Catalogue no. 12-001-X
ISSN 1492-0921

Survey Methodology

Dealing with small sample sizes, rotation group bias and discontinuities in a rotating panel design

by Jan A. van den Brakel and Sabine Krieg

Release date: December 17, 2015

Statistics Canada Statistique Canada

Canadä

How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, www.statcan.gc.ca.

You can also contact us by

email at STATCAN.infostats-infostats.STATCAN@canada.ca

telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following toll-free numbers:

- Statistical Information Service 1-800-263-1136
- National telecommunications device for the hearing impaired 1-800-363-7629
- Fax line 1-877-287-4369

Depository Services Program

- Inquiries line 1-800-635-7943
- Fax line 1-800-565-7757

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under "Contact us" > "Standards of service to the public."

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Standard table symbols

The following symbols are used in Statistics Canada publications:

.	not available for any reference period
..	not available for a specific reference period
...	not applicable
0	true zero or a value rounded to zero
0^s	value rounded to 0 (zero) where there is a meaningful distinction between true zero and the value that was rounded
^{p}	preliminary
^{r}	revised
x	suppressed to meet the confidentiality requirements of the *Statistics Act*
^{E}	use with caution
F	too unreliable to be published
*	significantly different from reference category ($p < 0.05$)

Published by authority of the Minister responsible for Statistics Canada

An HTML version is also available.

Cette publication est aussi disponible en français.

Survey Methodology, December 2015
Vol. 41, No. 2, pp. 267-296
Statistics Canada, Catalogue No. 12-001-X

Dealing with small sample sizes, rotation group bias and discontinuities in a rotating panel design

Jan A. van den Brakel and Sabine Krieg[1]

Abstract

Rotating panels are widely applied by national statistical institutes, for example, to produce official statistics about the labour force. Estimation procedures are generally based on traditional design-based procedures known from classical sampling theory. A major drawback of this class of estimators is that small sample sizes result in large standard errors and that they are not robust for measurement bias. Two examples showing the effects of measurement bias are rotation group bias in rotating panels, and systematic differences in the outcome of a survey due to a major redesign of the underlying process. In this paper we apply a multivariate structural time series model to the Dutch Labour Force Survey to produce model-based figures about the monthly labour force. The model reduces the standard errors of the estimates by taking advantage of sample information collected in previous periods, accounts for rotation group bias and autocorrelation induced by the rotating panel, and models discontinuities due to a survey redesign. Additionally, we discuss the use of correlated auxiliary series in the model to further improve the accuracy of the model estimates. The method is applied by Statistics Netherlands to produce accurate official monthly statistics about the labour force that are consistent over time, despite a redesign of the survey process.

Key Words: Common factor models; Kalman filter; Measurement bias; Small area estimation; Structural time series modelling; Survey sampling.

1 Introduction

Sample surveys of national statistical institutes are generally conducted repeatedly with the purpose of constructing time series that describe the evolution of finite population parameters of interest. Estimation techniques employed by national statistical institutes are largely design based. This implies that statistical inference is predominantly based on the stochastic structure of the sampling design, while statistical models only play a minor role. The general regression (GREG) estimator (Särndal, Swensson and Wretman 1992) is an example of this class of estimators. This estimator expands or weights the observations obtained in the sample with the so-called survey weights, such that the sum over the weighted observations is approximately design unbiased for the unknown population total. The survey weights are initially derived from the sampling design, by taking the weights equal to the inverse of the inclusion probabilities of the sampling units. In a second step these design-weights are calibrated, such that the sum over the weighted auxiliary variables in the sample equates to the known population totals. Under the model-assisted approach, the GREG estimator is derived from a linear regression model that specifies the relationship between the values of a certain target parameter and a set of auxiliary variables.

This class of estimators has nice properties that make them very attractive for use in a production process of compiling timely official statistics. GREG estimators are asymptotically design unbiased and consistent, see Isaki and Fuller (1982), and Robinson and Särndal (1983). This provides a form of robustness in the case of large sample sizes. If the underlying linear model of the GREG estimator explains the variation of the target parameter in the finite population reasonably well, then the use of

1. Jan A. van den Brakel, Methodology Department, Statistics Netherlands, PO Box 4481, 6401 CZ Heerlen, The Netherlands and Department of Quantitative Economics, Maastricht University School of Business and Economics, PO Box 616, 6200 MD Maastricht, The Netherlands. E-mail: ja.vandenbrakel@cbs.nl; Sabine Krieg, Methodology Department, Statistics Netherlands, PO Box 4481, 6401 CZ Heerlen, The Netherlands.

auxiliary information results in a reduction of the design variance and also decreases the bias due to selective non-response. Model misspecification might result in an increase of the design variance but the property that the GREG estimator is approximately design unbiased remains. From this point of view, the GREG estimator is robust against model misspecification. Additionally, these estimators only require one set of weights to estimate all possible target variables, which is an attractive practical advantage in multipurpose surveys.

Major drawbacks of GREG estimators are the relatively large design variances in the case of small sample sizes, and the fact that they do not handle measurement errors effectively. In such situations, model-based procedures can be used to produce more reliable estimates. These estimators employ sample information observed in other domains or previous time periods through an explicit statistical model, thus increasing the effective sample size in the separate domains and specific periods. In survey methodology, this type of estimation techniques is known as small area estimation, see Rao (2003) for a comprehensive overview. In this paper we describe an estimation approach, based on structural time series modelling, to deal with small sample sizes and problems with non-sampling errors in the Dutch Labour Force Survey (LFS).

Official monthly statistics about the Dutch labour force are based on the Dutch LFS. This survey is based on a rotating panel design. The responding households are interviewed five times at quarterly intervals, which implies that every month five panels are being interviewed. The estimation procedure of the LFS is based on the GREG estimator.

This paper solves three major problems encountered with this survey. The first problem is that the monthly sample size of the LFS is too small to rely on the GREG estimator to produce timely official monthly statistics about the employed and unemployed labour force. Therefore many national statistical institutes publish rolling quarterly figures about the labour force each month. Rolling quarterly figures have the obvious disadvantages that monthly seasonal patterns are smoothed out and that they are less timely since the monthly publications refer to the latest rolling quarter instead of the latest month.

The second problem is that there are substantial systematic differences between the subsequent panels due to mode and panel effects. This is a well-known problem for rotating panel designs, and in the literature this is referred to as rotation group bias (RGB), Bailar (1975). At the moment that the LFS changed from a cross-sectional survey to a rotating panel design in October 1999, the effects of the RGB on the outcomes of the LFS became very visible. This was the direct cause for developing procedures that account for this RGB.

The third problem is the systematic effect on the outcomes of the LFS due to a major redesign of the survey process in 2010. Redesigns generally affect the various non-sampling error sources in a survey process, and therefore result in systematic effects on the outcomes of a survey. In an ideal survey transition process, these so-called discontinuities are quantified in order to keep series consistent and preserve comparability of the outcomes over time. In this redesign the first panel under the old and the new design is conducted in parallel for a period of six months, which provides a direct estimate for the discontinuities in the first panel.

Pfeffermann (1991) proposed a multivariate structural time series model for rotating panels to borrow strength over time and to account for RGB in the level of monthly labour force series. Van den Brakel and Krieg (2009) applied this model to the LFS to estimate the monthly unemployment rate. They extended the model to account for RGB in the level and the seasonal patterns of the monthly unemployment rate series. Van den Brakel and Roels (2010) proposed an intervention analysis approach to estimate discontinuities due to a redesign of cross-sectional surveys, as an alternative for a parallel run.

In this paper, the model proposed by Pfeffermann (1991) is extended with this intervention approach and available auxiliary series. We describe how this model increases the precision of direct estimates by taking advantage of sample information from previous periods, and accounts for the autocorrelation in the sampling errors of the different panels, the RGB, and the discontinuities that arise by the change-over to a new survey process. We focus on how this model enables Statistics Netherlands to publish sufficiently reliable official monthly statistics about the labour force instead of rolling quarterly figures, commonly published by national statistical institutes. We illustrate how the model facilitates a smooth change-over to a new survey design by modelling discontinuities with intervention variables. An important question that will be addressed is how the information from the parallel run in the first panel can be used in the time series model. Finally we illustrate how available auxiliary information about the number of people that are formally registered at the employment office can be incorporated in the time series model to improve the estimates of the discontinuities as well as the precision of the model estimates.

The paper starts in Section 2 with a brief description of the LFS and the problems encountered with the chosen survey design. Section 3 describes the proposed time series model to estimate monthly labour force figures. Section 4 describes the implementation of the time series model before the redesign and compares the results of the time series model with the rolling quarterly figures. The introduction of the new survey design is accompanied by a parallel run of six months, which is described in Section 5. Six different methods are proposed to handle the problems with discontinuities induced by the redesign in Section 6. Results obtained with these methods are compared in Section 7, including a motivation for the method that is finally chosen to produce official statistics. The paper concludes with a discussion in Section 8.

2 Design of the Dutch Labour Force Survey

The objective of the Dutch LFS is to provide reliable information about the Dutch labour force. Each month a stratified two-stage cluster design of addresses is drawn. Strata are formed by geographical regions. Municipalities are considered as primary and addresses as secondary sampling units. All households residing at an address, up to a maximum of three, are included in the sample. Different subpopulations are oversampled to improve the accuracy of the official releases, for example, addresses where people live who are formally registered at the employment office, and subpopulations with low response rates.

Before 2000, the LFS was designed as a cross-sectional survey. Since October 1999, the LFS has been conducted as a rotating panel design. Until the redesign in 2010, data in the first panel were collected by means of computer assisted personal interviewing (CAPI). Respondents were re-interviewed four times at quarterly intervals by means of computer assisted telephone interviewing (CATI). During these re-interviews, a condensed questionnaire was used to establish changes in the labour market position of the respondents. The monthly gross sample size for the first panel averaged about 8,000 addresses commencing the moment that the LFS changed to a rotating panel design and gradually fell to about 6,500 addresses in 2012. The response rate is about 55% in the first panel and in the subsequent panels about 90% with respect to the responding households from the preceding panel.

The estimation procedure of the LFS starts with the GREG estimator. Inclusion probabilities reflect the sampling design and differences in response rates between geographic regions. The weighting scheme is based on a combination of different socio-demographic categorical variables. Key parameters of the LFS are the employed, unemployed and total labour force, which are defined as population totals. Another important parameter is the unemployment rate, which is defined as the ratio of the unemployed labour force to the total labour force.

