Discussion Paper

State Space Time Series Modelling of the Dutch Labour Force Survey: Model Selection and MSE Estimation, - Extended Version

2016 | 13

Oksana Bollineni-Balabay
Jan van den Brakel
Franz Palm

Structural time series models are known as a powerful technique for variance reduction in the framework of small area estimation (SAE) based on repeatedly conducted surveys. Statistics Netherlands currently uses a structural time series model for the publication of official monthly figures about the labour force. 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. This paper describes a simulation aimed at studying the properties of the model hyperparameters. The simulation of the distributions of the hyperparameters under different model specifications complements standard model diagnostics for state space models by providing additional insight into irrelevant hyperparameters in the model. Uncertainty around the model hyperparameters is another major issue. If it is not taken into account, the newly obtained variance estimates for the state variables of interest become negatively biased, particularly in short time series. In order to account for the negative bias in the MSE estimates of the Dutch Labour Force survey (DLFS) model, several estimation approaches known in the literature are considered. Apart from the MSE bias comparison, this paper also provides insight into the variances and MSEs of the MSE estimators considered. The results based on the DLFS suggest that the best performing approach we reviewed may correct for a 2-percent negative relative bias in 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 still about 22 percent smaller than the design-based standard errors.

1 Introduction

Figures on the labour force produced by national statistical institutes (NSIs) are generally based on Labour Force Surveys (LFS). There is an increasing interest to producing these indicators at a monthly frequency (EUROSTAT (2015)). Sample sizes are, however, hardly ever large enough even at the national level for producing sufficiently precise monthly labour force figures based on design-based estimators known from classical sampling theory (Särndal et al. (1992), Cochran (1977)). When figures are insufficiently precise, 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 and Molina (2015) and Pfeffermann (2013). Repeatedly conducted surveys in particular have a potential for improvement within the framework of structural time series (STS) or multilevel time series models.

STS models, as well as multilevel models, usually contain unknown hyperparameters that have to be estimated, which translates into larger standard errors of the domain predictions. If this uncertainty (here and further in this paper referred to as hyperparameter uncertainty) is not taken into account, estimated MSEs of the domain predictors become negatively biased. Within the framework of multilevel models, taking the hyperparameter uncertainty into account is viewed as a necessary and common practice. It is routinely performed when those models are estimated with the empirical best linear unbiased predictor (EBLUP) or the hierarchical Bayesian (HB) approach, see Rao and Molina (2015), Ch. 6-7, 10. STS models, in turn, are not as widely used in SAE as multilevel models. The Kalman filter, usually applied to fit STS models, ignores the hyperparameter uncertainty, and therefore produces negatively-biased MSE estimates. 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., Bollineni-Balabay

et al. (2016), Van den Brakel and Krieg (2015), Krieg and van den Brakel (2012), Van den Brakel and Krieg (2009), Pfeffermann and Bleuer (1993), Tiller (1992).

At Statistics Netherlands, a multivariate STS model is used to produce official monthly labour force figures for the Dutch Labour Force survey (DLFS). This model was originally proposed by Pfeffermann (1991). The DLFS survey is, as in many other countries, based on a rotating panel and features insufficiently large sample sizes for production of monthly figures. The STS model applied to the survey design-based estimates uses sample information from preceding time periods and accounts for different features 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 of the unemployed labour force are obtained (see Van den Brakel and Krieg (2009), Van den Brakel and Krieg (2015)). STS models are also applied in the production of official statistics at the US Bureau of Labor Statistics, Tiller (1992). Interest to this technique has been growing among several other NSIs spread around the world, for example at NSIs of Australia (Zhang and Honchar (2016)), Israel and the UK (ONS (2015)).

This paper presents an extended Monte-Carlo simulation study, where the DLFS model acts as the data generation process. Such a simulation is an insightful step into the process of model selection before the model is implemented in the production of official statistics. First, evaluating the hyperparameter distributions under different model specifications provides additional insight into the importance of retaining certain hyper-parameters in the model. Standard model diagnostics for state space models provide limited information on irrelevant hyperparameters. In case of model overspecification, not only may the distribution of redundant hyperparameter estimates largely deviate from normality, but estimation of other hyperparameters may also be disturbed. Therefore, even if the model diagnostics is satisfactory, it may still be wise to simulate the model and to examine the distribution of the maximum likelihood (ML) estimator of the model's hyperparameters.

Another aim of the simulation is evaluating to which extent uncertainty around the hyperparameter estimates affects estimation of the STS model-based MSEs. Ignoring the hyperparameter uncertainty in MSE estimation is only acceptable if the available time series are sufficiently long. Depending upon a particular application, the length of time required to be "sufficiently long" will vary. Most often, uninterrupted time series available at NSIs are relatively short, mainly due to survey redesigns. This paper illustrates how the hyperparameter uncertainty problem decays as the DLFS time series increase from 48 to 200 months. More importantly, this paper attempts to find the best MSE estimation method for the DLFS model. The literature offers several ways to account for the hyperparameter uncertainty in STS models: asymptotic approximation, bootstrapping and the full Bayesian approach (for the latter approach, see Durbin and Koopman (2012), Chapter 13). Among those 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.

The contribution of the paper is three-fold. First of all, it shows how the Monte Carlo simulation can be used to check for model overspecification (i.e. for redundant hyperparameters). Secondly, it suggests the best of the proposed approaches to MSE estimation for the DLFS and offers a more realistic evaluation of the variance reduction obtained with the STS model compared to the design-based approach. Finally, this Monte-Carlo study refutes 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. Apart from MSE bias comparison, this paper also provides insight into the variance and MSEs of these MSE estimators. To the best of our knowledge, the variability of

the above-mentioned bootstrap methods has not been studied yet.
The paper is structured as follows. Section 2 contains a description of the DLFS and the model currently used by Statistics Netherlands. Section 3 reviews the above-mentioned approaches to the MSE estimation. Details on the simulation and bootstrap setup specific to the DLFS are given in Section 4. Results of the simulation study with respect to model selection and to the performance of the MSE estimation methods are presented in Subsections 5.1 and 5.2, respectively. Section 6 contains concluding remarks.

2 The Dutch Labour Force survey

2.1 The DLFS design

The DLFS has been based on a rotating panel design since October 1999. Every month, a sample of addresses is drawn according to a stratified two-stage sample design. Strata are formed by geographical regions; municipalities are the primary sampling units and addresses are the secondary sampling units. All households residing on one address are included in the sample. In this paper, the DLFS data observed from January 2001 until June 2010 are considered. During this period, data in the first wave were collected by means of computer assisted personal interviewing (CAPI) by interviewers that visit sampled households at home. After a maximum of six attempts, an interviewer leaves a letter with the request to contact the interviewer by telephone to make an appointment for an interview. When a household member cannot be contacted, proxy interviewing is allowed by members of the same household. Households in which one or more of the selected persons do not respond for themselves or in a proxy interview, are treated as non-responding households. Respondents are re-interviewed four times at quarterly intervals. In these four subsequent waves, data are collected by means of computer assisted telephone interviewing (CATI). During these re-interviews, a condensed questionnaire is applied to establish any changes in the labour market position of the respondents. Proxy interviewing is also allowed during these re-interviews. Mobile phone numbers and secret land line numbers are collected in the first wave to avoid panel attrition. Commencing the moment the DLFS was conducted as a rotating panel design, the gross sample size was about 6200 addresses per month on average, with about 65% completely responding households. The response rates in the follow-up waves are about 90% compared to the preceding wave.
The general regression (GREG) estimator (Särndal et al. (1992)) is applied to estimate the total unemployed labour force. This estimator accounts for the complexity of the sample design and uses auxiliary information available from registers to correct, at least partially, for selective non-response. The STS model defined in the next subsection is applied to the unemployed labour force GREG estimates that are based on the data observed in each month and each wave. Let Y_t^{t-j} denote the GREG estimate of the total number of unemployed in month t based on the wave of respondents that entered the panel in month $t-j$. Five such estimates are obtained per month, each of them being respectively based on the sample that entered the survey in month $t-j, j=\{0,3,6,9,12\}$. The GREG estimator for this population total is defined as:

$$Y_t^{t-j} = \sum_{k \epsilon s} w_{k,t} \left(\sum_{i=1}^{n_{k,t}} y_{i,k,t} \right) \tag{2.1}$$

with $y_{i,k,t}$ representing the sample observations that are equal to 1 if the i-th person in the k-th household is unemployed, and zero otherwise; $n_{k,t}$ is the number of persons aged 15 or above in

the k-th household; $w_{k,t}$ are the regression weights for household k at time t. The method of Lemaître and Dufour (1987) is used to obtain equal weights for all persons within the same household:

$$w_{k,t} = \frac{1}{\pi_{k,t}}\left[1 + \left(\boldsymbol{X_t} - \sum_{k\epsilon s}\frac{\boldsymbol{x_{k,t}}}{\pi_{k,t}}\right)\left(\sum_{k\epsilon s}\frac{\boldsymbol{x_{k,t}}\boldsymbol{x'_{k,t}}}{\pi_{k,t}g_{k,t}}\right)^{-1}\frac{\boldsymbol{x_{k,t}}}{g_{k,t}}\right], \tag{2.2}$$

where $\pi_{k,t}$ is the inclusion probability of household k at time t, $g_{k,t}$ is the size of household k at time t; $\boldsymbol{x_{k,t}} = \sum_{i=1}^{n_{k,t}} \boldsymbol{x_{i,k,t}}$, with $\boldsymbol{x_{i,k,t}}$ being a J-dimensional vector with the weighting model auxiliary information on the i-th person in the k-th household at time t. Vector $\boldsymbol{X_t}$ contains population totals of auxiliary variables. The weighting model is defined by the following variables (with the number of categories in brackets): Age(5)Gender + Geographic Region(44) + Gender(2)×Age(21) + Age(5)×Marital Status(2) + Ethnicity(8), where × stands for interaction of variables, and Age(5)Gender is a variable classified into eight classes where Age has five categories, with the second, third and fourth age category being itemized into two genders.
The design-based variance estimator for the GREG estimate Y_t^{t-j} is approximated by:

$$\widehat{Var}(Y_t^{t-j}) = \sum_{h=1}^{H}\frac{n_{h,t}}{n_{h,t}-1}\left(\sum_{k=1}^{n_{h,t}}(w_{k,t}\hat{e}_{k,t})^2 - \frac{1}{n_{h,t}}\left(\sum_{k=1}^{n_{h,t}} w_{k,t}\hat{e}_{k,t}\right)^2\right), \tag{2.3}$$

where the GREG residuals are $\hat{e}_{k,t} = \sum_{i=1}^{n_{k,t}}(y_{i,k,t} - \boldsymbol{x'_{i,k,t}}\hat{\boldsymbol{\beta}}_{\boldsymbol{t}})$; $n_{h,t}$ is the number of households in stratum h (with H being the total number of strata); vector $\hat{\boldsymbol{\beta}}_{\boldsymbol{t}}$ is a Horvitz-Thompson type estimator for the regression coefficient that is obtained from regressing the target variable on the auxiliary variables from the sample.

2.2 The STS model for the DLFS

There are two reasons why Statistics Netherlands took a decision to switch to a time series model-based production approach in June 2010. One reason for that was inadequately small sample sizes for production of monthly estimates. With a net sample size of about 4000 households in the first wave on average, the GREG estimates of the unemployed labour force had a coefficient of variation of about 4% at the national level, which was considered to be too volatile for official statistical publications. In addition to that, monthly unemployment figures must be published for six domains based on a classification of gender and age. The design-based estimates of these domains feature much higher coefficients of variation. Another problem with the DLFS is the so-called rotation group bias (RGB), which refers to systematic differences between the estimates of different waves (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. See Van den Brakel and Krieg (2009) for a more detailed discussion.
Both problems can be solved with the help of an STS model that accounts for the RGB and makes use of the information accumulated over time to produce point-estimates with smaller standard errors. An STS model uses design-based estimates (of all the waves in the rotating panel design) as input and decomposes them into several unobserved components, whereupon the so-called signal - a more reliable series of point-estimates - can be obtained. The signal is usually extracted with the help of the Kalman filter and optionally with a smoother as well. The filter removes a great part of the population and sampling noise from the design-based estimates and produces point- and variance estimates for the signal and its unobserved components (trend, seasonal

etc.). As a result, not only are these signal point-estimates less volatile (see Fig. 2.1), their standard errors are usually substantially lower than the design-based standard errors (in the case of the DLFS, 24 percent smaller). Apart from that, an STS model can extract the RGB pattern from the design-based series in case of a rotating panel design.
In each month t, a five-dimensional vector $\boldsymbol{Y_t} = (Y_t^t\ Y_t^{t-3}\ Y_t^{t-6}\ Y_t^{t-9}\ Y_t^{t-12})'$ with the GREG estimates of the total number of the unemployed labour force based on the five waves is observed. Based on Pfeffermann (1991), Van den Brakel and Krieg (2009) developed the following model for the GREG estimates $\boldsymbol{Y_t}$:

