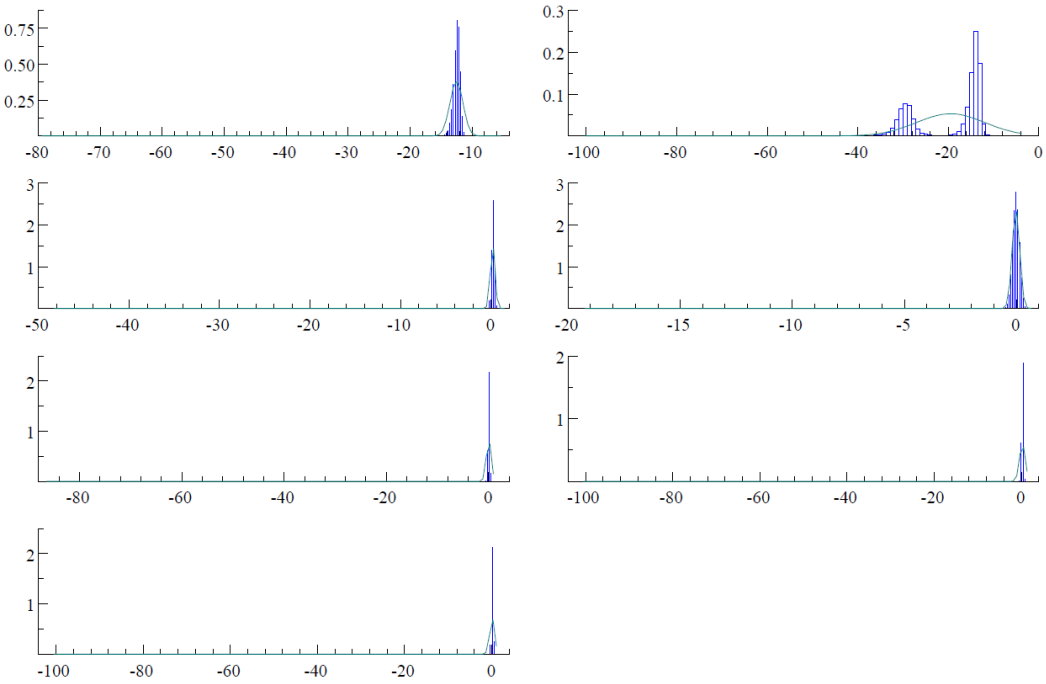

Figure B.2 Parameter distribution under Model 2, left to right: $ln(\hat{\sigma}_R^2), ln(\hat{\sigma}_\lambda^2), ln(\hat{\sigma}_{\nu_1}^2), ln(\hat{\sigma}_{\nu_2}^2), ln(\hat{\sigma}_{\nu_3}^2), ln(\hat{\sigma}_{\nu_4}^2), ln(\hat{\sigma}_{\nu_5}^2)$**; the normal density with the same mean and variance superimposed; 50000 simulations, T=114.**