Figure 2.1 illustrates the RGB for the unemployed labour force. The series of the GREG estimates of the first panel are compared with the average of the GREG estimates of the four subsequent panels. The GREG estimates for the unemployed labour force in the subsequent panels are systematically smaller than in the first panel. The RGB is a consequence of different non-sampling errors like selective non-response, panel attrition, mode-effects, effects due to differences between the CAPI questionnaire and the CATI questionnaire, and panel effects.

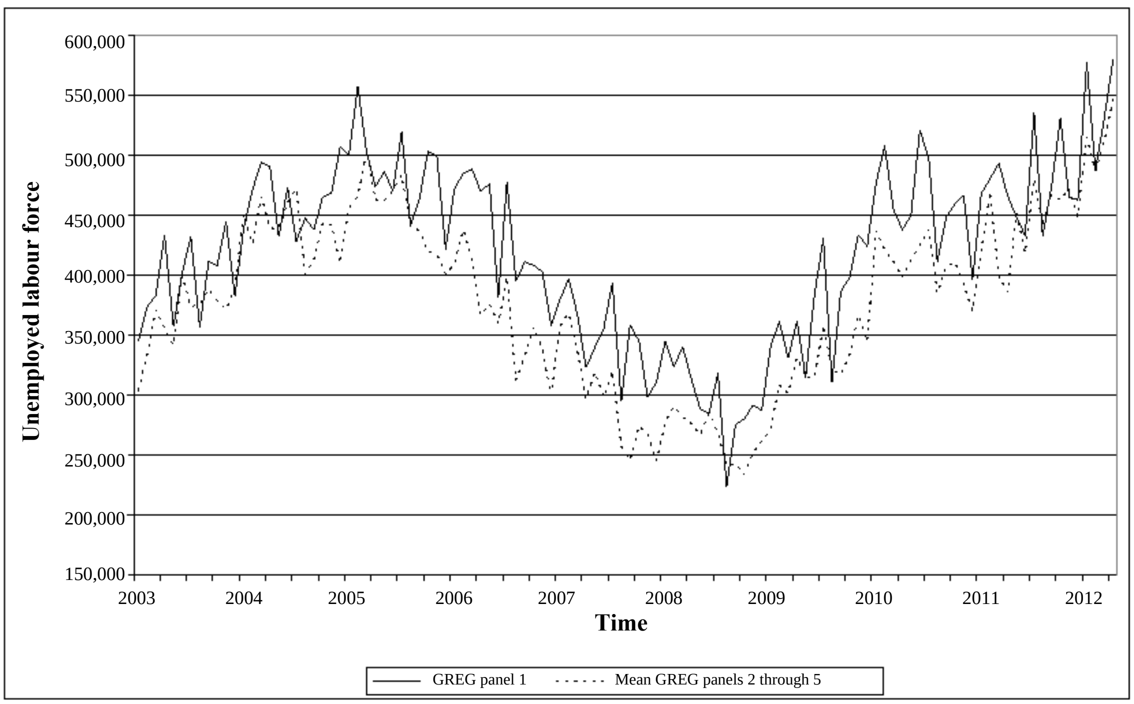


Figure 2.1 RGB unemployed labour force at the national level; comparison GREG estimates based on panel 1 with the mean of the series of the GREG estimates based on panel 2 through 5.

Until June 2010, rolling quarterly figures about the labour force were published each month. A rigid correction was applied to correct for the RGB. For the most important parameters, the ratio between the estimates based on the first panel only and the estimates based on all panels was computed using the data of the 12 preceding quarters. Estimates for the rolling quarterly figures were multiplied by this ratio to correct for RGB. In June 2010, a structural time series model was implemented to estimate model-based monthly figures instead of design-based rolling quarterly figures about the labour force. This model accounts for the RGB, and therefore replaces the ratio correction.

In 2010, a major redesign for the LFS started. The main objective of this redesign was to reduce the administration costs of this survey. This is accomplished by changing the data collection in the first panel from CAPI to a mixed data collection mode using CAPI and CATI. Households with a listed telephone number are interviewed by telephone, the remaining households are interviewed face-to-face. To make CATI data collection in the first panel feasible, the questionnaire for the first panel needed to be abridged since a telephone interview should not take longer than 15 to 20 minutes. Therefore parts of the questionnaire were transferred from the first to the second or the third panel. To avoid confounding real developments with systematic effects induced by the redesign, it is important to quantify these discontinuities and to account for these effects in the time series model.

3 Estimating monthly labour force figures

In this section a multivariate structural time series model is developed for the LFS data that are observed under the rotating panel design. The model deals with small sample sizes by borrowing strength over time to improve the precision of the GREG estimates, and accounts for the RGB as well as the autocorrelation between the subsequent panels of the rotating panel and models the discontinuities due to the redesign of the LFS in 2010.

Let $\hat{Y}_t^j$ denote the GREG estimate for the unknown population parameter, say θ_t, based on the j^{th} panel observed at time t, $j = 1, \ldots, 5.$ Since responding households are interviewed at quarterly intervals, it follows that the j^{th} panel at time t that was sampled for the first time at time $t - 3j + 3.$ Due to the applied rotation pattern, each month data are collected in five different panels and a vector $\hat{\mathbf{Y}}_t = \left(\hat{Y}_t^1, \hat{Y}_t^2, \hat{Y}_t^3, \hat{Y}_t^4, \hat{Y}_t^5\right)^T$ is observed. A five dimensional time series with GREG estimates for the monthly employed and unemployed labour force is obtained as a result. Pfeffermann (1991) proposed a multivariate structural time series model for this kind of time series to model the population parameter of interest, and to account for the RGB and the autocorrelation in the sampling errors. This approach is extended with an intervention component to model the discontinuities of the survey redesign. This results in the following time series model for the five series of GREG estimates:

$$\hat{\mathbf{Y}}_t = \mathbf{1}_5 \theta_t + \boldsymbol{\lambda}_t + \boldsymbol{\Delta}_t \boldsymbol{\beta} + \mathbf{e}_t, \tag{3.1}$$

with $\mathbf{1}_5$ a five dimensional vector with each element equal to one, $\boldsymbol{\lambda}_t = \left(\lambda_t^1, \lambda_t^2, \lambda_t^3, \lambda_t^4, \lambda_t^5\right)^T$ a vector with time dependent components that account for the RGB, $\boldsymbol{\Delta}_t = \text{Diag}\left(\delta_t^1, \delta_t^2, \delta_t^3, \delta_t^4, \delta_t^5\right)$ a diagonal matrix with dummy variables that change from zero to one at the moment that the survey changes from the old to the new design, $\boldsymbol{\beta} = \left(\beta^1, \beta^2, \beta^3, \beta^4, \beta^5\right)^T$ a five dimensional vector with regression coefficients, and $\mathbf{e}_t = \left(e_t^1, e_t^2, e_t^3, e_t^4, e_t^5\right)^T$ the corresponding survey errors for each panel estimate.

The population parameter θ_t in (3.1) can be decomposed in a trend component, a seasonal component, and an irregular component, i.e.,

$$\theta_t = L_t + S_t + \varepsilon_t. \tag{3.2}$$

Here L_t denotes a stochastic trend component, using the so-called smooth trend model,

$$\begin{aligned} L_t &= L_{t-1} + R_{t-1}, \\ R_t &= R_{t-1} + \eta_t, \\ E(\eta_t) &= 0, \quad \mathrm{Cov}(\eta_t, \eta_{t'}) = \begin{cases} \sigma_\eta^2 & \text{if } t = t' \\ 0 & \text{if } t \neq t'. \end{cases} \end{aligned} \tag{3.3}$$

A likelihood ratio test indicates that in this application the more general local linear trend model, which has a disturbance term for the slope parameter R_t as well as a disturbance term for the level parameter L_t, does not improve the fit to the data. Inclusion of a disturbance term for the level increases the log-likelihood of (3.1) with 0.05 units. This results in a likelihood ratio test statistic of 0.1. Under the null hypothesis that the level disturbance term is equal to zero, this test statistic is a chi-squared distributed random variable with 1 degree of freedom. As a result, this null hypothesis is accepted with a $p-$value of 0.75.

Furthermore, S_t denotes a trigonometric stochastic seasonal component,

$$S_t = \sum_{l=1}^{6} S_{l,t}, \tag{3.4}$$

where

$$\begin{aligned} S_{l,t} &= S_{l,t-1} \cos(h_l) + S_{l,t-1}^* \sin(h_l) + \omega_{l,t} \\ S_{l,t}^* &= S_{l,t-1}^* \cos(h_l) - S_{l,t-1} \sin(h_l) + \omega_{l,t}^*, \quad h_l = \frac{\pi l}{6}, \quad l = 1, \ldots, 6, \\ E(\omega_{l,t}) &= E(\omega_{l,t}^*) = 0, \\ \mathrm{Cov}(\omega_{l,t}, \omega_{l',t'}) &= \mathrm{Cov}(\omega_{l,t}^*, \omega_{l',t'}^*) = \begin{cases} \sigma_\omega^2 & \text{if } l = l' \text{ and } t = t' \\ 0 & \text{if } l \neq l' \text{ or } t \neq t' \end{cases}, \\ \mathrm{Cov}(\omega_{l,t}, \omega_{l,t}^*) &= 0, \quad \forall l, \forall t. \end{aligned} \tag{3.5}$$

Finally, ε_t denotes the irregular component, which contains the unexplained variation of the population parameter and is modelled as a white noise process:

$$E(\varepsilon_t) = 0, \quad \mathrm{Cov}(\varepsilon_t, \varepsilon_{t'}) = \begin{cases} \sigma_\varepsilon^2 & \text{if } t = t' \\ 0 & \text{if } t \neq t'. \end{cases} \tag{3.6}$$

It is not immediately obvious that the white noise component ε_t in (3.2) and the sampling errors $\mathbf{e}_t$ in (3.1) are both identifiable. The sampling errors can be separated from the white noise component because each sample is observed five times and because the variance of the sampling errors, as well as the autocorrelation in the sampling errors induced by the sample overlap of the panel, are calculated directly from the survey data. Details are explained below.

The trend (3.3) describes the gradual change in the population parameter, while the seasonal component (3.4) captures the systematic monthly deviations from the trend within a year. See e.g., Durbin and Koopman (2001) for details. Through component (3.2) values for θ_t are related to the population values from preceding periods. This component shows how sample information observed in preceding periods is used to improve the precision of the estimates for θ_t in a particular time period.