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

where $\mathbf{1_5}$ is a five-dimensional column vector of ones, ξ_t is the unknown (scalar) true population parameter, $\boldsymbol{\lambda_t}$ is a vector containing state variables for the RGB, and $\boldsymbol{e_t}$ is a vector of the survey errors that are correlated with their counterparts from previous waves (the structure will be presented later). It is assumed that the true population parameter can be decomposed in the following way: $\xi_t = L_t + \gamma_t + \varepsilon_t$, which is the sum of a stochastic trend L_t, a stochastic seasonal component γ_t, and an irregular component $\varepsilon_t \overset{iid}{\sim} N(0, \sigma_\varepsilon^2)$.
For the stochastic trend L_t, the so-called smooth-trend model is assumed:

$$L_t = L_{t-1} + R_{t-1},$$
$$R_t = R_{t-1} + \eta_{R,t},$$

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: $\eta_{R,t} \overset{iid}{\sim} N(0, \sigma_{\eta_R}^2)$.
In the case of monthly data, the seasonal component γ_t can be decomposed into six stochastic harmonics:

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

where each of these six harmonics follows the process:

$$\gamma_{t,l} = cos(h_l)\gamma_{t-1,l} + sin(h_l)\gamma_{t-1,l}^* + \omega_{t,l},$$
$$\gamma_{t,l}^* = -sin(h_l)\gamma_{t-1,l} + cos(h_l)\gamma_{t-1,l}^* + \omega_{t,l}^*,$$

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 possess the same variance within and across all the harmonics, such that:

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

To identify the component that models the RGB, it is assumed that the first wave is unbiased, as motivated in Van den Brakel and Krieg (2009). The RGBs for the follow-up waves are time-dependent and are modelled as random walk processes. The rationale behind this is that field-work procedures are subject to frequent changes. Apart from that, response rates change gradually over time. This makes the RGB time-dependent, as illustrated by Van den Brakel and Krieg (2015), Fig. 4.3. The RGB vector for the five waves can be written in the following form: $\boldsymbol{\lambda_t} = (0\ \lambda_t^{t-3}\lambda_t^{t-6}\lambda_t^{t-9}\lambda_t^{t-12})'$, with:

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

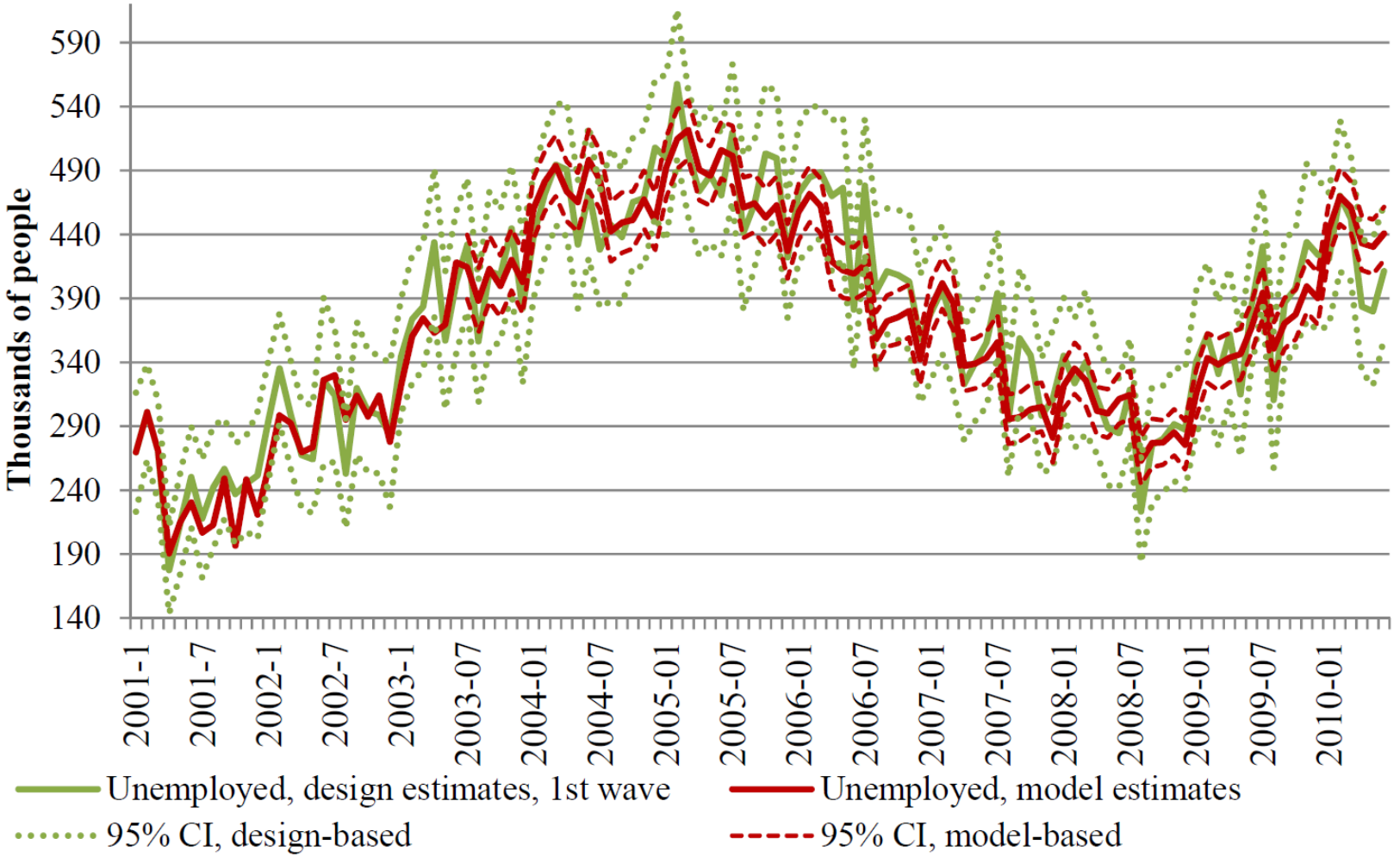

Figure 2.1 **Numbers of unemployed in the Netherlands on a monthly basis: design- and model-based estimates, together with their confidence intervals**

It is assumed that 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 contains the survey errors for the five GREG estimates, i.e. $\boldsymbol{e_t} = (e_t^t, e_t^{t-3}, e_t^{t-6}, e_t^{t-9}, e_t^{t-12})$. To account for the sampling error heterogeneity caused by changes in the survey design and in sample sizes over time, the sampling errors are modelled proportionally to the design-based standard errors according to the following measurement error model proposed by Binder and Dick (1990): $e_t^{t-j} = \tilde{e}_t^{t-j} z_t^{t-j}$, where $z_t^{t-j} = \sqrt{\widehat{Var}(Y_t^{t-j})}$. Here, $\widehat{Var}(Y_t^{t-j})$ are the design-based variance estimates obtained from the micro data using (2.3). They are treated as a priori known sampling variances in the STS model.
Since the sample in the first wave has no overlap with samples observed in the past, $\tilde{e}_t^t$ can be modelled as a white noise with $E(\tilde{e}_t^t) = 0$ and $Var(\tilde{e}_t^t) = \sigma_{v_t^t}^2$. The variance of the survey errors e_t^t will be equal to the variance of the GREG estimates if the maximum likelihood estimate of $\sigma_{v_t^t}^2$ is approximately equal to unity.
The survey errors in the follow-up waves are correlated with the survey errors from the preceding waves. This autocorrelation coefficient is estimated from the survey data using the approach proposed by Pfeffermann et al. (1998). The autocorrelation structure is modelled with an AR(1) model where the autocorrelation coefficient is obtained with the Yule-Walker equations (Van den Brakel and Krieg (2009)):

$$\tilde{e}_t^{t-j} = \rho \tilde{e}_{t-3}^{t-j} + v_t^{t-j}, \quad v_t^{t-j} \overset{iid}{\sim} N(0, \sigma_{v_t^{t-j}}^2), \; j = \{3, 6, 9, 12\}.$$

The first-order autocorrelation coefficient is common for all the four waves and equals 0.208. This estimate is used as a priori information in the model. Since $\tilde{e}_t^{t-j}$ is an AR(1) process, $Var(\tilde{e}_t^{t-j}) = \sigma_{v_t^{t-j}}^2/(1-\rho^2)$. The variance of the sampling error e_t^{t-j} is approximately equal to $\widehat{Var}(Y_t^{t-j})$ if the maximum likelihood estimate of $\sigma_{v_t^{t-j}}^2$ is approximately equal to $(1-\rho^2)$. Five different hyperparameters $\sigma_{v_t^{t-j}}^2$, $j = \{0, 3, 6, 9, 12\}$, are assumed for the survey error components of the five waves. These hyperparameters are estimated with the ML method.
Linear structural time series models with unobserved components are usually fitted with the help of the Kalman filter after putting them into a state space form:

$$\boldsymbol{\alpha_{t+1}} = \boldsymbol{T}\boldsymbol{\alpha_t} + \boldsymbol{\eta_t}, \tag{2.4}$$

$$\boldsymbol{Y_t} = \boldsymbol{Z_t}\boldsymbol{\alpha_t}. \tag{2.5}$$

Here, $\boldsymbol{T}$ and $\boldsymbol{Z_t}$ are known design matrices of the transition and measurement equations, respectively. The transition equation (2.4) describes how the state vector evolves over time, whereas the measurement equation (2.5) reflects the linear relationship between the observations and the state vector. The autoregressive structure of the 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. Here, the population parameter error term ε_t is also modelled as a state variable because it is shared by all the waves. These disturbance terms thus disappear from the measurement equation (2.5) and move to the state vector $\boldsymbol{\alpha_t}$. Each element of the zero-expectation vector $\boldsymbol{\eta_t}$ contains identically and independently distributed disturbance terms. The design-based standard error estimates $\sqrt{\widehat{Var}(Y_t^{t-j})}$ become time-varying elements of the matrix $\boldsymbol{Z_t}$ and are 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:

$$\begin{aligned}
&\boldsymbol{\alpha_t} = (\boldsymbol{\alpha_t^{\xi}}\ \boldsymbol{\alpha_t^{\lambda}}\ \boldsymbol{\alpha_t^{e}})', \text{ where}\\
&\boldsymbol{\alpha_t^{\xi}} = (L_t\ R_t\ \gamma_{t,1}\ \gamma^*_{t,1}...\gamma_{t,5}\ \gamma^*_{t,5}\ \gamma_{t,6}\ \varepsilon_t),\\
&\boldsymbol{\alpha_t^{\lambda}} = (\lambda_t^{t-3}\lambda_t^{t-6}\lambda_t^{t-9}\lambda_t^{t-12}),\\
&\boldsymbol{\alpha_t^{e}} = (\tilde{e}_t^t\ \tilde{e}_t^{t-3}\ \tilde{e}_t^{t-6}\ \tilde{e}_t^{t-9}\ \tilde{e}_t^{t-12}\ \tilde{e}_{t-2}^{t-2}\ \tilde{e}_{t-2}^{t-5}\ \tilde{e}_{t-2}^{t-8}\ \tilde{e}_{t-2}^{t-11}\ \tilde{e}_{t-1}^{t-1}\ \tilde{e}_{t-1}^{t-4}\ \tilde{e}_{t-1}^{t-7}\ \tilde{e}_{t-1}^{t-10}).
\end{aligned} \tag{2.6}$$

All the non-stationary state variables are initialised with a diffuse prior (i.e. with a zero-mean and a very large variance). The five survey error components $\tilde{e}_t^{t-j}, j = \{0, 3, 6, 9, 12\}$ and the population white noise ε_t are stationary state variables that 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 $\boldsymbol{\theta} = (\sigma^2_{\eta_R}\ \sigma^2_{\bar{\omega}}\ \sigma^2_{\varepsilon}\ \sigma^2_{\lambda}\ \sigma^2_{\nu_t^t}\ \sigma^2_{\nu_t^{t-3}}\ \sigma^2_{\nu_t^{t-6}}\ \sigma^2_{\nu_t^{t-9}}\ \sigma^2_{\nu_t^{t-12}}\ \rho)$, and the vector containing only the disturbance variances is called $\boldsymbol{\theta_\sigma} = (\sigma^2_{\eta_R}\ \sigma^2_{\bar{\omega}}\ \sigma^2_{\varepsilon}\ \sigma^2_{\lambda}\ \sigma^2_{\nu_t^t}\ \sigma^2_{\nu_t^{t-3}}\ \sigma^2_{\nu_t^{t-6}}\ \sigma^2_{\nu_t^{t-9}}\ \sigma^2_{\nu_t^{t-12}})$ (note that the variance of the last eight state variables is zero, see Appendix A). To avoid negative estimates, the disturbance variance hyperparameters in $\boldsymbol{\theta_\sigma}$ are estimated on a log-scale. The quasi-maximum likelihood method is used (see e.g. Harvey (1989)), where $\hat{\rho}$-estimates are treated as known.
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 Estimation Approaches