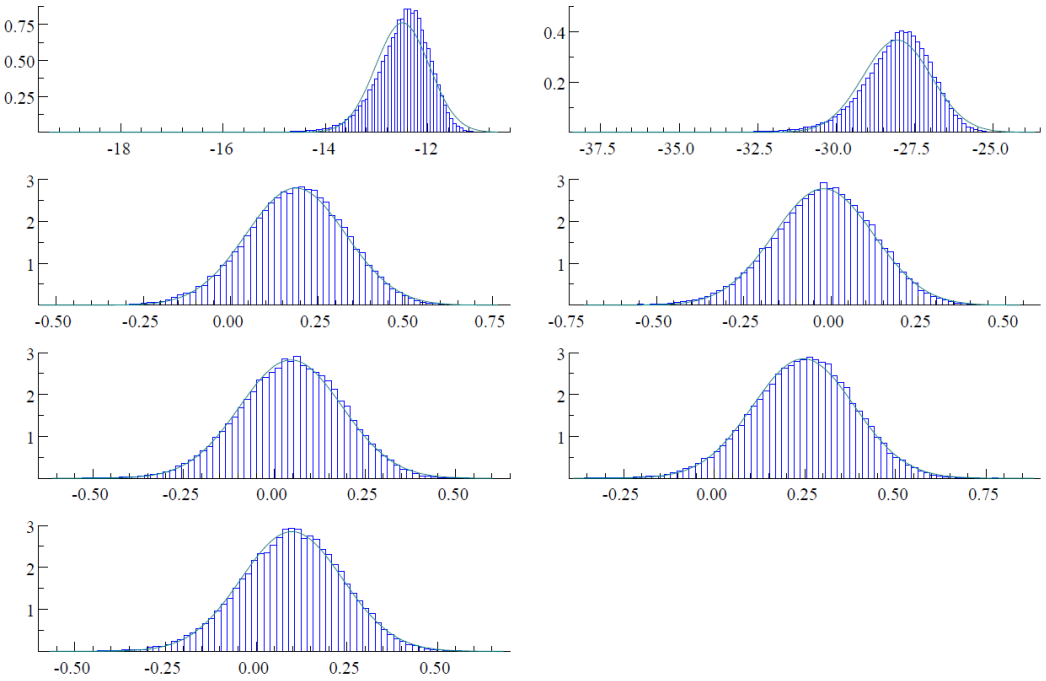

Figure B.3 Parameter distribution under Model 3, left to right: $ln(\hat{\sigma}_R^2), ln(\hat{\sigma}_\gamma^2), ln(\hat{\sigma}_{\nu_1}^2), ln(\hat{\sigma}_{\nu_2}^2), ln(\hat{\sigma}_{\nu_3}^2), ln(\hat{\sigma}_{\nu_4}^2), ln(\hat{\sigma}_{\nu_5}^2)$; the normal density with the same mean and variance superimposed; 50000 simulations, T=114.

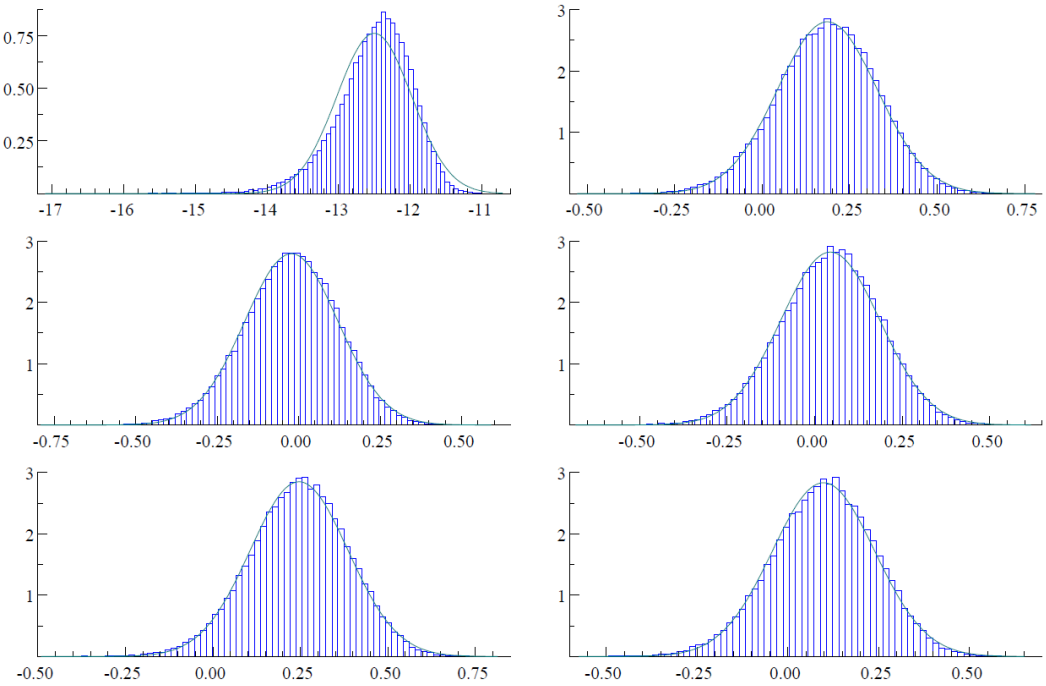

Figure B.4 Parameter distribution under Model 4, left to right: $ln(\hat{\sigma}_R^2), ln(\hat{\sigma}_{v_1}^2), ln(\hat{\sigma}_{v_2}^2), ln(\hat{\sigma}_{v_3}^2), ln(\hat{\sigma}_{v_4}^2), ln(\hat{\sigma}_{v_5}^2)$; the normal density with the same mean and variance superimposed; 50000 simulations, T=114.

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
2014–2015	2014 to 2015 inclusive
2014/2015	Average for 2014 to 2015 inclusive
2014/'15	Crop year, financial year, school year, etc., beginning in 2014 and ending in 2015
2012/'13–2014/'15	Crop year, financial year, etc., 2012/'13 to 2014/'15 inclusive

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

Publisher
Statistics Netherlands
Henri Faasdreef 312, 2492 JP The Hague
www.cbs.nl

Prepress
Studio BCO, Den Haag

Design
Edenspiekermann

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