The systematic differences between the subsequent panels, i.e., the RGB, are modelled in (3.1) with $\boldsymbol{\lambda}_t$. The absolute bias in the monthly labour force figures cannot be estimated from the sample data only. Therefore additional restrictions for the elements of $\boldsymbol{\lambda}_t$ are required to identify the model. Here it is assumed that an unbiased estimate for θ_t is obtained with the first panel, i.e., $\hat{Y}_t^1$. This implies that the first component of $\boldsymbol{\lambda}_t$ equals zero. The other elements of $\boldsymbol{\lambda}_t$ measure the time dependent differences with respect to the first panel. Contrary to Pfeffermann (1991), were time independent RGB is assumed, λ_t^j are modelled as random walks for $j = 2, 3, 4,$ and $5.$ As a result it follows that

$$\lambda_t^1 = 0, \quad \lambda_t^j = \lambda_{t-1}^j + \eta_{\lambda,j,t}, j = 2, 3, 4, 5, \tag{3.7}$$

$$E\left(\eta_{\lambda,j,t}\right) = 0, \quad \operatorname{Cov}\left(\eta_{\lambda,j,t}, \eta_{\lambda,j',t'}\right) = \begin{cases} \sigma_\lambda^2 & \text{if} \quad t = t' \quad \text{and} \quad j = j' \\ 0 & \text{if} \quad t \neq t' \quad \text{or} \quad j \neq j'. \end{cases}$$

The discontinuities induced by the redesign in 2010 are modelled with the third term in (3.1). The diagonal matrix $\boldsymbol{\Delta}_t$ contains five intervention variables:

$$\delta_t^j = \begin{cases} 0 & \text{if} \quad t < T_R^j \\ 1 & \text{if} \quad t \geq T_R^j \end{cases}, \text{for } j = 1, 2, \ldots, 5, \tag{3.8}$$

where T_R^j denotes the moment that panel j changes from the old to the new survey design. Under the assumption that (3.2) correctly models the evolution of the population variable, the regression coefficients in $\boldsymbol{\beta}$ can be interpreted as the systematic effects of the redesign on the level of the series observed in the five panels. The intervention approach with state-space models was originally proposed by Harvey and Durbin (1986) to estimate the effect of seat belt legislation on British road casualties. With step intervention (3.8) it is assumed that the redesign only has a systematic effect on the level of the series. Alternative interventions, e.g., for the slope or the seasonal components are also possible, see Durbin and Koopman (2001), Chapter 3. A redesign might not only affect the point estimates, but also the variance of the GREG estimates. This issue is discussed under the time series model for the survey errors.

Finally a time series model for the survey errors $\mathbf{e}_t$ in (3.1) is developed. The direct estimates for the design variances of the survey errors are available from the micro data and are incorporated in the time series model using the survey error model $e_t^j = k_t^j \tilde{e}_t^j$ where $k_t^j = \sqrt{\operatorname{V\hat{a}r}\left(\hat{Y}_t^j\right)},$ proposed by Binder and Dick (1990). Here $\operatorname{V\hat{a}r}\left(\hat{Y}_t^j\right)$ denotes the estimated variance of the GREG estimator. Choosing the survey errors proportional to the standard error of the GREG estimators allows for non-homogeneous variance in the survey errors, that arise e.g., due to the gradually decreasing sample size over the last decade.

The sample of the first panel has no sample overlap with panels observed in the past. Consequently, the survey errors of the first panel, e_t^1, are not correlated with survey errors in the past. It is, therefore, assumed that $\tilde{e}_t^1$ is white noise with $E\left(\tilde{e}_t^1\right) = 0$ and $\mathrm{Var}\left(\tilde{e}_t^1\right) = \sigma_{e1}^2$. As a result, the variance of the survey error equals $\mathrm{Var}\left(e_t^1\right) = \left(k_t^1\right)^2 \sigma_{e1}^2$, which is approximately equal to the direct estimate of the variance of the GREG estimate for the first panel if the maximum likelihood (ML) estimate for σ_{e1}^2 is close to one.

The survey errors of the second, third, fourth and fifth panel are correlated with survey errors of preceding periods. The autocorrelations between the survey errors of the subsequent panels are estimated from the survey data, using the approach proposed by Pfeffermann, Feder and Signorelli (1998). In this application it appears that the autocorrelation structure for the second, third, fourth and fifth panel can be modelled conveniently with an AR(1) model, van den Brakel and Krieg (2009). Therefore it is assumed that $\tilde{e}_t^j = \rho \tilde{e}_{t-3}^{j-1} + \nu_t^j$, with ρ the first order autocorrelation coefficient, $E\left(\nu_t^j\right) = 0$, and $\mathrm{Var}\left(\nu_t^j\right) = \sigma_{ej}^2$ for $j = 2, 3, 4, 5$. Since $\tilde{e}_t^j$ is an AR(1) process, $\mathrm{Var}\left(e_t^j\right) = \sigma_{ej}^2 \left(k_t^j\right)^2 \big/ \left(1 - \rho^2\right)$. As a result $\mathrm{Var}\left(e_t^j\right)$ is approximately equal to $\hat{\mathrm{Var}}\left(\hat{Y}_t^j\right)$ provided that the ML estimates for σ_{ej}^2 are close to $\left(1 - \rho^2\right)$.

The survey redesign in 2010 might affect the variance of the GREG estimates. Systematic differences in these variances are automatically taken into account, since they are used as a-priori information in the time series model for the survey error. An alternative possibility would be to allow for different values for σ_{ej}^2 before and after the survey redesign, which can be interpreted as an intervention on the variance hyperparameter of the survey error.

Auxiliary time series can be incorporated in the model to improve the estimates for the discontinuities. Reliable auxiliary series contain valuable information for correctly separating real developments from discontinuities in the intervention model. The auxiliary information will also increase the precision of the model estimates for the monthly unemployment figures. For the unemployed labour force, the number of people formally registered at the employment office is a potential auxiliary variable to be included in the model.

There are different ways to incorporate auxiliary information in the model. One straightforward possibility is to extend the time series model (3.2) for the population parameter of the LFS with a regression component for the auxiliary series, i.e., $\theta_t = L_t + S_t + bX_t + \varepsilon_t$, where X_t denotes the auxiliary series and b the regression coefficient. The major drawback of this approach is that the auxiliary series will partially explain the trend and seasonal effect in θ_t, leaving only a residual trend and seasonal effect for L_t and S_t. This hampers the estimation of a trend for the target variable.

An alternative approach, that allows the direct estimation of a filtered trend for θ_t, is to extend model (3.1) with the auxiliary series and model the correlation between the trends of the series of the LFS and the auxiliary series. This gives rise to the following model:

$$\begin{pmatrix} \mathbf{Y}_t \\ X_t \end{pmatrix} = \begin{pmatrix} \mathbf{1}_5 \theta_t^{\mathrm{LFS}} \\ \theta_t^R \end{pmatrix} + \begin{pmatrix} \boldsymbol{\lambda}_t \\ 0 \end{pmatrix} + \begin{pmatrix} \boldsymbol{\Delta}_t \boldsymbol{\beta} \\ 0 \end{pmatrix} + \begin{pmatrix} \mathbf{e}_t \\ 0 \end{pmatrix}. \tag{3.9}$$

The series of the LFS and the auxiliary series from the register both have their own population parameter that can be modelled with two separate time series models, i.e., $\theta_t^z = L_t^z + S_t^z + \varepsilon_t^z,$ where $z = \text{LFS}$ or $z = R$ (R stands for register), defined similarly to (3.2). Since the auxiliary series is based on a registration, this series does not have a RGB, a discontinuity at the moment that the LFS is redesigned or a survey error component.

The model allows for correlation between the disturbances of the slope of the trend component of the LFS and the auxiliary series. This results in the following definition for the smooth trend model for the LFS and the auxiliary series:

$$\begin{aligned}
L_t^z &= L_{t-1}^z + R_{t-1}^z, \\
R_t^z &= R_{t-1}^z + \eta_t^z, \\
E\left(\eta_t^z\right) &= 0, \\
\operatorname{Cov}\left(\eta_t^z, \eta_{t'}^z\right) &= \begin{cases} \sigma_{\eta^z}^2 & \text{if} \quad t = t' \\ 0 & \text{if} \quad t \neq t' \end{cases}, \quad z = \text{LFS}, R, \\
\operatorname{Cov}\left(\eta_t^{\text{LFS}}, \eta_{t'}^R\right) &= \begin{cases} \vartheta \sigma_{\eta \text{LFS}} \sigma_{\eta R} & \text{if} \quad t = t' \\ 0 & \text{if} \quad t \neq t', \end{cases}
\end{aligned}$$

with ϑ the correlation coefficient between these series. The correlation between both series is determined by the model. If the model detects a strong correlation, then the trends of both series will develop into the same direction more or less simultaneously. Model (3.9) does not allow for correlation between the disturbances of the seasonal component of the LFS series and the auxiliary series. Both series have their own seasonal component S_t^z defined by (3.5). In a similar way both series have their own white noise ε_t^z for the unexplained variation, which are assumed to be uncorrelated and are defined by (3.6).

Models (3.1) and (3.9) explicitly account for discontinuities in the different panels through the intervention component. Estimates for the target variables, obtained with these models, are therefore not affected by the systematic effect of the change-over. As a result, the models correct for the discontinuities induced by the redesign. Model estimates for the target variables can be interpreted as the results observed under the old method, also after the change-over to the new survey design. The discontinuity of the first panel must be added to the model estimates for the target variables to produce figures that can be interpreted as being obtained under the new design.

The general way to proceed is to express the model in the so-called state-space representation and apply the Kalman filter to obtain optimal estimates for the state variables, see e.g., Durbin and Koopman (2001). It is assumed that the disturbances are normally distributed. Under this assumption, the Kalman filter gives optimal estimates for the state vector and the signals. Estimates for state variables for period t based on the information available up to and including period t are referred to as the filtered estimates. The filtered estimates of past state vectors can be updated if new data become available. This procedure is referred to as smoothing and results in smoothed estimates that are based on the completely observed time series. In this application, interest is mainly focussed on the filtered estimates, since they are based on the complete set of information that would be available in the regular production process to produce a model-based estimate for month t.