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

$$\begin{aligned}
&\hat{\boldsymbol{\alpha}}_{t|t} = \boldsymbol{T}\hat{\boldsymbol{\alpha}}_{t-1|t-1} + \boldsymbol{K_t}\boldsymbol{\epsilon_t},\\
&\boldsymbol{P_{t|t}} = \boldsymbol{P_{t|t-1}} - \boldsymbol{P_{t|t-1}}\boldsymbol{Z_t'}\boldsymbol{K_t'},\\
&\boldsymbol{P_{t|t-1}} = \boldsymbol{T}\boldsymbol{P_{t-1|t-1}}\boldsymbol{T'} + \boldsymbol{Q},
\end{aligned}$$

where $'$ stands for a transpose, $\hat{\boldsymbol{\alpha}}_{t|t}$ and $\boldsymbol{P_{t|t}} = \boldsymbol{E_t}[(\hat{\boldsymbol{\alpha}}_{t|t} - \boldsymbol{\alpha_t})(\hat{\boldsymbol{\alpha}}_{t|t} - \boldsymbol{\alpha_t})']$ 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 are usually referred to as *filtered* estimates. Matrix $\boldsymbol{P_{t|t-1}}$ is the state covariance matrix of the predictor $\hat{\boldsymbol{\alpha}}_{t|t-1}$ at time $t-1$: $\boldsymbol{P_{t|t-1}} = \boldsymbol{E_t}[(\hat{\boldsymbol{\alpha}}_{t|t-1} - \boldsymbol{\alpha_t})(\hat{\boldsymbol{\alpha}}_{t|t-1} - \boldsymbol{\alpha_t})']$. $\boldsymbol{K_t}$ is the so-called Kalman gain: $\boldsymbol{K_t} = \boldsymbol{P_{t|t-1}}\boldsymbol{Z_t'}\boldsymbol{F_t^{-1}}$, $\boldsymbol{\epsilon_t} = \boldsymbol{Y_t} - \boldsymbol{Z_t}\hat{\boldsymbol{\alpha}}_{t|t-1}$ are the innovations and $\boldsymbol{F_t}$ is the innovation covariance matrix: $\boldsymbol{F_t} = \boldsymbol{Z_t}\boldsymbol{P_{t|t-1}}\boldsymbol{Z_t'}$ (note that the covariance matrix of the measurement equation error terms is absent here because those error terms have been placed in the state vector).
The MSE 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 joint distribution of the state vector and the Y-values at time t, provided this expectation exists. In practice, the true hyperparameter vector is replaced by its estimate $\hat{\boldsymbol{\theta}}$ 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:

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

which is larger than the MSE in (3.1) and can be decomposed as the sum of the filter uncertainty and parameter uncertainty, provided the error terms are normal:

$$\begin{aligned}\boldsymbol{MSE}_{t|t} =& \boldsymbol{E}_t[(\hat{\boldsymbol{\alpha}}_{t|t}(\boldsymbol{\theta}) - \boldsymbol{\alpha}_t)(\hat{\boldsymbol{\alpha}}_{t|t}(\boldsymbol{\theta}) - \boldsymbol{\alpha}_t)'] + \\ & \boldsymbol{E}_t[(\hat{\boldsymbol{\alpha}}_{t|t}(\hat{\boldsymbol{\theta}}) - \hat{\boldsymbol{\alpha}}_{t|t}(\boldsymbol{\theta}))(\hat{\boldsymbol{\alpha}}_{t|t}(\hat{\boldsymbol{\theta}}) - \hat{\boldsymbol{\alpha}}_{t|t}(\boldsymbol{\theta}))'].\end{aligned} \tag{3.2}$$

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

For the parametric bootstrap, the necessary disturbances for state recursions (2.4) and (2.5) are drawn from their joint conditional multivariate normal density $\boldsymbol{\eta}_t \overset{iid}{\sim} MN(\mathbf{0}, \boldsymbol{\Omega})$. Non-parametric bootstrap, in turn, has an advantage of not depending on any particular assumption about this joint distribution. In this case, a bootstrap sample of standardised innovations $\{\boldsymbol{\epsilon}_{r+1}^{\boldsymbol{b,St}}, \ldots, \boldsymbol{\epsilon}_T^{\boldsymbol{b,St}}\}$ is obtained by sampling with replacement from $\{\boldsymbol{\epsilon}_{r+1}^{\boldsymbol{St}}(\hat{\boldsymbol{\theta}}), \ldots, \boldsymbol{\epsilon}_T^{\boldsymbol{St}}(\hat{\boldsymbol{\theta}})\}$ where $\boldsymbol{\epsilon}_t^{\boldsymbol{St}} = \boldsymbol{F}_t^{-1/2}(\hat{\boldsymbol{\theta}})\boldsymbol{\epsilon}_t(\hat{\boldsymbol{\theta}})$ are standardized innovations from the Kalman filter recursions based on the original ML estimates of the hyperparameters. In the DLFS model, the first $r = 13$ time points are not subject to resampling and constitute the so-called diffuse sample (this is the time needed to construct a proper distribution for the non-stationary state variables; see Koopman (1997) for initialisation of non-stationary state variables). A bootstrap observation set $\{\boldsymbol{Y}_1^{\boldsymbol{b}}, \ldots, \boldsymbol{Y}_T^{\boldsymbol{b}}\}$ is then constructed by running the so-called innovation form of the Kalman filter:

$$\hat{\boldsymbol{\alpha}}_{t|t}^{\boldsymbol{b}} = \hat{\boldsymbol{\alpha}}_{t|t-1}^{\boldsymbol{b}} + \boldsymbol{K}_t(\hat{\boldsymbol{\theta}})\boldsymbol{F}_t^{1/2}(\hat{\boldsymbol{\theta}})\boldsymbol{\epsilon}_t^{\boldsymbol{b,St}},$$
$$\hat{\boldsymbol{Y}}_t = \boldsymbol{Z}_t\hat{\boldsymbol{\alpha}}_{t|t-1}^{\boldsymbol{b}} + \boldsymbol{F}_t^{1/2}(\hat{\boldsymbol{\theta}})\boldsymbol{\epsilon}_t^{\boldsymbol{b,St}}, t = d+1, \ldots, T.$$

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, observations Y_t^{t-j} in $\boldsymbol{Y}_t$ are reexpressed in a way that they can be treated as univariate time series. Then, for each of these series, the Kalman gain $\boldsymbol{K}_t$ is a vector, whereas F_t is a scalar.

If an STS model contains non-stationary components, as is the case with the DLFS model, the generated series are likely to diverge from the original dataset they have been bootstrapped from, $\{\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_T\}$. Therefore, a special procedure is required for bootstrap samples to be brought in accordance with the pattern of the given dataset. This can be done with the help of the simulation smoother algorithm developed by Durbin and Koopman (2002). Let $\{\boldsymbol{Y}_1^{\boldsymbol{b},\dagger}, \ldots, \boldsymbol{Y}_T^{\boldsymbol{b},\dagger}\}$ denote the initially generated bootstrap series that have to undergo correction with the simulation smoother. At the first step, the differences between the original series $\{\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_T\}$ and the bootstrap series $\{\boldsymbol{Y}_1^{\boldsymbol{b},\dagger}, \ldots, \boldsymbol{Y}_T^{\boldsymbol{b},\dagger}\}$ - generated either parametrically or non-parametrically - are calculated. As the next step, smoothed state variables of the series of these differences are extracted. The means of these smoothed states are added to $\{\boldsymbol{Y}_1^{\boldsymbol{b},\dagger}, \ldots, \boldsymbol{Y}_T^{\boldsymbol{b},\dagger}\}$, which results in corrected bootstrap series, further denoted as $\{\boldsymbol{Y}_1^{\boldsymbol{b}}, \ldots, \boldsymbol{Y}_T^{\boldsymbol{b}}\}$ (see Koopman, Shephard, and Doornik (2008), Ch.8.4.2. for description). 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.

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

bootstrap) and of 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 MSE estimation conditional on the data, which means that bootstrap hyperparameters are further applied to the original data series for obtaining bootstrap estimates of the state variables. Bootstrapping can be done both parametrically and non-parametrically, following the steps below:

1. Estimate the model and obtain the hyperparameter estimates $\hat{\boldsymbol{\theta}}$.
2. Generate a bootstrap sample $\{\boldsymbol{Y_1^b}, \ldots, \boldsymbol{Y_T^b}\}$ using $\hat{\boldsymbol{\theta}}$, either parametrically or non-parametrically, as described in the introduction to this section. If the model is non-stationary, the bootstrap sample has to be corrected with the help of the simulation smoother, as described in the introduction to this section.
3. The bootstrap dataset $\{\boldsymbol{Y_1^b}, \ldots, \boldsymbol{Y_T^b}\}$ is used to obtain both the survey error autocorrelation parameter estimates $\hat{\rho}^b$ and bootstrap ML estimates $\hat{\boldsymbol{\theta}}_{\boldsymbol{\sigma}}^{\boldsymbol{b}}$. Thereafter, the Kalman filter is launched using the original series $\{\boldsymbol{Y_1}, \ldots, \boldsymbol{Y_T}\}$ and the newly-estimated $\hat{\boldsymbol{\theta}}^{\boldsymbol{b}}$, which produces $\hat{\boldsymbol{\alpha}}_{\boldsymbol{t|t}}(\hat{\boldsymbol{\theta}}^{\boldsymbol{b}})$ and $\boldsymbol{P}_{\boldsymbol{t|t}}(\hat{\boldsymbol{\theta}}^{\boldsymbol{b}})$.
4. Then, steps 2-3 are repeated B times.
5. Having obtained B bootstrap replicates, the MSE can be estimated in the following way:

$$\widehat{\boldsymbol{MSE}}_{\boldsymbol{t|t}}^{\boldsymbol{RR}} = \frac{\boldsymbol{1}}{\boldsymbol{B}}\sum_{\boldsymbol{b=1}}^{\boldsymbol{B}} \boldsymbol{P}_{\boldsymbol{t|t}}(\hat{\boldsymbol{\theta}}^{\boldsymbol{b}}) + \frac{\boldsymbol{1}}{\boldsymbol{B}}\sum_{\boldsymbol{b=1}}^{\boldsymbol{B}} [\hat{\boldsymbol{\alpha}}_{\boldsymbol{t|t}}(\hat{\boldsymbol{\theta}}^{\boldsymbol{b}}) - \bar{\boldsymbol{\alpha}}_{\boldsymbol{t|t}}][\hat{\boldsymbol{\alpha}}_{\boldsymbol{t|t}}(\hat{\boldsymbol{\theta}}^{\boldsymbol{b}}) - \bar{\boldsymbol{\alpha}}_{\boldsymbol{t|t}}]', \tag{3.3}$$

where $\bar{\boldsymbol{\alpha}}_{\boldsymbol{t|t}} = \frac{\boldsymbol{1}}{\boldsymbol{B}}\sum_{\boldsymbol{b=1}}^{\boldsymbol{B}} \hat{\boldsymbol{\alpha}}_{\boldsymbol{t|t}}(\hat{\boldsymbol{\theta}}^{\boldsymbol{b}})$.
Equation (3.3) is applied both for the parametric and non-parametric bootstrap MSE-estimators (denoted hereafter 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 expectation of the squared error loss in (3.2) is taken unconditionally on the data, i.e. the bootstrap state variables under this expectation are obtained from the bootstrap dataset $\{\boldsymbol{Y_1^b}, \ldots, \boldsymbol{Y_T^b}\}$, instead of the original data $\{\boldsymbol{Y_1}, \ldots, \boldsymbol{Y_T}\}$ as in Rodriguez and Ruiz (2012). Unlike Rodriguez and Ruiz (2012), Pfeffermann and Tiller (2005) approximate the true MSE up to the order of $O(1/T^2)$ - a property that is theoretically proven in Pfeffermann and Tiller (2005), Appendix C. Using results in Hall and Martin (1988), they show that the true MSE, being

$$\boldsymbol{MSE}_{\boldsymbol{t|t}} = \boldsymbol{P}_{\boldsymbol{t|t}}(\boldsymbol{\theta}) + \boldsymbol{E}[(\hat{\boldsymbol{\alpha}}_{\boldsymbol{t|t}}(\hat{\boldsymbol{\theta}}) - \hat{\boldsymbol{\alpha}}_{\boldsymbol{t|t}}(\boldsymbol{\theta}))(\hat{\boldsymbol{\alpha}}_{\boldsymbol{t|t}}(\hat{\boldsymbol{\theta}}) - \hat{\boldsymbol{\alpha}}_{\boldsymbol{t|t}}(\boldsymbol{\theta}))'], \tag{3.4}$$

can be estimated by its bootstrap analogues as follows:

$$\boldsymbol{P}_{\boldsymbol{t|t}}(\boldsymbol{\theta}) = \boldsymbol{2P}_{\boldsymbol{t|t}}(\hat{\boldsymbol{\theta}}) - \frac{\boldsymbol{1}}{\boldsymbol{B}}\sum_{\boldsymbol{b=1}}^{\boldsymbol{B}} \boldsymbol{P}_{\boldsymbol{t|t}}(\hat{\boldsymbol{\theta}}^{\boldsymbol{b}}) + \boldsymbol{O}\left(\boldsymbol{1/T^2}\right), \tag{3.5}$$

$$\begin{aligned}&\boldsymbol{E}[(\hat{\boldsymbol{\alpha}}_{\boldsymbol{t|t}}(\hat{\boldsymbol{\theta}}) - \hat{\boldsymbol{\alpha}}_{\boldsymbol{t|t}}(\boldsymbol{\theta}))(\hat{\boldsymbol{\alpha}}_{\boldsymbol{t|t}}(\hat{\boldsymbol{\theta}}) - \hat{\boldsymbol{\alpha}}_{\boldsymbol{t|t}}(\boldsymbol{\theta}))'] = \\ &\frac{\boldsymbol{1}}{\boldsymbol{B}}\sum_{\boldsymbol{b=1}}^{\boldsymbol{B}} [\hat{\boldsymbol{\alpha}}_{\boldsymbol{t|t}}^{\boldsymbol{b}}(\hat{\boldsymbol{\theta}}^{\boldsymbol{b}}) - \hat{\boldsymbol{\alpha}}_{\boldsymbol{t|t}}^{\boldsymbol{b}}(\hat{\boldsymbol{\theta}})][\hat{\boldsymbol{\alpha}}_{\boldsymbol{t|t}}^{\boldsymbol{b}}(\hat{\boldsymbol{\theta}}^{\boldsymbol{b}}) - \hat{\boldsymbol{\alpha}}_{\boldsymbol{t|t}}^{\boldsymbol{b}}(\hat{\boldsymbol{\theta}})]' + \boldsymbol{O}\left(\boldsymbol{1/T^2}\right).\end{aligned} \tag{3.6}$$