The analysis is conducted with software developed in OxMetrics in combination with the subroutines of SsfPack 3.0, see Doornik (2009) and Koopman, Shephard and Doornik (2008). All state variables are non-stationary with the exception of the survey errors. The non-stationary variables are initialised with a diffuse prior, i.e., the expectation of the initial states is equal to zero and the initial covariance matrix of the states is diagonal with large diagonal elements. The survey errors are stationary and therefore initialised with a proper prior. The initial values for the survey errors are equal to zero and the covariance matrix is available from the aforementioned model for the survey errors. In Ssfpack 3.0 an exact diffuse log-likelihood function is obtained with the procedure proposed by Koopman (1997).

4 Implementation

In this section we compare the results obtained with the time series model with the GREG estimator for the period before the change-over to the new design, since rolling quarterly data are not calculated during and after the implementation of the new design. Since June 2010 model (3.1) has been applied to produce official monthly figures about the unemployed labour force, the employed labour force and the total labour force at the national level, and for six domains (men and women in three age classes). The model is applied to each variable separately. Estimates are computed as the sum of the trend and the seasonal effects, which is further referred to as the signal. Furthermore, trend estimates are published, replacing previous seasonally corrected figures. The first years of the GREG series are used to obtain stable estimates for the state variables of model (3.1). At the moment of implementation, a series of monthly figures starting in January 2003 is published.

Table 4.1 provides an overview of the ML estimates of the hyperparameters and the autocorrelation in the survey errors. The assumptions underlying the state-space model are evaluated by testing whether the standardized innovations are standard normally and independently distributed, Durbin and Koopman (2001), Section 4.2.4. Bowman-Shenton normality tests, $F-$tests for heteroscedasticity, $QQ-$plots, plots of standardized innovations and sample correlograms indicate that these assumptions are not violated under model (3.1).

Table 4.1
ML estimates of hyperparameters for monthly unemployed labour force figures before the survey redesign. Values are expressed as standard deviations

Standard deviation		National level	Men 15-24	Women 15-24	Men 25-44	Women 25-44	Men 45-64	Women 45-64
Slope	$(\hat{\sigma}_\eta)$	2,079	248	179	724	463	412	228
Seasonal	$(\hat{\sigma}_\omega)$	0.02	0.00	0.00	0.00	0.00	0.04	0.22
RGB	$(\hat{\sigma}_\lambda)$	905	941	468	268	669	3	335
White noise	$(\hat{\sigma}_\varepsilon)$	6,884	1,528	3,521	4,359	4,294	3,329	2
Survey error panel 1	$(\hat{\sigma}_{e1})$	1.07	0.98	1.11	1.04	0.89	0.99	1.14
Survey error panel 2	$(\hat{\sigma}_{e2})$	0.99	0.95	1.03	1.03	0.94	1.17	1.02
Survey error panel 3	$(\hat{\sigma}_{e3})$	1.01	1.06	1.12	1.03	0.96	1.04	0.92
Survey error panel 4	$(\hat{\sigma}_{e4})$	1.13	1.07	1.21	0.99	0.96	0.99	0.95
Survey error panel 5	$(\hat{\sigma}_{e5})$	1.06	1.00	1.03	0.99	0.99	1.08	0.87
Autocorrelation	$(\hat{\rho})$	0.21	0.13	0.12	0.39	0.22	0.44	0.38

The hyperparameter estimates for the survey errors for panel 2, 3, 4 and 5 are divided by $\left(1-\hat{\rho}^2\right)$. Therefore hyperparameters for the survey errors are, as expected, around 1.

In Figure 4.1, the filtered estimates for the monthly unemployed labour force at the national level based on model (3.1) are compared with the monthly GREG estimates and with the rolling quarterly GREG figures. Both GREG estimates are corrected for RGB using the ratio correction described in Section 2. The three series are at the same level, since they are calibrated to the level of the first panel. The series of the monthly GREG estimates has more pronounced peaks and dips than the filtered estimates. Under the times series model these fluctuations are partially considered as survey errors and filtered from the GREG estimates. The rolling quarterly figures have a less pronounced seasonal pattern, since monthly patterns are averaged over three subsequent months.

Figure 4.2 compares the filtered trend estimates with the seasonally adjusted estimates of the rolling quarterly data for the unemployed labour force at the national level. The seasonally adjusted rolling quarterly data, computed by X-12-ARIMA (U.S. Census Bureau 2009), were published before the new estimation method was implemented, and are available until May 2010. They are computed as the original estimates minus the seasonal effects. Besides the trend, they also include the sampling errors and other irregularities. Seasonally adjusted rolling quarterly figures and the filtered trend therefore measure slightly differently defined concepts. After the implementation of the time series model, the seasonally adjusted figures are replaced by the filtered trend, so it is interesting to compare the differences between both figures mainly to judge how large the consequences are for the users of these data.

There are some minor differences in the levels of the series in Figures 4.1 and 4.2. They are the result of large sampling errors and differences between the methods used to account for RGB. Firstly, the monthly GREG estimates and the rolling quarterly GREG estimates are more sensitive to large sampling errors. This in contrast with the time series model that filters the survey errors from the GREG estimates.

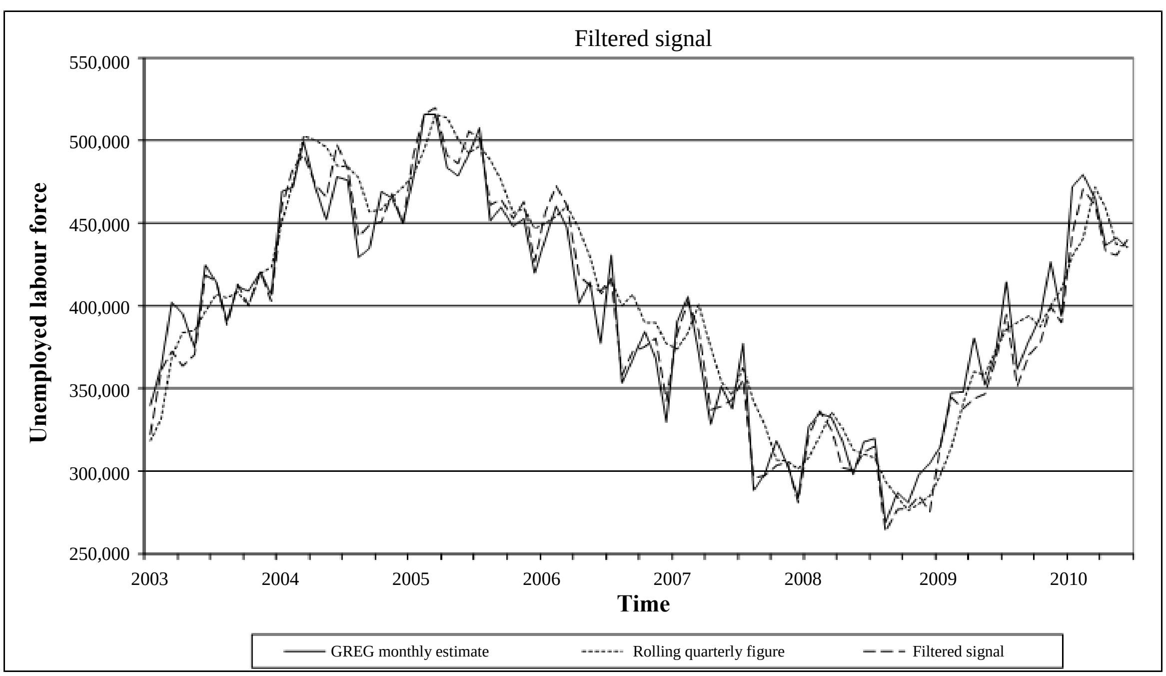


Figure 4.1 Monthly GREG estimates, rolling quarterly GREG estimates and monthly filtered model estimates, unemployed labour force at the national level.

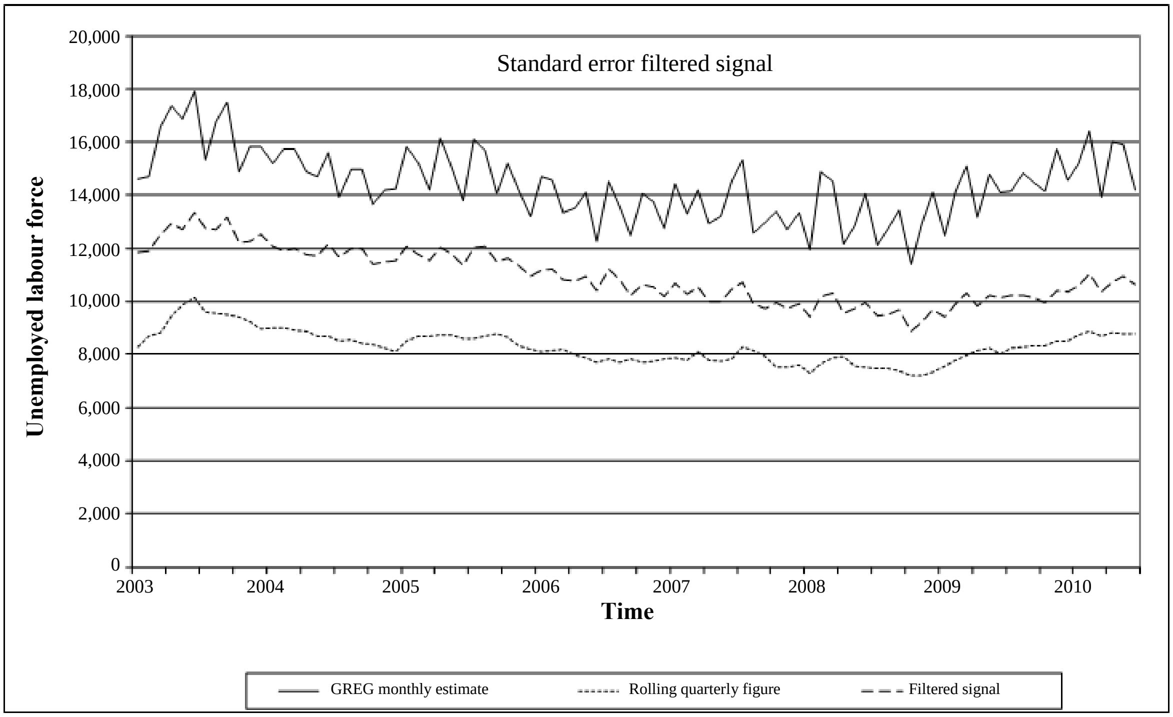


Figure 4.1 (cont.) **Standard errors monthly GREG estimates, rolling quarterly GREG estimates and monthly filtered model estimates, unemployed labour force at the national level.**

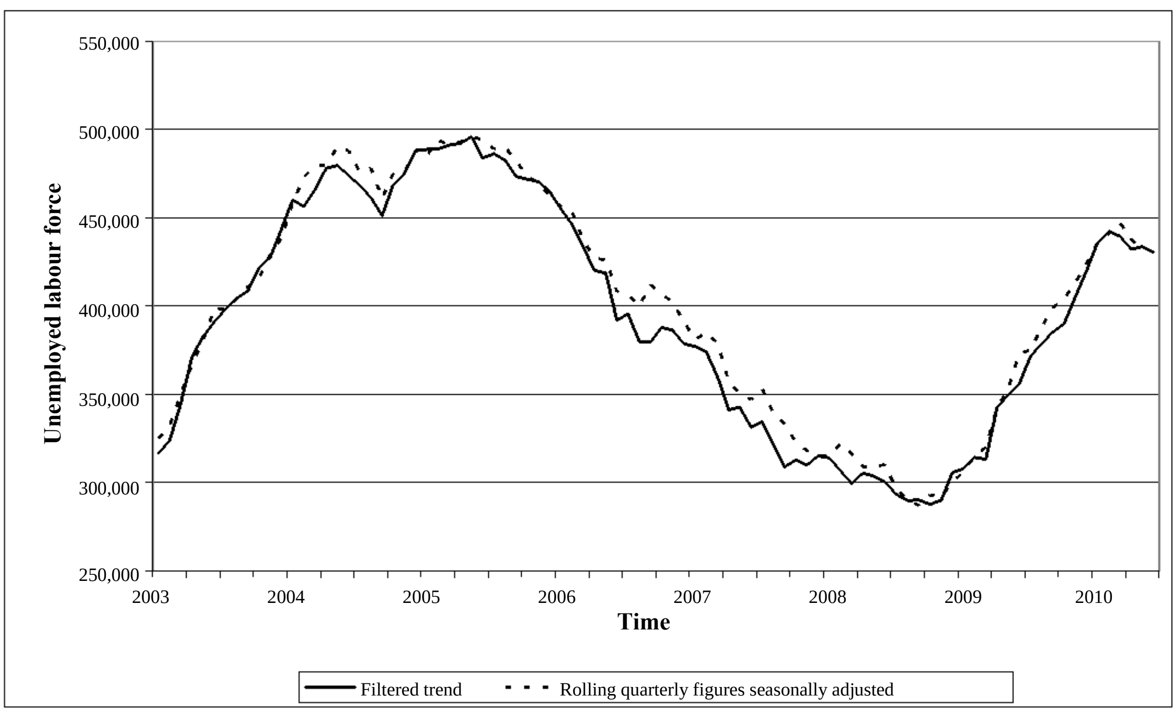


Figure 4.2 **Seasonally adjusted rolling quarterly figures and monthly filtered trend estimates, unemployed labour force at national level.**

Secondly, the RGB correction for the monthly GREG estimates and the rolling quarterly figures are based on a rigid and untested assumption of a constant ratio over a period of three years, see Section 2. In the time series model, the RGB is modelled as differences between the panels and is allowed to change gradually over time, see equation (3.7). Filtered estimates for the RGB in the monthly unemployed labour force at national level are plotted in Figure 4.3. This figure shows that the assumption of a constant ratio over a period of three years is not tenable, since the absolute value of the RGB increases in a period that the unemployed labour force decreases. It is therefore unlikely that the ratio used to correct the rolling quarterly figures is constant over three year periods. The model evaluation does not indicate that the assumptions underlying time series model (3.1) are not met. It can therefore be expected that a more reliable RGB correction is obtained with the time series modelling approach.

Thirdly, the methodology of X-12-ARIMA assumes that there is no autocorrelation in the sampling errors. This assumption is clearly not met in a rotating panel. Pfeffermann et al. (1998) showed that the use of X-12-ARIMA to series with autocorrelated survey errors results in spurious trend estimates. This partially explains the differences between the filtered trend and the seasonally adjusted rolling quarterly data in Figure 4.2.

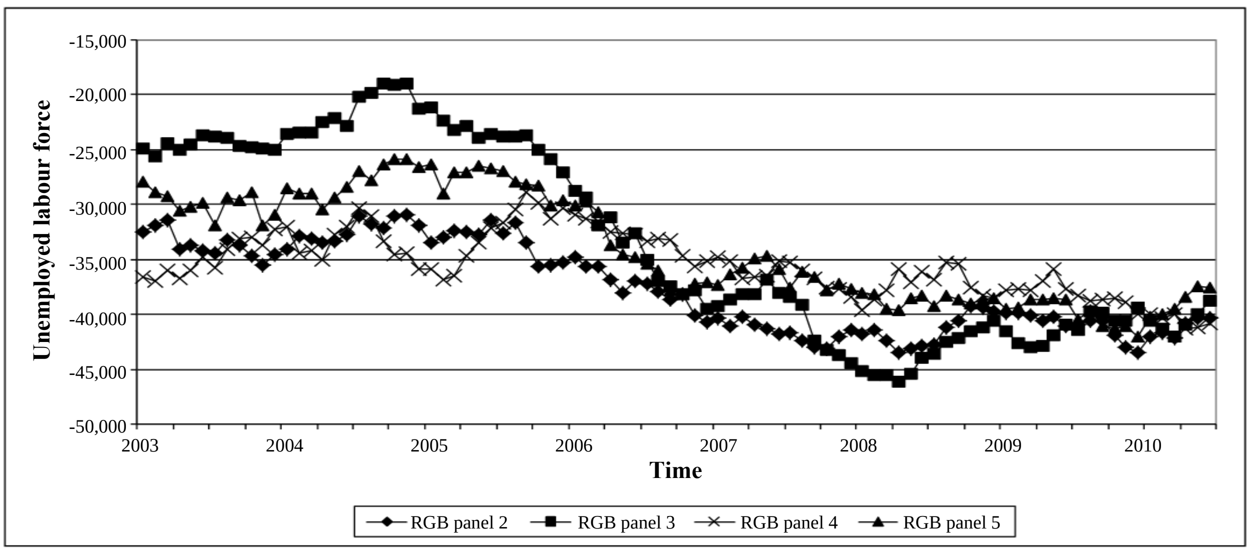


Figure 4.3 Filtered estimates for RGB in the monthly unemployed labour force at national level.

The standard errors of the monthly GREG estimates and the rolling quarterly figures are based on the variance of the Taylor approximation of the GREG estimator, Särndal et al. (1992), Chapter 6. The ratio used to correct for RGB is assumed to be known, although it is based on the samples of three years. The standard errors of the filtered estimates ignore the uncertainty of using ML estimates for the hyperparameters. Table 4.2 compares the means of the standard errors over the last 24 months for the three considered methods for the unemployed labour force, at the national level and for the six domains. Figure 4.1 compares the standard errors at the national level for the three methods for the entire series. In all cases, the precision of the monthly GREG estimates has been substantially improved by the time series model. The rolling quarterly figures have smaller standard errors than the model estimates in almost all cases. For the domains men $15-24$ and women $45-64,$ the precision of the model estimates and of the rolling quarterly figures are similar. Nevertheless, the time series model produces sufficiently reliable monthly estimates to replace the rolling quarterly figures by monthly figures. This circumvents the aforementioned disadvantages of the rolling quarterly figures. Moreover it is not straightforward how rolling quarterly figures can be corrected for RGB in combination with discontinuities induced by the redesign in 2010.

Table 4.2
Mean standard errors unemployed labour force over 24 months (July 2008 – June 2010)

	National level	Men 15-24	Women 15-24	Men 25-44	Women 25-44	Men 45-64	Women 45-64
Rolling quarterly estimate	8,118	3,126	2,831	4,041	3,809	3,452	3,260
Monthly GREG estimate	14,172	5,448	4,885	7,083	6,662	6,046	5,676
Model estimate	10,082	3,247	3,439	5,075	4,749	4,119	3,269
Ratio model and rolling quarterly figure	1.24	1.04	1.21	1.26	1.25	1.19	1.00
Ratio model and monthly GREG estimate	0.71	0.60	0.70	0.72	0.71	0.68	0.58

An artefact of applying model (3.1) to each variable and domain separately is that the sum over the domain estimates is not exactly equal to the estimate at the national level and that the sum of the employed and unemployed labour force is not exactly equal to the total labour force for each domain and at the national level. With the GREG estimator these estimates are consistent by definition, since one set of weights is used to compile all required estimates. The aforementioned restrictions for the model estimates are restored through an appropriate Lagrange function, which distributes the discrepancies over the model estimates proportional to their MSE estimates. Details are given in the Appendix. Finally, unemployment rates are obtained as the ratio of the model estimate for the unemployed labour force to the total labour force for the six domains and the national level.

The model-based domain estimates for the monthly employed and unemployed labour force are included as a weighting term in the GREG estimator for the quarterly and yearly releases. This enforces consistency between monthly, quarterly, and yearly labour force figures and corrects for the RGB in the GREG estimates of the quarterly and yearly labour force figures.

5 Redesign of the Dutch Labour Force Survey

The LFS was redesigned in 2010, as described in Section 2. Discontinuities induced by this redesign were quantified by conducting the first panel under the old and new design in parallel for a period of six months, from January through June 2010. Each month two separate samples with the regular monthly sample size were drawn from the target population according to the sample design of the LFS. One sample was assigned to the old and one to the new LFS design. This made a direct estimate possible for the discontinuities for the main parameters in the first panel.

Mainly due to budget constraints, the subsequent panels were not conducted in parallel under the old and the new design. Possible discontinuities were quantified using the intervention approach described in Section 3. In the time series model, the outcomes of the subsequent panels are benchmarked to the level of the first panel. It is therefore crucial that the first panel is measured as accurately as possible, including possible discontinuities due to a redesign. Therefore it was decided to conduct a sufficiently large parallel run for the first panel, and use the intervention approach for the remaining panels. The estimates for the discontinuities from the parallel run as well as the intervention variables of the time series model are the effect of all factors that changed simultaneously in the redesign of the survey.