Equations (3.5) and (3.6) correspond to the first and the second terms of equation (3.2), respectively. The resulting MSE-estimator below has a bias of order $O\left(1/T^2\right)$:

$$\widehat{\boldsymbol{MSE}}_{t|t}^{\boldsymbol{PT}} = 2\boldsymbol{P}_{t|t}(\hat{\boldsymbol{\theta}}) - \frac{\boldsymbol{1}}{\boldsymbol{B}}\sum_{\boldsymbol{b=1}}^{\boldsymbol{B}} \boldsymbol{P}_{t|t}(\hat{\boldsymbol{\theta}}^{\boldsymbol{b}}) + \\ + \frac{1}{B}\sum_{\boldsymbol{b=1}}^{\boldsymbol{B}} [\hat{\boldsymbol{\alpha}}_{t|t}^{\boldsymbol{b}}(\hat{\boldsymbol{\theta}}^{\boldsymbol{b}}) - \hat{\boldsymbol{\alpha}}_{t|t}^{\boldsymbol{b}}(\hat{\boldsymbol{\theta}})][\hat{\boldsymbol{\alpha}}_{t|t}^{\boldsymbol{b}}(\hat{\boldsymbol{\theta}}^{\boldsymbol{b}}) - \hat{\boldsymbol{\alpha}}_{t|t}^{\boldsymbol{b}}(\hat{\boldsymbol{\theta}})]'. \tag{3.7}$$

Equation (3.7) is applied both for the parametric and non-parametric bootstrap MSE-estimators (denoted further as MSE^{PT1} and MSE^{PT2}, respectively). MSE-calculation in (3.7) requires two Kalman filter runs for every bootstrap series. In the first run, $\hat{\boldsymbol{\alpha}}_{t|t}^{\boldsymbol{b}}(\hat{\boldsymbol{\theta}}^{\boldsymbol{b}})$ is estimated from the bootstrap data set $\{\boldsymbol{Y}_{\boldsymbol{1}}^{\boldsymbol{b}}, \ldots, \boldsymbol{Y}_{\boldsymbol{T}}^{\boldsymbol{b}}\}$ and the bootstrap parameters $\hat{\boldsymbol{\theta}}^{\boldsymbol{b}}$. In this run, $\boldsymbol{P}_{t|t}(\hat{\boldsymbol{\theta}}^{\boldsymbol{b}})$ can also be obtained based on $\hat{\boldsymbol{\theta}}^{\boldsymbol{b}}$, since matrix $\boldsymbol{P}_{t|t}$ does not depend on the data. The second Kalman filter run is needed to produce the state estimates $\hat{\boldsymbol{\alpha}}_{t|t}^{\boldsymbol{b}}(\hat{\boldsymbol{\theta}})$ based on $\{\boldsymbol{Y}_{\boldsymbol{1}}^{\boldsymbol{b}}, \ldots, \boldsymbol{Y}_{\boldsymbol{T}}^{\boldsymbol{b}}\}$ 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 $\boldsymbol{P}_{t|t}(\hat{\boldsymbol{\theta}})$ for future use in (3.7).
2. Using either the parametric or non-parametric method, generate a bootstrap sample $\{\boldsymbol{Y}_{\boldsymbol{1}}^{\boldsymbol{b}}, \ldots, \boldsymbol{Y}_{\boldsymbol{T}}^{\boldsymbol{b}}\}$ and apply the simulation smoother correction to it if the model is non-stationary, as described in the introduction to Section 3.
3. Estimate bootstrap hyperparameter estimates $\hat{\boldsymbol{\theta}}^{\boldsymbol{b}}$ from the newly generated bootstrap dataset. Run the Kalman filter once to get $\hat{\boldsymbol{\alpha}}_{t|t}^{\boldsymbol{b}}(\hat{\boldsymbol{\theta}}^{\boldsymbol{b}})$ and $\boldsymbol{P}_{t|t}(\hat{\boldsymbol{\theta}}^{\boldsymbol{b}})$, and another time to obtain $\hat{\boldsymbol{\alpha}}_{t|t}^{\boldsymbol{b}}(\hat{\boldsymbol{\theta}})$, as described under (3.7).
4. Repeat steps 2-3 B times.
5. Estimate the MSE using (3.7).

Pfeffermann and Tiller (2005) note that, in the case of the parametric bootstrap, the second Kalman filter run can be avoided because the true state vector is generated (and thus known) for every bootstrap series. Thus, the state estimates $\hat{\boldsymbol{\alpha}}_{t|t}^{\boldsymbol{b}}(\hat{\boldsymbol{\theta}})$ in (3.7) can be replaced by the true vector $\boldsymbol{\alpha}_{\boldsymbol{t}}^{\boldsymbol{b}}$ to obtain the following MSE estimator:

$$\widehat{\boldsymbol{MSE}}_{t|t}^{\boldsymbol{PT1}} = \boldsymbol{P}_{t|t}(\hat{\boldsymbol{\theta}}) - \frac{\boldsymbol{1}}{\boldsymbol{B}}\sum_{\boldsymbol{b=1}}^{\boldsymbol{B}} \boldsymbol{P}_{t|t}(\hat{\boldsymbol{\theta}}^{\boldsymbol{b}}) + \frac{\boldsymbol{1}}{\boldsymbol{B}}\sum_{\boldsymbol{b=1}}^{\boldsymbol{B}} [\hat{\boldsymbol{\alpha}}_{t|t}^{\boldsymbol{b}}(\hat{\boldsymbol{\theta}}^{\boldsymbol{b}}) - \boldsymbol{\alpha}_{\boldsymbol{t}}^{\boldsymbol{b}}][\hat{\boldsymbol{\alpha}}_{t|t}^{\boldsymbol{b}}(\hat{\boldsymbol{\theta}}^{\boldsymbol{b}}) - \boldsymbol{\alpha}_{\boldsymbol{t}}^{\boldsymbol{b}}]'. \tag{3.8}$$

In this formulation, there is only one $\boldsymbol{P}_{t|t}(\hat{\boldsymbol{\theta}})$ in the right-hand side of (3.8). This is due to the fact that the new term $\boldsymbol{E}_{\boldsymbol{B}}[\hat{\boldsymbol{\alpha}}_{t|t}^{\boldsymbol{b}}(\hat{\boldsymbol{\theta}}^{\boldsymbol{b}}) - \boldsymbol{\alpha}_{\boldsymbol{t}}^{\boldsymbol{b}}][\hat{\boldsymbol{\alpha}}_{t|t}^{\boldsymbol{b}}(\hat{\boldsymbol{\theta}}^{\boldsymbol{b}}) - \boldsymbol{\alpha}_{\boldsymbol{t}}^{\boldsymbol{b}}]'$, corresponding to the last term on the right-hand side of (3.8), can itself be decomposed, in the same fashion as in (3.4), into the measure of parameter uncertainty $\boldsymbol{E}_{\boldsymbol{B}}[\hat{\boldsymbol{\alpha}}_{t|t}^{\boldsymbol{b}}(\hat{\boldsymbol{\theta}}^{\boldsymbol{b}}) - \hat{\boldsymbol{\alpha}}_{t|t}^{\boldsymbol{b}}(\hat{\boldsymbol{\theta}})][\hat{\boldsymbol{\alpha}}_{t|t}^{\boldsymbol{b}}(\hat{\boldsymbol{\theta}}^{\boldsymbol{b}}) - \hat{\boldsymbol{\alpha}}_{t|t}^{\boldsymbol{b}}(\hat{\boldsymbol{\theta}})]'$ and the filter uncertainty term $\boldsymbol{P}_{t|t}^{\boldsymbol{b}}(\hat{\boldsymbol{\theta}}) = \boldsymbol{E}[\hat{\boldsymbol{\alpha}}_{t|t}^{\boldsymbol{b}}(\hat{\boldsymbol{\theta}}) - \boldsymbol{\alpha}_{\boldsymbol{t}}^{\boldsymbol{b}}][\hat{\boldsymbol{\alpha}}_{t|t}^{\boldsymbol{b}}(\hat{\boldsymbol{\theta}}) - \boldsymbol{\alpha}_{\boldsymbol{t}}^{\boldsymbol{b}}]'$, $\hat{\boldsymbol{\theta}}$ being the true parameter vector the bootstrap state variables $\boldsymbol{\alpha}_{\boldsymbol{t}}^{\boldsymbol{b}}$ are generated with. However, the bootstrap average term $\frac{1}{B}\sum_{\boldsymbol{b=1}}^{\boldsymbol{B}}[\hat{\boldsymbol{\alpha}}_{t|t}^{\boldsymbol{b}}(\hat{\boldsymbol{\theta}}) - \boldsymbol{\alpha}_{\boldsymbol{t}}^{\boldsymbol{b}}][\hat{\boldsymbol{\alpha}}_{t|t}^{\boldsymbol{b}}(\hat{\boldsymbol{\theta}}) - \boldsymbol{\alpha}_{\boldsymbol{t}}^{\boldsymbol{b}}]'$ replacing $\boldsymbol{P}_{t|t}(\hat{\boldsymbol{\theta}})$ 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 decomposition of the term $\boldsymbol{E}_{\boldsymbol{B}}[\hat{\boldsymbol{\alpha}}_{t|t}^{\boldsymbol{b}}(\hat{\boldsymbol{\theta}}^{\boldsymbol{b}}) - \boldsymbol{\alpha}_{\boldsymbol{t}}^{\boldsymbol{b}}][\hat{\boldsymbol{\alpha}}_{t|t}^{\boldsymbol{b}}(\hat{\boldsymbol{\theta}}^{\boldsymbol{b}}) - \boldsymbol{\alpha}_{\boldsymbol{t}}^{\boldsymbol{b}}]'$ according to (3.2) will also contain a non-zero expectation of the cross-terms: $\boldsymbol{E}\{[\hat{\boldsymbol{\alpha}}_{t|t}^{\boldsymbol{b}}(\hat{\boldsymbol{\theta}}) - \boldsymbol{\alpha}_{\boldsymbol{t}}^{\boldsymbol{b}}][\hat{\boldsymbol{\alpha}}_{t|t}^{\boldsymbol{b}}(\hat{\boldsymbol{\theta}}^{\boldsymbol{b}}) - \hat{\boldsymbol{\alpha}}_{t|t}^{\boldsymbol{b}}(\hat{\boldsymbol{\theta}})]\}$. 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 $\frac{\boldsymbol{1}}{\boldsymbol{B}}\sum_{\boldsymbol{b=1}}^{\boldsymbol{B}}[\hat{\boldsymbol{\alpha}}_{t|t}^{\boldsymbol{b}}(\hat{\boldsymbol{\theta}}) - \boldsymbol{\alpha}_{\boldsymbol{t}}^{\boldsymbol{b}}][\hat{\boldsymbol{\alpha}}_{t|t}^{\boldsymbol{b}}(\hat{\boldsymbol{\theta}}) - \boldsymbol{\alpha}_{\boldsymbol{t}}^{\boldsymbol{b}}]'$ exhibited large departures (in both directions) 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 term $\frac{1}{B}\sum_{b=1}^{B}[\hat{\boldsymbol{\alpha}}_{t|t}^{b}(\hat{\boldsymbol{\theta}})-\boldsymbol{\alpha}_{t}^{b}][\hat{\boldsymbol{\alpha}}_{t|t}^{b}(\hat{\boldsymbol{\theta}})-\boldsymbol{\alpha}_{t}^{b}]'$ that replaces $\boldsymbol{P}_{t|t}(\hat{\boldsymbol{\theta}})$ in (3.8) may not be sufficiently close to it. For this reason, both parametric (denoted as PT1) and non-parametric (PT2) bootstraps in this application rely on the estimator in (3.7).
A few words have to be said about the role of the simulation smoother of Durbin and Koopman (2002). We suggest that it should be used at the bootstrap series generation step. Without it, the bootstrap hyperparameter distribution obtained from uncorrected series in a non-stationary model could be very different from what it should be for a particular realisation of the data at hand. At least in the case of the DLFS, omitting the simulation smoother step resulted in bootstrap hyperparameter distributions having a much wider range than that in distributions obtained with the help of the simulation smoother. Moreover, such bootstrap hyperparameter distributions obtained from uncorrected series in the DLFS are centred around values that are much larger than hyperparameter values the series have been generated with. This results in an excessively large bootstrap average $\frac{1}{B}\sum_{b=1}^{B}\boldsymbol{P}_{t|t}(\hat{\boldsymbol{\theta}}^{b})$ (relatively to $\boldsymbol{P}_{t|t}(\hat{\boldsymbol{\theta}})$) and, subsequently, in MSE-estimates that are even lower than the naive ones. The term $\frac{1}{B}\sum_{b=1}^{B}[\hat{\boldsymbol{\alpha}}_{t|t}^{b}(\hat{\boldsymbol{\theta}}^{b})-\hat{\boldsymbol{\alpha}}_{t|t}^{b}(\hat{\boldsymbol{\theta}})][\hat{\boldsymbol{\alpha}}_{t|t}^{b}(\hat{\boldsymbol{\theta}}^{b})-\hat{\boldsymbol{\alpha}}_{t|t}^{b}(\hat{\boldsymbol{\theta}})]'$ also becomes very unstable over time and excessively large compared to when the simulation smoother is used, but that does not compensate for the negative bias obtained from (3.7) without the simulation smoother.