In the parallel run, 19,150 responding households under the old design and 16,906 responding households under the new design were obtained. Table 5.1 compares the field work results of the new and old design, both for households with and without a listed phone number. Overall, the response rate is lower for households without a listed phone number. This can be explained by the fact that this part of the population typically consists of hard to reach groups like young people and migrants. Furthermore, the response rate is lower under CATI than under CAPI for households with a listed phone number. Both the percentages of no contact and of frame errors increase substantially when using CATI instead of CAPI. Frame errors under CAPI are mostly non-existing or unoccupied addresses, under CATI they are mostly closed phone lines. Other non-response includes, for example, illness.

Table 5.2 summarizes the estimation results of the parallel run for the unemployed labour force. At the national level, the change-over to the new design resulted in an increase of about 55,000 in the monthly unemployed labour force figures. The differences fluctuated considerably over the six months of the parallel run, probably caused by the large sampling errors of the GREG estimates. A strong increase in the differences was observed in the last two months of the parallel run, particularly at the national level. This can be explained partially by the low response under the new design during these two months.

The decision was made to produce official monthly figures using the data obtained under the old design until June 2010. After completion of the parallel run, all the available data obtained under the new design were used to compile official monthly figures. So since July 2010, the data in the first panel have been based on the new design from January 2010, while the data in the second panel are based on the new design from April 2010, and the data in the third panel are based on the new design from July 2010 and so on.

Table 5.1
Overview fieldwork results of the parallel run first panel

OLD						
	CAPI - phone		**CAPI – no phone**		**total**	
Category	households	%	households	%	households	%
Total	20,813	100.0%	14,469	100.0%	35,282	100.0%
Frame errors	769	3.7%	1,039	7.2%	1,808	5.1%
Not approached	618	3.0%	463	3.2%	1,081	3.1%
Language problems	390	1.9%	878	6.1%	1,268	3.6%
Refusal	4,909	23.6%	3,112	21.5%	8,021	22.7%
No contact	889	4.3%	1,455	10.1%	2,344	6.6%
Other non-response	921	4.4%	689	4.8%	1,610	4.6%
Complete response	12,317	59.2%	6,833	47.2%	19,150	54.3%
NEW						
	CATI		**CAPI**		**total**	
Category	households	%	households	%	households	%
Total	20,234	100.0%	13,345	100.0%	33,579	100.0%
Frame errors	1,539	7.6%	982	7.4%	2,521	7.5%
Not approached	1	0.0%	428	3.2%	429	1.3%
Language problems	317	1.6%	788	5.9%	1,105	3.3%
Refusal	4,545	22.5%	2,903	21.8%	7,448	22.2%
No contact	2,233	11.0%	1,333	10.0%	3,566	10.6%
Other non-response	963	4.8%	641	4.8%	1,604	4.8%
Complete response	10,636	52.6%	6,270	47.0%	16,906	50.3%

To analyse differences in response distributions between the old and the new design, results must be compared column-wise.

Table 5.2
Comparison of GREG estimates new and old design for monthly unemployed labour force figures, first panel (×1,000), standard errors in brackets, significant difference at a 5% significance level indicated with *

	National level	Men 15-24	Women 15-24	Men 25-44	Women 25-44	Men 45-64	Women 45-64
Monthly unemployed labour force new design – mean over January-June							
	475	67	56	103	101	80	68
Difference new and old design monthly unemployed labour force							
Mean January – June	55*(17)	19*(6)	7 (6)	-1 (9)	20*(8)	6 (8)	4 (7)
Difference per month							
January	56 (39)	13 (14)	1 (14)	-15 (21)	-16 (18)	52*(18)	22 (15)
February	38 (42)	41*(16)	9 (17)	-10 (22)	24 (21)	-41*(18)	15 (18)
March	1 (41)	-2 (15)	-11 (13)	-18 (21)	29 (21)	6 (19)	-4 (14)
April	55 (40)	-2 (13)	17 (17)	17 (21)	36 (20)	0 (17)	-13 (16)
May	70 (44)	20 (15)	17 (13)	12 (27)	14 (21)	4 (20)	3 (15)
June	110*(41)	41*(15)	10 (14)	6 (21)	35 (18)	13 (20)	5 (17)

6 Accounting for discontinuities in the time series model

The parallel run showed that the redesign resulted in discontinuities in the series of the monthly figures about the labour force. To avoid severe model misspecification, the intervention term $\boldsymbol{\Delta}_t\boldsymbol{\beta}$ has to be included in model (3.1). An additional question is how the available information about the discontinuities in the first panel, obtained with the parallel run, can be used efficiently in the time series model. Six different methods to use the available information from the parallel run in model (3.1) and (3.9) are discussed.

Method 1: Model (3.1) with a diffuse prior for all intervention variables.

The time independent regression coefficients of the intervention variables for all five panels are included in the state vector and initialised with a diffuse prior, as described by Durbin and Koopman (2001), Subsection 6.2.2. The Kalman filter can be applied straightforwardly to obtain estimates for the regression coefficients. This approach ignores the information about the discontinuities that is available from the parallel run. In this application, this approach is interesting since comparing the time series model estimate for the discontinuity in the first panel with the direct estimates obtained with the parallel run illustrates how well discontinuities can be estimated with the intervention approach.

Method 2: Model (3.1) with an exact prior for the intervention variable of the first panel.

The direct estimates of the discontinuities from the parallel run are incorporated into the model by using an informative prior for the initialization of β^1. This can be done by using these estimates in the initial state vector for β^1 and their estimated variances as an uncertainty measure for β^1 in the covariance matrix of the initial state vector.

Method 3: Model (3.1) where the regression coefficient of the intervention variable for the first panel equals the average direct estimate for the discontinuity obtained with the parallel run.

Another possibility of using the direct estimate of the discontinuities in the first panel as a-priori information in model (3.1), is to assume that the regression coefficient for the intervention in the first panel is time independent and equal to the average value of the observed discontinuity in the parallel run, i.e.,

$$\bar{\beta}^{1} = \frac{1}{6}\sum_{t=\tilde{t}}^{\tilde{t}+5}\left(\hat{Y}_{t}^{t,\mathrm{New}} - \hat{Y}_{t}^{t,\mathrm{Old}}\right),$$

where $\tilde{t}$ denotes the start of the parallel run in January 2010. In this case the direct estimate for the discontinuity is treated as if it is a fixed value, known in advance. This approach ignores the uncertainty of using a survey estimate for the discontinuity.

Method 4: As method 3, but with a time dependent regression coefficient for the intervention variable of the first panel.

The direct estimates for the discontinuities fluctuate considerably over the six months of the parallel run, see Table 5.2. To have a smooth transition from the old to the new design, an alternative for method 3 is considered where during the parallel run, the regression coefficient of the first panel is time dependent and equals the observed monthly discontinuities. For the period after the parallel run, this regression coefficient is equal to the average value of the observed discontinuity in the parallel run, i.e.,

$$\beta_{t}^{1} = \begin{cases} \hat{Y}_{t}^{t,\mathrm{New}} - \hat{Y}_{t}^{t,\mathrm{Old}} & \text{if} \quad t \in [\tilde{t}, \ldots, \tilde{t} + 5] \\ \bar{\beta}^{1} & \text{if} \quad t > \tilde{t} + 5. \end{cases}$$

This method comes down to replacing the observations under the new design by the observations under the old design during the parallel run and assumes that the results under the old design are more reliable during this period. Similar to method 3, the uncertainty of using a survey estimate for the discontinuity is ignored.

The four methods can be applied to model (3.9) that is extended with an auxiliary series about the number of people formally registered at the employment office. The following two methods are considered:

Method 5: Equals Method 1 applied, to model (3.9).

Method 6: Equals Method 4 applied, to model (3.9).

In practise, method 1 would be considered if no parallel run is available. In the case of a well conducted parallel run, method 2 is probably the most natural approach, because the sample estimate for the discontinuity together with its uncertainty are used as prior information in the model. The sample information that becomes available after the parallel run under the new design is still used to improve the estimate of the discontinuity. Methods 3 and 4 are considered as alternatives for method 2 for getting a smoother transition from the estimates obtained until June 2010 under the old design to the estimates under the new design, starting in July 2010. Method 3 might work well if the variation between the monthly estimates for the discontinuity during the parallel run is small. In the case of large fluctuations between the monthly discontinuities, method 4 might be considered because during the parallel run each monthly deviation of the estimate under the new design is nullified with the time dependent discontinuities. Method 4 will therefore result in the smoothest transition.

In the case of strong and reliable auxiliary information, each method can be combined with model (3.9). It is a requirement, however, that the evolution of this auxiliary series is not influenced by factors that are unrelated to the real developments of the labour market. Method 5 would be considered if no parallel run is available. The auxiliary series might result in more precise estimates for the discontinuity and the trend and signal of the unemployed labour force. In the case of a parallel run, method 2 in combination with model (3.9) is probably the most natural approach for similar reasons as mentioned before (results not presented). Method 6 can be used to get a smoother transition from the old to the new design and more precise estimates for the trend and the signal of the unemployed labour force by taking advantage of the available auxiliary information. For similar reasons method 3 can be combined with model (3.9) (results not presented).

7 Results

7.1 Estimation results for the national level

Results are presented for the monthly unemployed labour force figures at the national level. The filtered estimates for the discontinuities in panels 1 and 2 are plotted in Figure 7.1. Figure 7.2 compares the filtered estimates of the RGB in panel 2 under the six different methods from January 2006 until March 2012, and the filtered RGB obtained under the old data until June 2010. Results for the other panels are similar and therefore omitted. Figure 7.3 compares the filtered trend estimates under the six different methods from July 2009 until March 2012, with the filtered trend estimates obtained under the old data until June 2010.

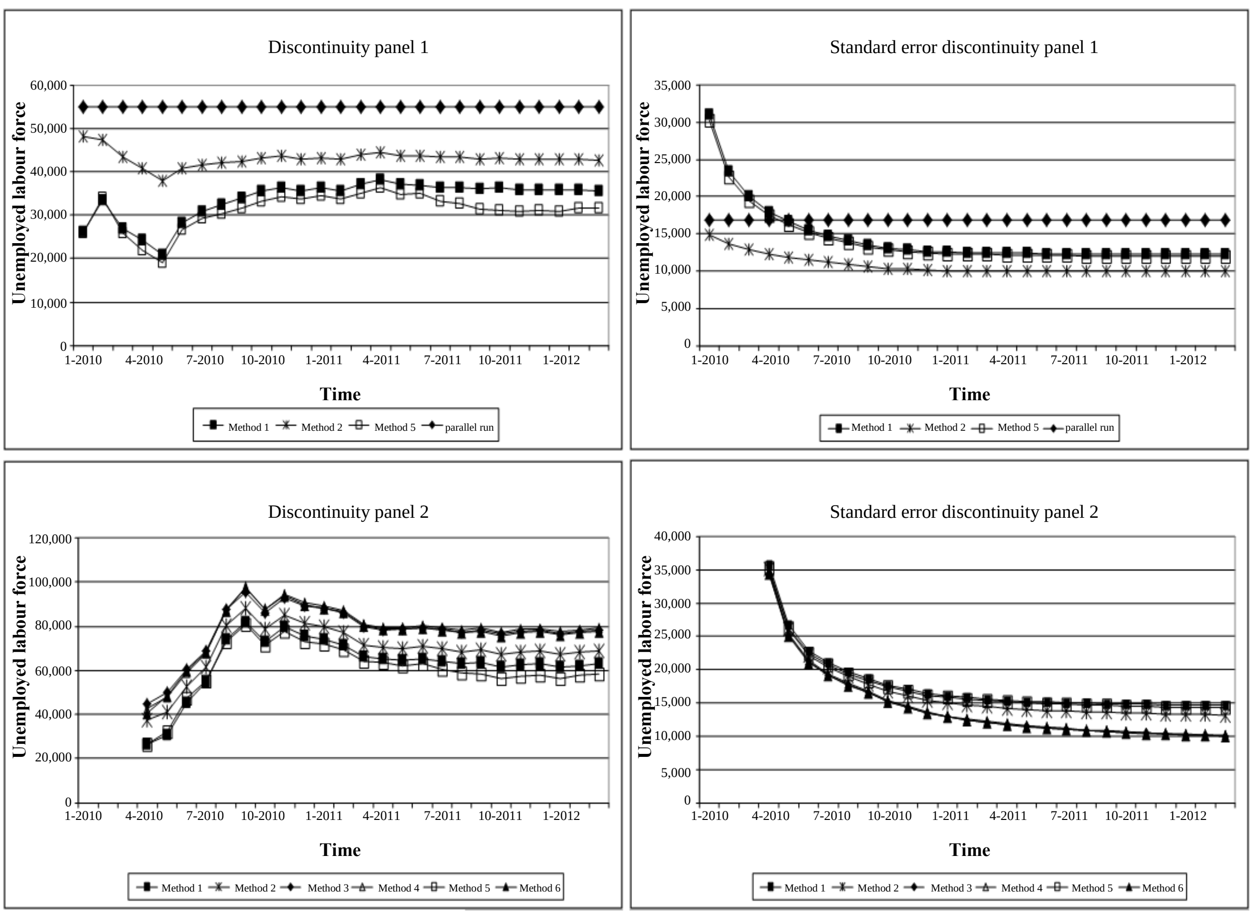
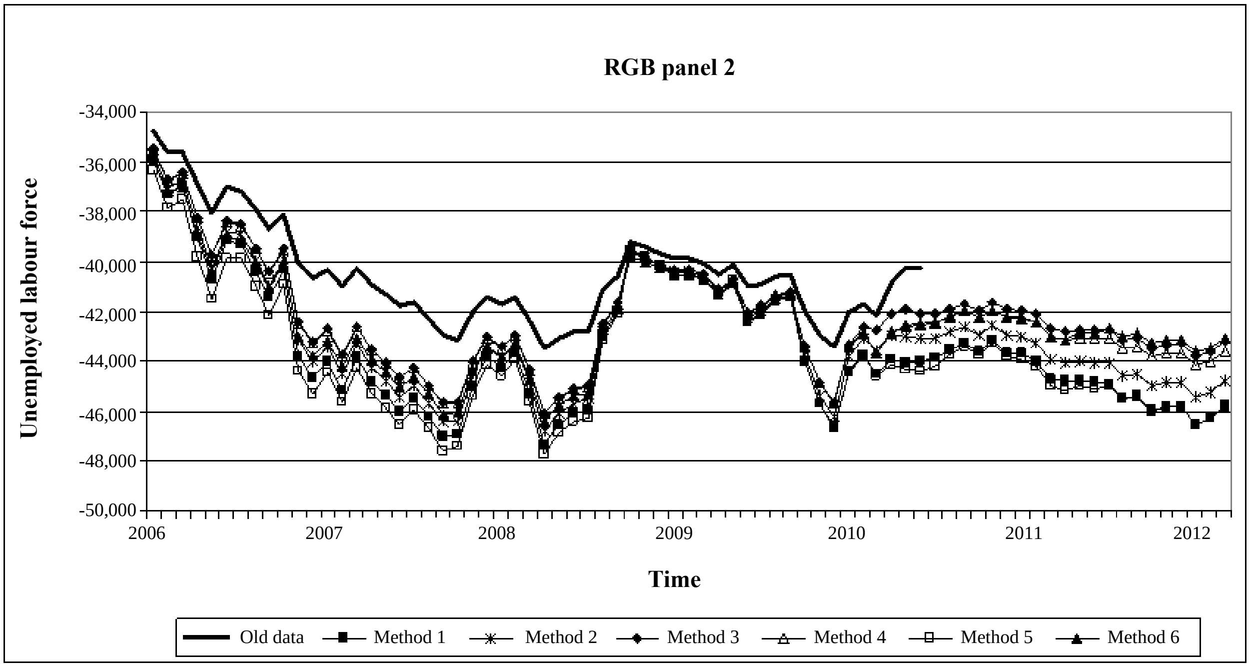


Figure 7.1 Filtered estimates for discontinuities and their standard errors January 2010 – March 2012, panel 1 and 2 for monthly unemployed labour force at national level.



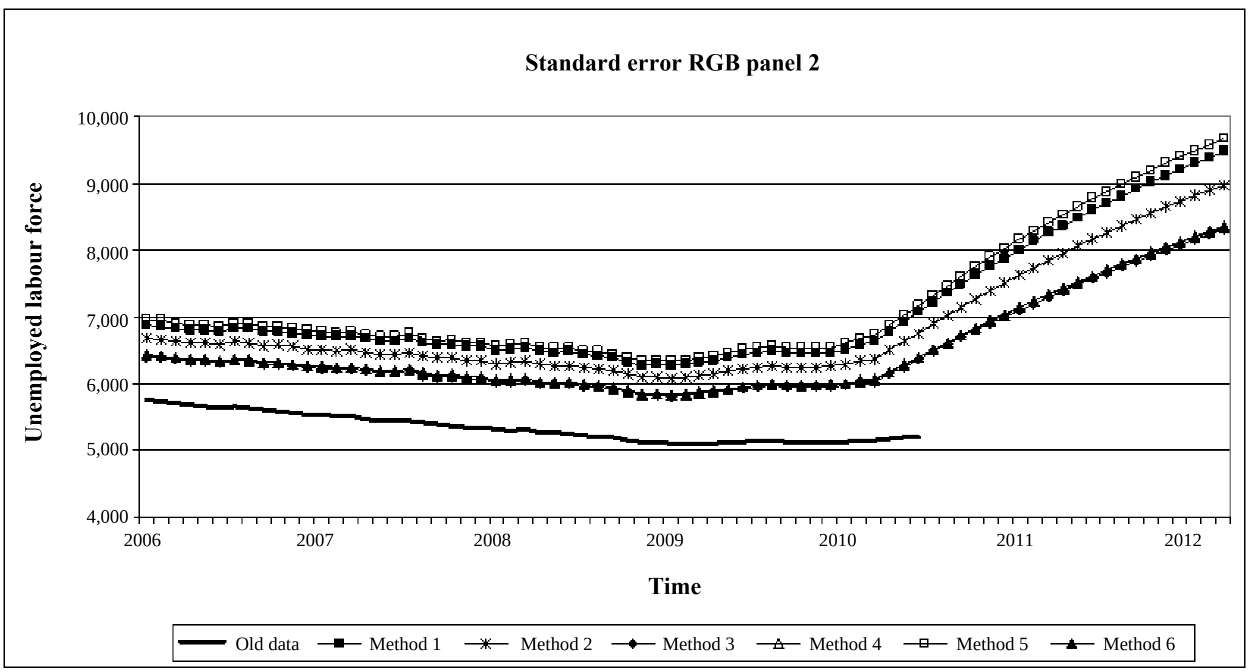


Figure 7.2 **Filtered estimates for RGB and their standard errors panel 2 for monthly unemployed labour force at national level January 2006 – March 2012 for six different methods that account for discontinuities and the old data.**