3.3 Asymptotic Approximation

An asymptotic approximation (AA) to the true MSE in equation (3.2) was developed by Hamilton (1986) and can be expressed as an expectation over the hyperparameter joint asymptotic distribution $\pi(\boldsymbol{\theta}|\boldsymbol{Y})$, conditional on the given dataset $\boldsymbol{Y}\equiv\{\boldsymbol{Y_1},\ldots,\boldsymbol{Y_T}\}$. In the present application, the part of the hyperparameter vector that is estimated by the ML-method, $(\boldsymbol{\theta_\sigma})$, depends on the value of the autoregressive parameter ρ. Therefore, the hyperparameter joint asymptotic distribution has the following form: $\pi(\boldsymbol{\theta}|\boldsymbol{Y})=\pi(\rho|\boldsymbol{Y})\pi(\boldsymbol{\theta_\sigma}|\rho,\boldsymbol{Y})$. The MSE is approximated as follows:

$$\begin{aligned}\boldsymbol{MSE}_{t|t}=&\boldsymbol{E}_{\pi(\boldsymbol{\theta}|\boldsymbol{Y})}[\boldsymbol{P}_{t|t}(\boldsymbol{\theta},\boldsymbol{Y})]+\\&\boldsymbol{E}_{\pi(\boldsymbol{\theta}|\boldsymbol{Y})}[(\hat{\boldsymbol{\alpha}}_{t|t}(\boldsymbol{\theta},\boldsymbol{Y})-\hat{\boldsymbol{\alpha}}_{t|t}(\boldsymbol{Y}))(\hat{\boldsymbol{\alpha}}_{t|t}(\boldsymbol{\theta},\boldsymbol{Y})-\hat{\boldsymbol{\alpha}}_{t|t}(\boldsymbol{Y}))'],\end{aligned}\tag{3.9}$$

where $\boldsymbol{E}_{\pi(\boldsymbol{\theta}|\boldsymbol{Y})}$ is an expectation taken over the hyperparameter joint asymptotic distribution $\pi(\boldsymbol{\theta}|\boldsymbol{Y})$, and $\hat{\boldsymbol{\alpha}}_{t|t}(\boldsymbol{Y})$ are the state vector estimates when the hyperparameters are not known (i.e. $\boldsymbol{E}_{\pi(\boldsymbol{\theta}|\boldsymbol{Y})}[\hat{\boldsymbol{\alpha}}_{t|t}(\boldsymbol{\theta},\boldsymbol{Y})]$).
Distribution $N(\hat{\rho},Var(\hat{\rho}))$ is taken to denote ρ's asymptotic distribution $\pi(\rho|\boldsymbol{Y})$, from which random ρ-realisations are drawn. Generally, the sampling distribution of the correlation coefficient has a complex form, but it may be well approximated by a normal distribution, which was the case in this application (the normal distribution fitted both the simulated and the bootstrap ρ-distribution very well). From equation (3) in Bartlett (1946), and using the fact that the autoregressive coefficient in an AR(1) process is equal to the correlation for lag 1, the variance estimator of $\hat{\rho}$ becomes: $Var(\hat{\rho})\approx(1-\hat{\rho}^2)1/T$. In the case of the DLFS, where $\hat{\rho}=0.208$, this means that $\widehat{Var}(\hat{\rho})\approx 0.96(1/T)$. Taking into account the fact that ρ's standard error is used for making draws from the asymptotic distribution, and that the square root is a concave function, the sample standard deviation would be an underestimate. Therefore, making ρ-draws by means of $1/\sqrt{T}$ as the asymptotic distribution's standard deviation would be a reasonable choice.
After a ρ-value is drawn from $\pi(\rho|\boldsymbol{Y})$, the other hyperparameters are re-estimated to obtain $\hat{\boldsymbol{\theta}}_{\boldsymbol{\sigma}}|\rho$ and the information matrix $\hat{\boldsymbol{I}}(\hat{\boldsymbol{\theta}}_{\boldsymbol{\sigma}}|\rho_{\pi(\rho|\boldsymbol{Y})})$. Finally, a $\boldsymbol{\theta_\sigma}$-draw is made from distribution $\sqrt{T}(\boldsymbol{\theta_\sigma}-\hat{\boldsymbol{\theta}}_{\boldsymbol{\sigma}}|\rho_{\pi(\rho|\boldsymbol{Y})})\sim MN(\boldsymbol{0},T\hat{\boldsymbol{I}}^{-1}(\hat{\boldsymbol{\theta}}_{\boldsymbol{\sigma}}|\rho_{\pi(\rho|\boldsymbol{Y})}))$. The Kalman filter is run again using ρ- and

$\boldsymbol{\theta_\sigma}$-realisations to obtain the state estimates $\hat{\boldsymbol{\alpha}}_{\boldsymbol{t|t}}(\boldsymbol{\theta}_{\boldsymbol{\pi(\theta|Y)}}, \boldsymbol{Y})$ and their MSEs $\boldsymbol{P}_{\boldsymbol{t|t}}(\boldsymbol{\theta}_{\boldsymbol{\pi(\theta|Y)}})$ ($\boldsymbol{\theta}_{\boldsymbol{\pi(\theta|Y)}}$-draws are further denoted as $\boldsymbol{\theta^a}$). The procedure is repeated until B $\boldsymbol{\theta^a}$-draws are obtained, whereafter (3.9) is obtained 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 $\sqrt{T}(\boldsymbol{\theta}_{\boldsymbol{\pi(\theta|Y)}} - \hat{\boldsymbol{\theta}}) \sim MN(\mathbf{0}, T\hat{\boldsymbol{I}}^{-1}(\hat{\boldsymbol{\theta}}))$.
The first term in (3.9) can be approximated by the average value of the Kalman filter variance $\boldsymbol{P}_{\boldsymbol{t|t}}$ across B realizations of the hyperparameter vector, and the second term by the variance of the state vector estimates across the same B realisations. An asymptotic approximation for the MSE could therefore be obtained in the following way:

$$\widehat{\boldsymbol{MSE}}^{\boldsymbol{AA}}_{\boldsymbol{t|t}} = \frac{1}{B}\sum_{\boldsymbol{a=1}}^{\boldsymbol{B}} \boldsymbol{P}_{\boldsymbol{t|t}}(\boldsymbol{\theta^a}) + \\ + \frac{1}{B}\sum_{\boldsymbol{a=1}}^{\boldsymbol{B}} [\hat{\boldsymbol{\alpha}}_{\boldsymbol{t|t}}(\boldsymbol{\theta^a}, \boldsymbol{Y}) - \bar{\hat{\boldsymbol{\alpha}}}_{\boldsymbol{t|t}}][\hat{\boldsymbol{\alpha}}_{\boldsymbol{t|t}}(\boldsymbol{\theta^a}, \boldsymbol{Y}) - \bar{\hat{\boldsymbol{\alpha}}}_{\boldsymbol{t|t}}]'. \tag{3.10}$$

where $\boldsymbol{\theta^a}$ is the a-th draw from the $\pi(\boldsymbol{\theta}|\boldsymbol{Y})$ asymptotic distribution. As Hamilton (1986) suggests, the sample average $\bar{\hat{\boldsymbol{\alpha}}}_{\boldsymbol{t|t}} = \frac{\boldsymbol{1}}{\boldsymbol{B}}\sum_{\boldsymbol{a=1}}^{\boldsymbol{B}} \hat{\boldsymbol{\alpha}}_{\boldsymbol{t|t}}(\boldsymbol{\theta}^{\boldsymbol{a}}_{\boldsymbol{\pi(\theta|Y)}}, \boldsymbol{Y})$ can replace $\hat{\boldsymbol{\alpha}}_{\boldsymbol{t|t}}(\boldsymbol{Y})$ in (3.9). Further, he states, such a decomposition of the total uncertainty into the filter and parameter uncertainty resembles the well-known decomposition: $var(X) = E[var(X|Y)] + var[E(X|Y)]$. Obviously, this MSE-estimator is entirely based on the assumption of asymptotic normality of 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 happen 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 very 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 Estimation

The performance of the five MSE estimation 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 lengths 48 and 80 months, and on 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 DLFS model used by Statistics Netherlands for the number of unemployed. MSEs for each of these series are estimated 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 delivered by the five methods and 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, at least 10000 simulations are needed.

Equations (2.4)-(2.5) and the estimated parameter vector $\hat{\boldsymbol{\theta}}$ are used to generate artificial data series $\boldsymbol{Y}_t^s, s = 1, \ldots S$, S=1000. State disturbances (recall that survey errors are also modelled as state variables) are randomly drawn from their joint normal distribution $N(\mathbf{0}, \boldsymbol{\Omega}(\hat{\boldsymbol{\theta}}_{\boldsymbol{\sigma}}))$, and series are generated using the Kalman filter recursion. Since the system is non-stationary, the generated series $\boldsymbol{Y}_t^s$ may take on negative or implausibly large numbers of the unemployed. In order to avoid an excessively large number of series with negative values, the state variables recursion is launched from the states' smoothed estimates at one of the highest points of the observed series. Further, the first 30 time points are discarded in order to prevent that the series start at the same time-point. With an assumption that unemployment in the Netherlands will not exceed 15 percent of the total labour force, the simulation data set is restricted to contain only series with values between 0 and 1 mln of unemployed (this value comprised about 15 percent of the Dutch labour force in 2010); other series are discarded. 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-based standard errors z_t^{t-j}, which are the time-dependent elements of the design-matrix $\boldsymbol{Z}_t$.

Every series of simulated GREG point-estimates needs its own series of simulated design-based standard error estimates, z_t^{t-j}s. The original known design-based standard error estimates $\sqrt{\widehat{Var}(Y_t^{t-j})}$ would not be suitable for this simulation because the sampling error variance is proportional to the corresponding point-estimate. The following variance function is used to generate design-based variances for the simulated series of point-estimates (see Appendix B for details):

$$\begin{aligned} ln[\widehat{Var}(Y_t^t)] = ln[(z_t^t)^2] &= c + \beta^0 ln(l_t^t) + \epsilon_t^t,\ \epsilon_t^t \sim N(0, (\sigma_\epsilon^0)^2); \\ ln[\widehat{Var}(Y_t^{t-j})] = ln[(z_t^{t-j})^2] &= \psi^j ln[(z_{t-3}^{t-j})^2] + \beta^j ln(l_t^{t-j}) + \epsilon_t^{t-j}, \\ &\epsilon_t^{t-j} \sim N(0, (\sigma_\epsilon^j)^2), j = \{3, 6, 9, 12\}. \end{aligned} \tag{4.1}$$

The regression coefficients in (4.1) are time-invariant and are obtained by regressing $ln(z_t^{t-j})^2$ on $ln(l_t^t)$ and $ln((z_{t-3}^{t-j})^2)$ from the original DLFS series. The superscripts are used to denote the wave these coefficients belong to. The coefficient estimates are presented in Table 4.1. They are used to generate design-based standard errors z_t^{t-j} for every simulated series and for each series length considered.

Table 4.1 Regression estimates for the design-based 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
$\hat{\psi}^j$	-	0.859	0.786	0.749	0.751
$\hat{\sigma}_\epsilon^j$	0.202	0.265	0.228	0.225	0.267

The simulation proceeds as follows. For each s, 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-based standard errors $z_{t,s}^{t-j}$ according to the process defined by (4.1). 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. All these are saved in $\hat{\boldsymbol{\theta}}_{\boldsymbol{s}}$ and used by the Kalman filter to produce the state estimates $\hat{\boldsymbol{\alpha}}_{\boldsymbol{t,s}}$. Further, $\hat{\boldsymbol{\theta}}_{\boldsymbol{s}}$ is used to generate bootstrap samples. Note that the same set of design-based standard errors $\boldsymbol{z}_{\boldsymbol{t,s}}$ is used to generate all bootstrap series within simulation s.

In order to obtain the true MSEs, the DLFS model is simulated a large number of times ($M = 50000$), with each of these replications being restricted to the same limits as before, i.e. between zero and 1 mln of the 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:

$$\boldsymbol{MSE}_{\boldsymbol{t}}^{\boldsymbol{true}} = \frac{\mathbf{1}}{\boldsymbol{M}} \sum_{\boldsymbol{m=1}}^{\boldsymbol{M}} [(\hat{\boldsymbol{\alpha}}_{\boldsymbol{m,t}}(\hat{\boldsymbol{\theta}}_{\boldsymbol{m}}) - \boldsymbol{\alpha}_{\boldsymbol{m,t}})(\hat{\boldsymbol{\alpha}}_{\boldsymbol{m,t}}(\hat{\boldsymbol{\theta}}_{\boldsymbol{m}}) - \boldsymbol{\alpha}_{\boldsymbol{m,t}})'] . \tag{4.2}$$

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

5 Model Specification and MSE Estimation

5.1 Alternative Model Specifications for the DLFS

Model adequacy in the STS framework is usually examined by means of formal diagnostic tests for normality, homogeneity and independence of the standardised innovations. These tests do not detect any problems with the DLFS model. However, asymptotic confidence intervals of the hyperparameters, together with their distribution across a large number of the DLFS model replications give a clear indication that the model tends to be overspecified in a sense that some state variables may be modelled as time-invariant as their hyperparameters tend to zero. Therefore, replicating the model in a Monte-Carlo simulation can be a very useful tool in checking for redundant hyperparameters. Redundant parametrization in nested models is usually checked through the likelihood-ratio test. However, the outcome of this test will depend on the particular point-estimate of the suspect hyperparameter, rather than on the whole distribution of the hyperparameter in question.

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 implausibly large variance and disturbed estimation of other marginally significant hyperparameters (the seasonal and RGB disturbance variances) in the case of the DLFS. Removing the irregular component ε_t from the model has mitigated the instability in the two above-mentioned hyperparameters. 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. Alterations of unemployment levels are usually gradual and therefore must be largely incorporated into the stochastic trend movements. The other three models are special cases of Model 1, i.e. all with the irregular component ε_t removed (see Table 5.1).

Table 5.1 Hyperparameters estimated in the four versions of the DLFS model; the disturbance variances are estimated on a log-scale

Models	Description	Parameters estimated
M1	complete model	$\rho, \sigma^2_{\eta_R}, \sigma^2_{\bar{\omega}}, \sigma^2_{\eta_\lambda}, \sigma^2_{\nu_t^t}, \sigma^2_{\nu_t^{t-3}}, \sigma^2_{\nu_t^{t-6}}, \sigma^2_{\nu_t^{t-9}}, \sigma^2_{\nu_t^{t-12}}$
M2	seasonal time-independent	$\rho, \sigma^2_{\eta_R}, \sigma^2_{\eta_\lambda}, \sigma^2_{\nu_t^t}, \sigma^2_{\nu_t^{t-3}}, \sigma^2_{\nu_t^{t-6}}, \sigma^2_{\nu_t^{t-9}}, \sigma^2_{\nu_t^{t-12}}$
M3	RGB time-independent	$\rho, \sigma^2_{\eta_R}, \sigma^2_{\bar{\omega}}, \sigma^2_{\nu_t^t}, \sigma^2_{\nu_t^{t-3}}, \sigma^2_{\nu_t^{t-6}}, \sigma^2_{\nu_t^{t-9}}, \sigma^2_{\nu_t^{t-12}}$
M4	seasonal, RGB fixed	$\rho, \sigma^2_{\eta_R}, \sigma^2_{\nu_t^t}, \sigma^2_{\nu_t^{t-3}}, \sigma^2_{\nu_t^{t-6}}, \sigma^2_{\nu_t^{t-9}}, \sigma^2_{\nu_t^{t-12}}$

The rationale behind studying the other three models becomes clear when inspecting the hyperparameter distribution under Model 1 after a large number of replications. The simulation

has shown that variances of the stochastic terms of the seasonal and, in particular, of the RGB component are often estimated to be close to zero. This causes bi-modality in the distribution of these variance estimates with a significant mass concentrated close to zero. Apart from that, an attempt to estimate both $ln(\hat{\sigma}_{\omega}^2)$ and $ln(\hat{\sigma}_{\eta_\lambda}^2)$, as in Model 1, distorts the other hyperparameters' ML estimator distribution that is expected to be normal. For instance, normality in $ln(\hat{\sigma}_{\nu_t^{t-6}}^2)$, $ln(\hat{\sigma}_{\nu_t^{t-9}}^2)$ and $ln(\hat{\sigma}_{\nu_t^{t-12}}^2)$ is severely violated with extreme outliers and/or a huge kurtosis (see Fig. C.1 in Appendix C, where the x-axis is extended due to the outliers), while the corresponding variances are less likely to exhibit extreme values as they are supposed to fluctuate around 1. Making the seasonal component time-invariant, as in Model 2, hardly changes the situation for the trend 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. C.2). By contrast, under both models where the RGB-component is fixed over time (Models 3 and 4), all hyperparameter estimates corresponding to the survey error component have turned out to be normally distributed, see Fig. C.3 and Fig. C.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_{ω}^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. However, since the RGB itself depends on the numbers of unemployed, its variance hyperparameter is retained for production purposes at Statistics Netherlands to secure sufficient flexibility against considerable changes in the underlying process.

The distribution of the estimator of the survey error autoregressive parameter ρ across the 1000 simulated series does not seem to be affected by model reformulations: it approaches the normal distribution quite closely and ranges between 0 and 0.4 when $T = 114$, which is in line with the approximation of its asymptotic distribution mentioned in Subsection 3.3. The range is slightly wider for the shorter time series and narrower when $T = 200$. The simulation procedure described in the previous section and the analysis of bootstrap methods that follows is performed separately for all the four models.

5.2 MSE Estimation

The focus of this simulation study is MSE estimation for the trend and for the population signal, the latter being the sum of the trend and seasonal components. The performance of the Kalman filter at estimated parameter values, as well as of the five MSE estimation methods mentioned in Section 3 is evaluated by use of the MSE relative bias. First, the filtered MSE estimates from (3.3), (3.7) and (3.10) are averaged over 1000 simulations (where the average is denoted with a bar: $\overline{MSE}_{t|t}$, whereas the Kalman filter MSE estimates are averaged over 10000 simulations, as mentioned at the beginning of Section 4. These averaged filtered MSE estimates for Model 3 (except for the AA-method, see below why) are depicted in Fig. 5.1 - 5.4 for $T = 48$, $T = 80$, $T = 114$, and $T = 200$, respectively, skipping the first $d = 30$ time points of the sample (d should exceed the diffuse sample r, to eliminate the filter initialisation effect). Note that all the analysis is based on filtered, rather than smoothed estimates, because filtered estimates better mimic the process of official figures production. The percentage relative bias is calculated as $RB_t^f = 100\left(\overline{MSE}_{t|t}^f / MSE_{t|t}^{true} - 1\right)$, where f defines a particular estimation method and $MSE_{t|t}^{true}$ is defined in (4.2). 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 5.2, 5.3, 5.4, and 5.5.

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 $\boldsymbol{I}^{-1}(\hat{\boldsymbol{\theta}}_{\boldsymbol{\sigma}}|\boldsymbol{\rho}^{\boldsymbol{a}}))$ 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 method, but significantly worse than the PT2 method's performance.

The main conclusions from the simulation study are as follows.

1. For $T = 48$, and when averaged over time (starting from $t = 31$), the relative bias of the signal MSE obtained with the use of the Kalman filter is around -7 percent. This bias tends to decrease as the series length increases. The Kalman filter (KF) bias is quite small for the case of $T = 200$, such that none of the estimation methods offers an improvement over the KF-based MSE-estimates. One could still apply the best estimation method with positive biases in order to get a range of values containing the true MSE.

2. As can be immediately observed, the use of the RR-bootstrap results in a negative bias, whereas the use of the PT-method produces a positive bias. Contrary to 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-based MSE estimates, both the parametric and non-parametric ones, have larger negative biases than the uncorrected KF-estimates across all the models and series lengths (except for RR2 in Model 4 when $T = 48$, and in Model 1 when $T = 80$ and $T = 114$). While the PT-bootstrap method is shown 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. The PT-methods have never produced negative biases for the DLFS, which makes these methods conservative (except in Model 4 when $T = 48$, with the negative bias still being smaller than that of the Kalman filter). Another striking outcome for $T = 48$ is that the PT2 positive bias and the RR negative bias take on very large values in Model 3. However, with such a short series length and with so many non-stationary components like in the DLFS model, it is difficult to obtain reliable estimates from non-parametric bootstrap methods, since the burn-in period (or the diffuse sample) necessary for the non-parametric generation of the series takes more than a quarter of the series length (13 months out of 48).

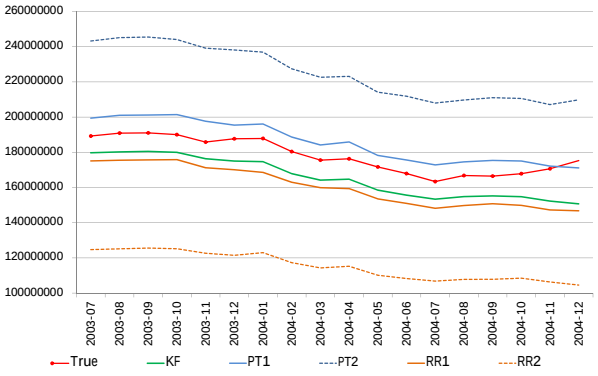

Figure 5.1 Signal MSE comparison for Model 3, T=48 months

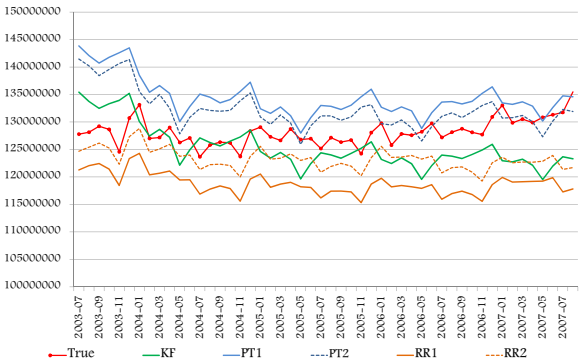

Figure 5.2 Signal MSE comparison for Model 3, T=80 months

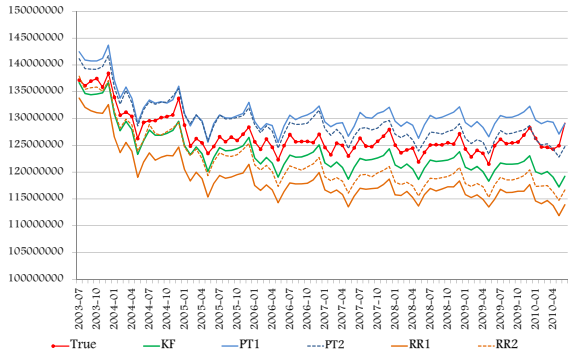

Figure 5.3 Signal MSE comparison for Model 3, T=114 months

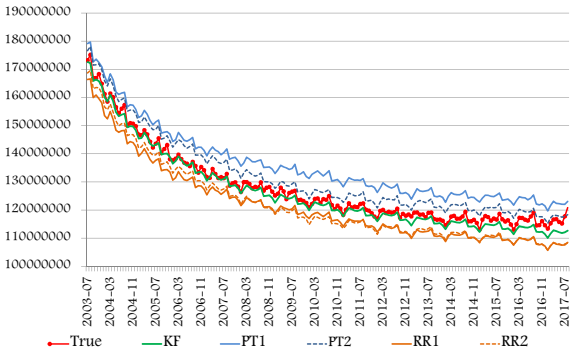

Figure 5.4 Signal MSE comparison for Model 3, T=200 months

Table 5.2 Average percent bias of the MSE estimators under the DLFS model, $t = \{31, ..., T\}$, T=48

	Signal				Trend				Seasonal			
Models	M1	M2	M3	M4	M1	M2	M3	M4	M1	M2	M3	M4
KF	N/A	N/A	-7.1	-7.6	N/A	N/A	-6.5	-6.6	N/A	N/A	-6.7	-7.0
PT1	N/A	N/A	**4.4**	**1.4**	N/A	N/A	8.7	6.4	N/A	N/A	4.9	2.4
PT2	N/A	N/A	26.2	-4.4	N/A	N/A	22.4	-3.1	N/A	N/A	25.6	-4.6
RR1	N/A	N/A	-9.8	-10.8	N/A	N/A	-13.9	-13.8	N/A	N/A	-9.5	-10.1
RR2	N/A	N/A	-35.3	-5.6	N/A	N/A	-29.9	-3.2	N/A	N/A	-29.7	-5.1

Table 5.3 Average percent bias of the MSE estimators under the DLFS model, $t = \{31, ..., T\}$, 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	-4.2	5.3	-4.1	-4.6	-5.4	18.6	-4.7	-4.1	-4.3

Table 5.4 Average percent bias of the MSE estimators under the DLFS model, $t = \{31, ..., T\}$, 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 5.5 Average percent bias of the MSE estimators under the DLFS model, $t = \{31, ..., T\}$, 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

Table 5.6 Average estimated variance and MSE of the MSE estimators for the numbers of unemployed under the DLFS model (divided by 10^{15}), $t = \{31, ..., T\}$, T=48

	Signal				Trend				Seasonal			
Models	M3		M4		M3		M4		M3		M4	
	Var_{MSE}	MSE_{MSE}	Var_{MSE}	MSE_{MSE}	Var_{MSE}	MSE_{MSE}	Var_{MSE}	MSE_{MSE}	Var_{MSE}	MSE_{MSE}	Var_{MSE}	MSE_{MSE}
PT1	3.39	3.46	3.64	3.66	3.61	3.83	3.67	3.81	0.59	0.61	0.64	0.65
PT2	5.03	7.26	3.03	3.10	4.02	5.27	2.56	2.61	1.00	1.50	0.52	0.54
RR1	2.51	2.83	2.68	3.06	2.03	2.51	2.13	2.62	0.44	0.51	0.48	0.55
RR2	1.59	5.93	2.74	2.85	1.52	3.97	2.50	2.56	0.55	1.28	0.50	0.52

Table 5.7 Average estimated variance and MSE of the MSE estimators for the numbers of unemployed under the DLFS model (divided by 10^{15}), $t = \{31, ..., T\}$, T=80

	Signal				Trend				Seasonal			
Models	M3		M4		M3		M4		M3		M4	
	Var_{MSE}	MSE_{MSE}	Var_{MSE}	MSE_{MSE}	Var_{MSE}	MSE_{MSE}	Var_{MSE}	MSE_{MSE}	Var_{MSE}	MSE_{MSE}	Var_{MSE}	MSE_{MSE}
PT1	2.24	2.29	2.43	2.52	1.82	1.91	1.97	2.09	0.27	0.30	0.27	0.31
PT2	2.20	2.23	2.14	2.18	1.71	1.74	1.66	1.69	0.27	0.28	0.27	0.29
RR1	1.86	1.95	1.74	1.82	1.42	1.56	1.33	1.46	0.22	0.23	0.22	0.23
RR2	1.98	2.01	1.94	1.97	1.57	1.60	1.49	1.54	0.23	0.23	0.23	0.23

3. For the series of lengths $T = 114$ and $T = 80$, the positive biases produced by the PT2-method slightly exceed the KF-biases in absolute value in 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 value. For $T = 48$, bootstrap results are presented only for Models 3 and 4 (Models 1 and 2 that tend to be overspecified are not considered due to numerical problems). As could be expected, the biases are larger for this series length: the negative KF and RR biases become larger in absolute value, and so do the PT positive biases, with an exception of the above-mentioned result for PT2 in Model 4.
The signal MSE of Model 3, which could be considered as the better option for the production of official DLFS figures, is best estimated by the PT2 approach, with the relative bias of 1.4 and 1.9 percent for $T = 80$ and $T = 114$, respectively. The PT2-bootstrap method also seems to be the best method for $T = 200$, but, as already noted, the negative KF biases are already quite small for series of this length. For very short series like $T = 48$, the parametric PT1 bootstrap seems to be the best option.
4. For both the PT- and RR-methods (except for RR2 in Model 4, $T = 48$), the absolute values of the 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 distorted normality of the error distribution in the models. Therefore, non-parametric bootstraps should be preferred when time series are of a reasonable length.
5. Apart from the bias of the MSE estimators, their variability may also give important insights into their reliability. To our knowledge, this has not been yet presented in the statistical literature. Tables 5.6 and 5.7 contain variances and MSEs of the four bootstrap MSE estimators for the signal, trend, and seasonal components for the two most interesting series lengths: $T = 48$ and $T = 80$ months (Models 1 and 2, as well as the asymptotic approximation, are not considered due to the aforementioned numerical problems). For both Model 3 and Model 4, the MSEs of the two PT MSE estimators are slightly larger than those MSEs of the two RR MSE estimators. The RR methods' seemingly superior performance, reflected by smaller MSEs of MSEs, is due to the smaller variance of the RR MSE estimators. The biases, however, are large enough to bring the MSEs of MSE almost to the level of those of the PT estimators. More importantly, the biases of the RR MSE estimators are mostly negative, often exceeding those of the Kalman filter. This phenomenon makes RR-bootstraps hardly applicable in practice.
In order to see if the STS model-based approach still offers more precise predictors than the design-based variance estimates even after correcting for the hyperparameter uncertainty in the original series ($T = 114$), percentage reductions in the standard errors of the GREG estimates are presented in Table 5.8. These reductions result from applying the DLFS model to the GREG estimates without correcting for hyperparameter uncertainty (KF in Table 5.8 stands for the Kalman filter), as well as after the uncertainty has been taken into account by use of the five MSE estimation 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 signal 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 5.8 as well.
With a 18 percent reduction in the GREG SEs, it may first seem that the AA-method accounts for the parameter uncertainty best out of all other methods, as it exhibits the smallest reduction in the GREG standard error estimates. However, keeping in mind that the AA may on average have some severe positive biases (compare 18 to 20 percent reduction based on the true MSEs), especially in the case of small hyperparameters, one should feel more confident with the use of the PT1 parametric approach that offers about a 20 percent reduction in the estimated GREG

standard errors. This means that the model-based approach even after accounting for the hyperparameter uncertainty offers a significant variance reduction compared to the traditional design-based approach.

Table 5.8 Reductions in the GREG SE estimates of the original DLFS series, averaged over time (d=30), and the percentage increase in the KF-based SEs after applying the MSE correction (in parentheses).

	Model 1	Model 2	Model 3	Model 4
KF	-24.1	-24.1	-24.5	-24.5
True	-20.0 (5.56)	-20.1 (5.5)	-20.6 (5.4)	-20.7 (5.3)
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 Concluding remarks

Flawless model-based estimation is not restricted to model construction and model diagnostics. This work stresses the importance of two additional stages: model simulation and alleviating a negative bias in MSE estimates.

The simulation presented in this paper illustrates that attempting to estimate redundant hyperparameters may render the other hyperparameters unreliable, as reflected by heavy distortions in the normality of the distribution of the ML estimator. For the DLFS model, the simulation shows that it may be worth considering a more restricted version of the model, where the variance of the RGB and population noise are set equal to zero.

Neglected hyperparameter uncertainty in STS model-based MSEs becomes a problem when series are of a relatively short length. This causes the MSE estimates of domain predictors to be negatively biased, which may be a serious issue when it comes to such important economic indicators as unemployment. The literature on STS 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 predictors. Within multilevel models, that are a standard tool in SAE, it is common practice to account for the hyperparameter uncertainty. Whether the MSE bias is substantial depends on the model structure in combination with the time series length and can be checked through a simulation for a particular application.

The literature offers several procedures to correct for the negative bias in the MSE estimates produced by STS models. The present work is aimed at establishing the best estimation 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. The simulation study reveals that the asymptotic approximation is not applicable to cases where hyperparameter estimates are 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 less dependent on normality assumptions about the error distribution, perform better than their parametric counterparts under both Pfeffermann and Tiller (2005) and Rodriguez and Ruiz (2012) methods, except when the time series is too short ($T = 48$ in this paper). 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 (in absolute terms) than those produced by the

Kalman filter. This is contrary to the claim of Rodriguez and Ruiz (2012) about the superiority of their method for short time series. Apparently, their findings are purely heuristic and are based on a simple model (random walk plus noise), while Pfeffermann and Tiller (2005) prove that their bootstrap approach produces MSE estimates with a bias of correct order. The variance and MSE of the bootstrap MSE estimators has also been studied in this paper. The results show that the PT MSE estimators feature larger variances; their MSEs, however, only slightly exceed the MSEs of the RR MSE estimators (MSEs of the RR-based MSE are 28 to 8 percent lower than the PT-based MSEs, depending on the model and time series length). More importantly, the tendency of the RR MSE estimators to have negative biases, sometimes exceeding those of the Kalman filter, renders these bootstrap methods inapplicable. Hence, the PT-methods should be generally considered for other survey data too, despite the fact that these methods may occasionally be outperformed by the RR-methods.
For very short time series (of 48 months) that may occur in practice when an NSI switches to the STS-based approach, the relative bias of the signal MSE produced by the Kalman filter is around 7 percent. Non-parametric bootstraps do not seem to be an option for this short length for a model of the presented complexity, with non-stationary state variables present, due to a quite extensive burn-in period necessary for generation of bootstrap series. The PT parametric bootstrap, however, corrects the negatively biased MSE up to a small positive bias (1.4 to 4.4 percent, depending on the model). For the present series length of 114 months, the negative MSE bias can be reduced from about -2.4 to 1.9 percent with the non-parametric method of Pfeffermann and Tiller (2005) in the model with the RGB hyperparameter set to zero. Even with this slight positive bias, the standard errors of the GREG estimates are reduced by about 22 percent. In general, the biases in the Kalman filter MSE estimates are relatively small in the DLFS application, therefore it may be deemed sufficient to rely on these naive MSE estimates for publication purposes.

Acknowledgements

We thank the national statistical office of the Netherlands, Statistics Netherlands, for funding this research, as well as the Associate Editor and the two anonymous reviewers for careful reading of this manuscript and valuable comments. The views expressed in this paper are those of the authors and do not necessarily reflect the policy of Statistics Netherlands.

Appendices

A State Space Form of the DLFS Model

The state space representation of the DLFS model consists of the transition equation (2.4) and the measurement equation (2.5), with the state vector defined in (2.6). The design matrix $\boldsymbol{Z_t}$ in the measurement equation is a composite of 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 matrix of a dimension specified in the subscript, with each element equal to zero,

$\boldsymbol{Z_t^\xi} = [\mathbf{1_5} \otimes (1\,0\,1\,0\,1\,0\,1\,0\,1\,0\,1\,0\,1\,1)]$ selects the level L_t, six seasonal harmonics $\gamma_{t,1}, \ldots \gamma_{t,6}$ and the population parameter white noise ε_t 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-based 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 design matrix $\boldsymbol{T}$ of the transition equation is defined as:

$$\boldsymbol{T} = Blockdiag(\boldsymbol{T^L}\ \boldsymbol{T^\gamma}\ 0\ \boldsymbol{T^\lambda}\ \boldsymbol{T^e}),$$

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

The covariance matrix of the state stochastic terms is diagonal:

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

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

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

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 Variance Function for Simulation

This appendix contains details on how the variance function in (4.1) is derived from the well-known expression for the variance of the estimated population total of a binary response variable: $\widehat{Var}(Y_t) = N_t^2\,(1 - n_t/N_t)\,p_t(1-p_t)/(n_t - 1)$, 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 Y_t/N_t. First take logs on both sides of this expression:

$$ln(\widehat{Var}(Y_t)) = 2lnN_t + ln\,(N_t - n_t/N_t) - ln(n_t - 1) + lnp_t + ln(1 - p_t). \tag{B.1}$$

Taking $ln(n_t - 1) \approx lnn_t$ and assuming that unemployment p_t does not reach high values, which allows us to neglect the term $ln(1 - p_t)$, results in the following :

$$\begin{aligned} ln(\widehat{Var}(Y_t)) &= 2lnN_t - lnN_t + ln(N_t - n_t) - lnn_t + lnp_t; \\ &= lnN_t - lnn_t + ln(N_t - n_t) + ln(Y_t/N_t). \end{aligned} \tag{B.2}$$

Take $ln(N_t - n_t) \approx lnN_t$, then:

$$\begin{aligned} ln(\widehat{Var}(Y_t)) &= ln(N_t/n_t) + lnN_t + lnY_t - lnN_t; \\ &= -ln(n_t/N_t) + lnY_t. \end{aligned} \tag{B.3}$$

The dependence of the design-based variance estimator on the numbers of unemployed, Y_t, and on n_t/N_t can be reflected in the following variance function:

$$ln(\widehat{Var}(Y_t)) \approx \alpha ln\,(n_t/N_t) + \beta ln(Y_t). \tag{B.4}$$

In this simulation, signals $l_t^{t-j} = L_t + \gamma_t + \lambda_t^{t-j}$ act as a proxy for Y_t^{t-j}. The term $ln\,(n_t/N_t)$ is assumed to be constant over time and is represented by intercept c in the expression for the first wave (see the first equation of (4.1)). As can be seen from the second equation in (4.1), the design-based variances in waves 2 through 5 depend on the design-based variances of the preceding waves (i.e. of waves 1 though 4) observed three months ago, 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.

C Simulated Density Functions of the Hyperparameters under the Four Versions of the DLFS Model

This appendix presents the hyperparameter density functions obtained from simulations where the four versions of the DLFS model (see Table 5.1) act as the data generating process. The x-axes depict variance hyperparameters on a log-scale, while the y-axes stand for frequencies. The x-axis may be extended due to outliers.

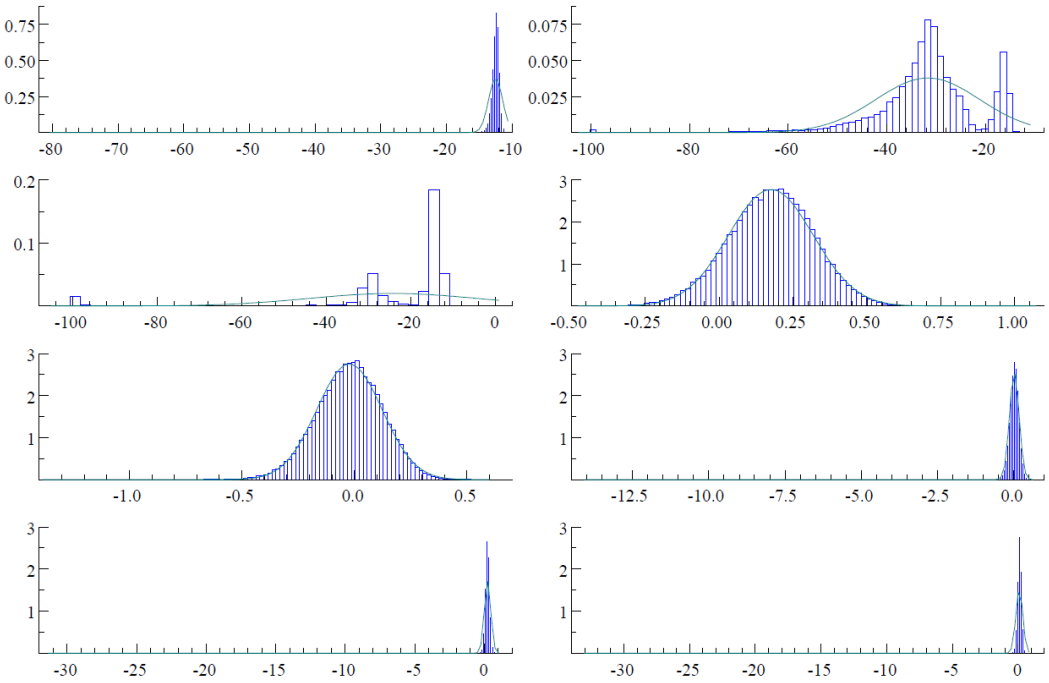

Figure C.1 Hyperparameter distribution under the complete DLFS model (Model 1), left to right on the x-axes: $ln(\hat{\sigma}_R^2), ln(\hat{\sigma}_\gamma^2), ln(\hat{\sigma}_\lambda^2), ln(\hat{\sigma}_{v_t^t}^2), ln(\hat{\sigma}_{v_t^{t-3}}^2), ln(\hat{\sigma}_{v_t^{t-6}}^2), ln(\hat{\sigma}_{v_t^{t-9}}^2), ln(\hat{\sigma}_{v_t^{t-12}}^2)$**; the normal density with the same mean and variance superimposed; 50000 simulations, T=114.**

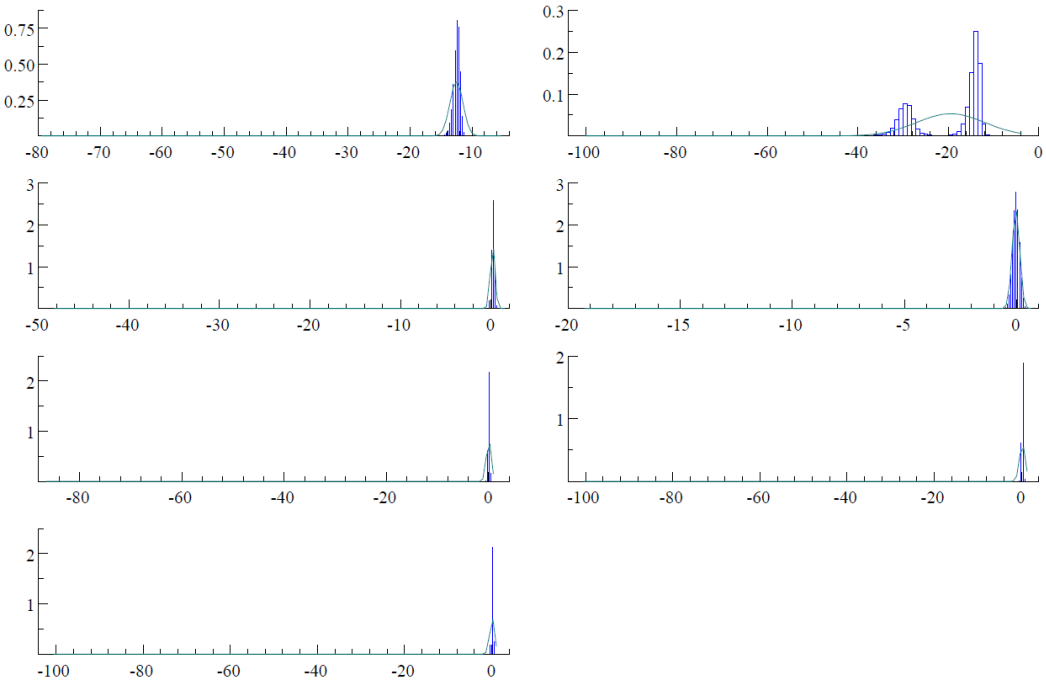

Figure C.2 Hyperparameter distribution under Model 2, left to right on the x-axes: $ln(\hat{\sigma}_R^2), ln(\hat{\sigma}_\lambda^2), ln(\hat{\sigma}_{\nu_t^t}^2), ln(\hat{\sigma}_{\nu_t^{t-3}}^2), ln(\hat{\sigma}_{\nu_t^{t-6}}^2), ln(\hat{\sigma}_{\nu_t^{t-9}}^2), ln(\hat{\sigma}_{\nu_t^{t-12}}^2)$; the normal density with the same mean and variance superimposed; 50000 simulations, T=114.

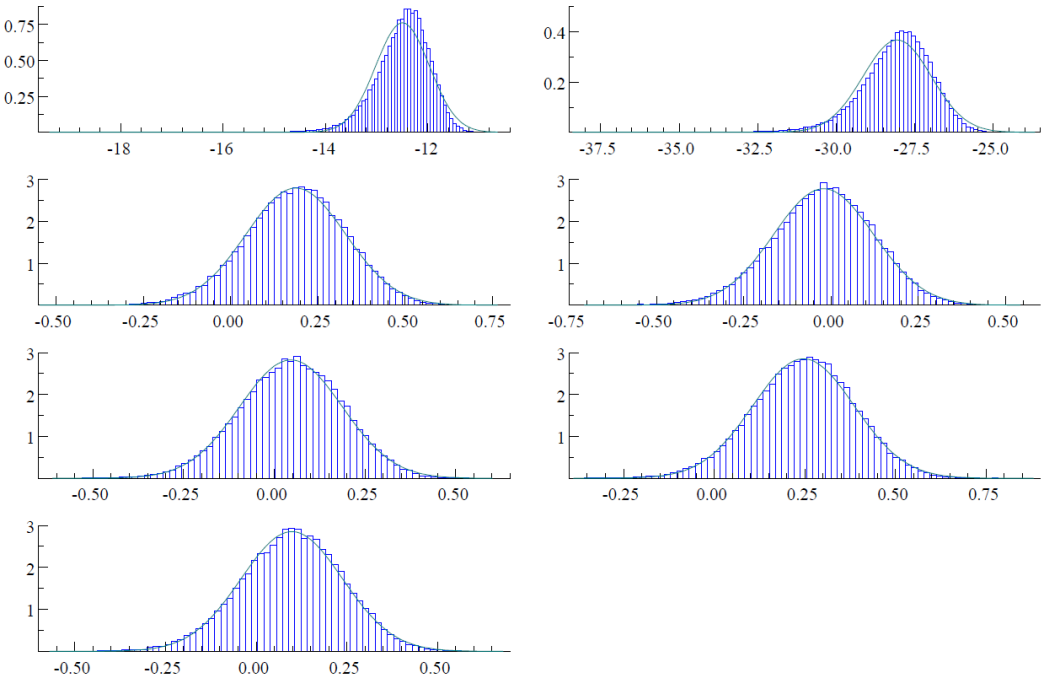

Figure C.3 Hyperparameter distribution under Model 3, left to right on the x-axes: $ln(\hat{\sigma}_R^2), ln(\hat{\sigma}_\gamma^2), ln(\hat{\sigma}_{\nu_t^t}^2), ln(\hat{\sigma}_{\nu_t^{t-3}}^2), ln(\hat{\sigma}_{\nu_t^{t-6}}^2), ln(\hat{\sigma}_{\nu_t^{t-9}}^2), ln(\hat{\sigma}_{\nu_t^{t-12}}^2)$; the normal density with the same mean and variance superimposed; 50000 simulations, T=114.

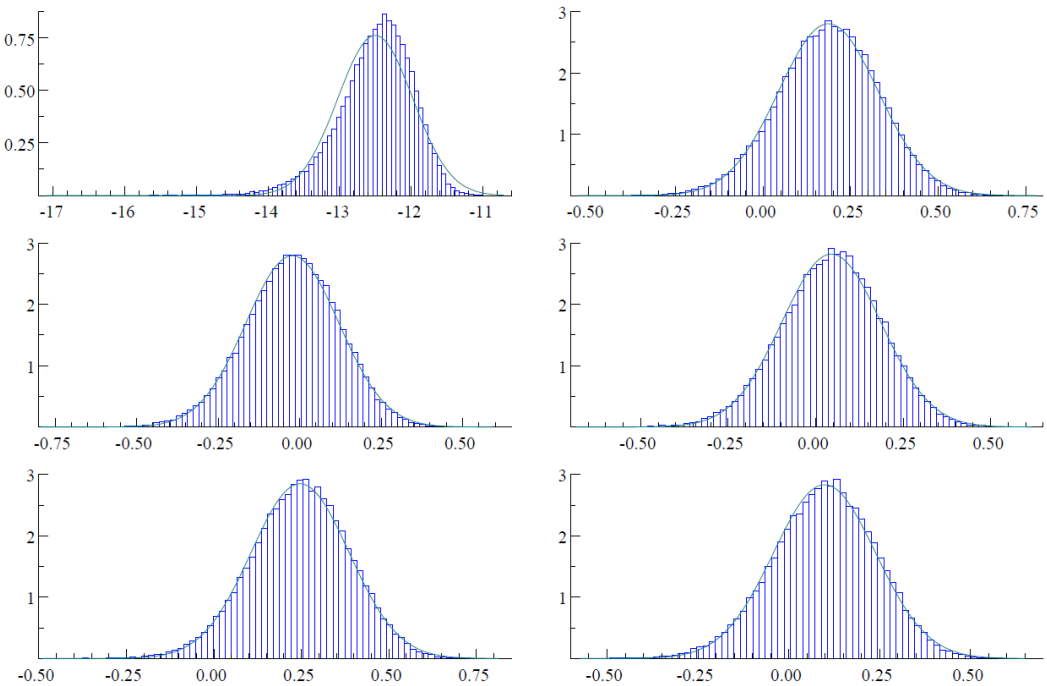

Figure C.4 Hyperparameter distribution under Model 4, left to right on the x-axes: $ln(\hat{\sigma}_R^2)$, $ln(\hat{\sigma}_{\nu_t^t}^2)$, $ln(\hat{\sigma}_{\nu_t^{t-3}}^2)$, $ln(\hat{\sigma}_{\nu_t^{t-6}}^2)$, $ln(\hat{\sigma}_{\nu_t^{t-9}}^2)$, $ln(\hat{\sigma}_{\nu_t^{t-12}}^2)$; the normal density with the same mean and variance superimposed; 50000 simulations, T=114.

References

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

Bartlett, M. S. (1946). On the theoretical specification and sampling properties of autocorrelated time-series. *Supplement to the Journal of the Royal Statistical Society* 8, 27--41.

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

Bollineni-Balabay, O., Van den Brakel, J., and Palm, F. (2016). Multivariate state space approach to variance reduction in series with level and variance breaks due to survey redesigns. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 179, 377--402.

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.

Durbin, J. and Koopman, S. J. (2012). *Time Series Analysis by State Space Methods*. Number 38. Oxford University Press.

EUROSTAT (2015). Task Force on monthly unemployment - revised report. Working Group Labour Market Statistics.

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.

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

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., 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, 2918--2933.

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

Lemaître, G. and Dufour, J. (1987). An integrated method for weighting persons and families. *Survey Methodology* 13, 199--207.

ONS (2015). A State Space Model for LFS Estimates: agreeing the target and dealing with wave specific bias. Report of the 29-th meeting of the Government Statistical Service Methodology Advisory Committee. http://www.ons.gov.uk/ons/guide-method/method-quality/advisory-committee/previous-meeting-papers-and-minutes/mac-29-papers.pdf.

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. (2013). New important developments in small area estimation. *Statistical Science* 28, 40--68.

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. and Molina, I. (2015). *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. (2009). 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. (2015). Dealing with small sample sizes, rotation group bias and discontinuities in a rotating panel design. *Survey Methodology* 41, 267--296.

Zhang, M. and Honchar, O. (2016). Predicting survey estimates by state space models using multiple data sources. *Paper for the Australian Bureau of Statistics' Methodology Advisory Committee*.

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

Prepress: Statistics Netherlands, Grafimedia
Design: Edenspiekermann

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

Where to order
verkoop@cbs.nl
Fax +31 45 570 62 68
ISSN 1572-0314

xxxxxxxxxxxxx