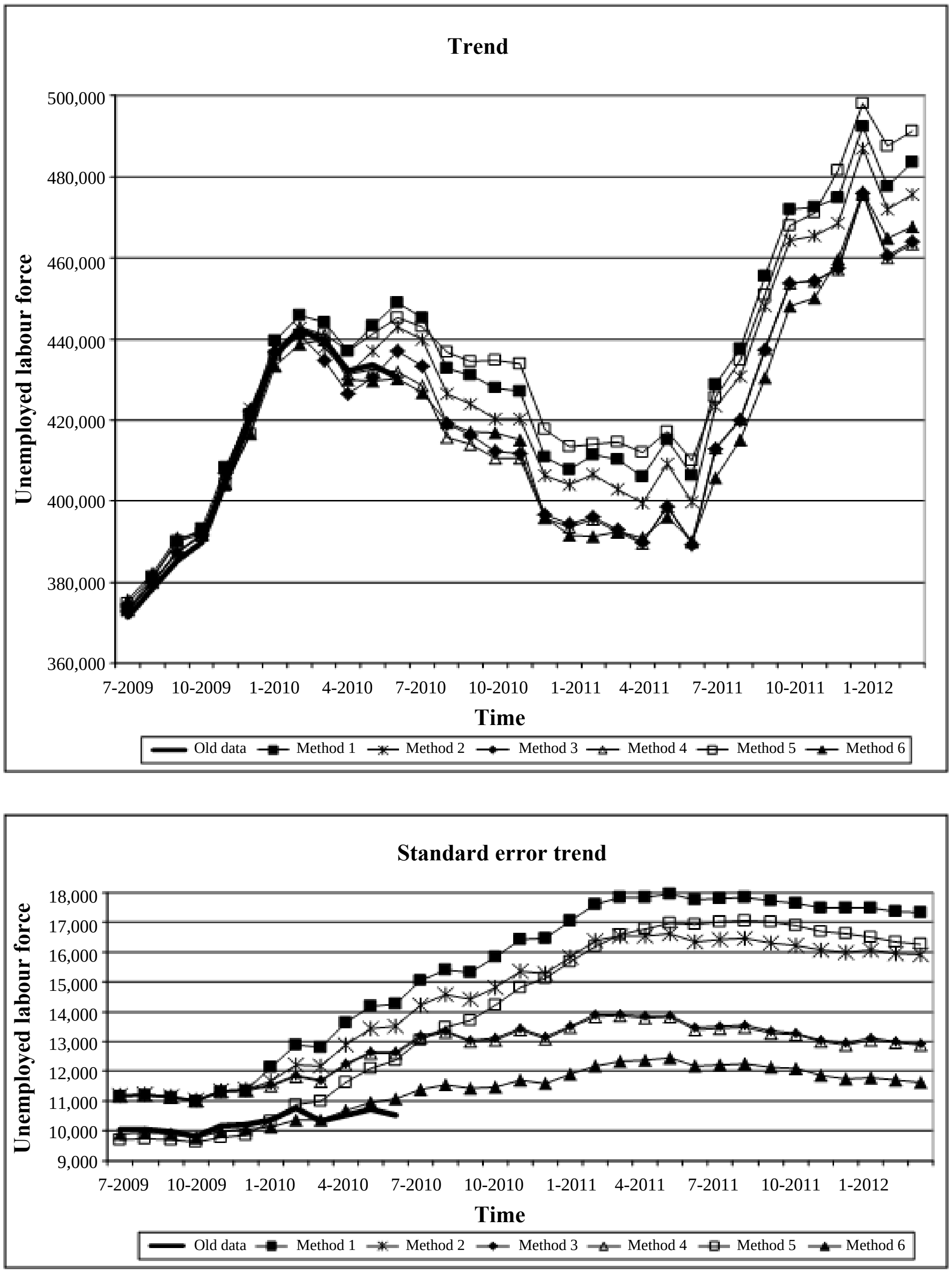


Figure 7.3 **Filtered trend of monthly unemployed labour force at the national level and their standard errors July 2009 - March 2012 for six different methods that account for discontinuities and the old data.**

Figure 7.1 shows that the different methods lead to different estimates for the discontinuities. The filtered estimates for the regression coefficient of the intervention variable in the first panel are systematically smaller than the direct estimate obtained in the parallel run. The smallest estimate is obtained if a diffuse prior is used to initialise this regression coefficient (method 1 and 5). Extending the model with an auxiliary series resulted in a slightly smaller estimate (compare method 1 and 5). Using the direct estimate from the parallel run as an exact prior for the regression coefficient (i.e., method 2) resulted, as expected, in an estimate that is closer to the direct estimate obtained with the parallel run.

The standard errors of the regression coefficients of the interventions follow a smooth exponentially decreasing pattern. Already five months after the change-over to the new design, the standard errors of the regression coefficients initialised with a diffuse prior became smaller than the standard error of the direct estimate for the discontinuity obtained in the parallel run. The standard errors of the regression coefficients initialised with an exact prior were, as expected, immediately smaller than the standard error of the direct estimate.

The estimated discontinuities in panel 2 through 5 follow the same pattern as the estimates observed in panel 1. Methods with small estimates for the discontinuity in panel 1, also have the smallest estimates in the subsequent panels and vice versa. As described below, the estimate of the discontinuity in the first panel strongly influences the estimated level of the trend. This explains why the method used to quantify the discontinuity in the first panel also influences the estimated discontinuities in the subsequent panels. Extending the model with an auxiliary series hardly affects the estimated discontinuities (method 6 versus 3 and 4, method 5 versus 2). On average the estimated regression coefficients become more or less stable about one year after the change-over. By using the exact prior in the first panel (method 2), a stable estimate for the discontinuity in the first panel is obtained after about half a year. The auxiliary series, on the other hand, do not decrease the required period to obtain a stable estimate.

The filtered estimates for the discontinuities are affected by the model choice of the RGB. Since the model for the RGB is time dependent, the filtered estimates for the RGB may partially absorb the discontinuities induced by the redesign. Therefore the filtered estimates for the regression coefficients do not reflect the absolute effect of the redesign. They nevertheless avoid model misspecification due to discontinuities in the input series. More realistic estimates for the discontinuities are obtained with a model were the RGB is time invariant $(\text{i.e., } \sigma_\lambda = 0)$. Under this model, the estimated discontinuities for the first panel indeed increase with about 7,000 persons under method 1, 2, and 5 and come closer to the direct estimate for the discontinuity observed in the parallel run (results not presented).

The standard errors of the regression coefficients in panel 2 through 5 are affected by the method used to estimate the discontinuity in the first panel. Method 3, 4 and 6, which use the direct estimate from the parallel run for the discontinuity in the first panel have the smallest standard errors and are more or less equal. Method 1 and 5, which use a diffuse prior for the regression coefficient for the discontinuity in the first panel, have the largest standard errors for the discontinuities in the subsequent panels. Method 2, which uses an exact prior in the first panel, has standard errors that are somewhere in between.

Figure 7.2 shows that the filtered RGB is also influenced by the intervention term and the method used to estimate the discontinuity in the first panel. Most striking is the difference between the RGB with the data observed under the old approach only, and the RGB obtained with the six methods that include the data under the new approach, during the period before the change-over to the new design. These differences can be explained with differences between the ML estimates for the hyperparameter of the RGB $(\hat{\sigma}_\lambda)$. Adding the data observed under the new design and augmenting the model with an appropriate intervention term increases $\hat{\sigma}_\lambda$ with a factor of about 1.4 (compare Tables 4.1 and 7.2).

After the change-over to the new design, the estimates for the RGB become less volatile than in the period before the change-over. The level of the RGB after the change-over also depends on the method used to quantify the discontinuity in the first panel. As will be explained below, the value for the discontinuity in the first panel determines the level of the trend in the first panel and therefore also the relative bias, i.e., the RGB, in the subsequent panels with respect to the first panel.

The evolution of the standard errors of the filtered RGB shows a smooth pattern. The standard errors for the RGB under the old design are substantially smaller since the ML estimate for the hyperparameter is smaller compared to the methods that include the data observed under the new design. The introduction of the five intervention variables, starting in January 2010, introduced additional uncertainty in the estimated RGB. As a result the standard errors consistently increased after January 2010. It is remarkable that they did not stabilize within the observed period, like the standard errors of the trends (see below). This might be caused by the fact that the discontinuities simultaneously influence the intervention variables and the RGB parameters and could be an indication that the model has difficulties separating both effects with a model that allows for time dependent RGB. A model with constant RGB has indeed a constant standard error for the RGB after the change-over. The problems with model identification increased with the flexibility of the RGB.

The order of the standard errors of the RGB under the six methods is equal to the results observed for the standard error of the discontinuities. Similar results hold for the RGB in the other three panels.

Figure 7.3 shows that the level of the trend (and also the signal) strongly depends on the choice of the method used to estimate the discontinuities. Larger estimates for the discontinuities resulted in smaller levels for the trend and vice versa. The evolution of the trend is more or less similar under the six methods.

Before the change-over, the standard errors of the trend under the new design were larger compared to the method that only uses the old data, with the exception of method 5 and 6, which are based on the model extended with an auxiliary variable. This difference can be attributed to the increased flexibility of the RGB as described before. Methods 5 and 6 have more or less the same standard error as the method based on the old data only. The disturbance terms of the slope of the auxiliary series and the monthly unemployed labour force were strongly correlated (about 0.9). This resulted in a substantial decrease of the standard error of the filtered trend and neutralized the increase of the standard error due to the increased flexibility of the RGB.

Each time a panel changes to the new design, the standard error of the filtered trend increases under each of the six methods and stabilizes after the change-over in the fifth panel. Methods 1 and 5, which use a diffuse prior for the discontinuity in the first panel showed the largest increases in the standard error at each time a new intervention variable modelled the change-over to the new design in a panel. Recall from Figure 7.1 that the standard errors for the discontinuities in the five panels are the largest under these two methods. The standard error for the trend under method 5 is smaller than in method 1, since this method takes advantage of a strongly correlated auxiliary series. Method 2, which uses an exact prior, follows more or less the same pattern, but had smaller increases in the standard error. Methods 3, 4, and 6, which use the direct estimate for the discontinuity in the first panel, had the smallest increase in the standard error of the trend, since they had the smallest standard error for the four discontinuities in panel 2 through 5 and ignored the standard error of the direct estimate for the discontinuity in the first panel. The standard errors for method 3 and 4 were equal. The standard errors for method 6 were smaller since this method benefited from the correlated auxiliary series.

We do not present the results for filtered slopes and seasonals but just mention that the standard errors of these state variables are not affected by the change-over to the new design in the different panels.

7.2 Estimation results for domains

Roughly speaking, similar results are observed for the six domains. Table 7.1 summarizes the trend and the discontinuities in the first panel with their standard errors averaged over the last 12 months of the six domains and the national level for the six methods. For method 5 and 6 the ML estimates for the correlation between the disturbances of the slopes are also included. The differences between the direct estimates for the discontinuities and the regression coefficients of the intervention in the first panel are in some cases larger compared to the national level. This can be expected since the sample size in the domains is smaller, resulting in less precise direct estimates for the discontinuities.

Methods 5 and 6, which take advantage of a correlated auxiliary series, showed a stronger decrease for some of the domains of the standard error of the filtered trend compared to the national level. In these cases, the ML estimates of the correlation were larger and sometimes equal to one, which implies that the trend of the auxiliary series and the unemployed labour force are or tend to be cointegrated.

Table 7.1
Trend and discontinuities panel 1 averaged over the last 12 months of the national level and the six domains for the six different methods used to quantify the discontinuity in the first panel. Standard errors between brackets

Parameter	Method	National level	Men 15-24	Women 15-24	Men 25-44	Women 25-44	Men 45-64	Women 45-64
Trend	1	452 (18)	58 (5)	40 (5)	78 (8)	100 (7)	87 (7)	82 (6)
	2	445 (16)	53 (5)	41 (5)	83 (8)	95 (7)	85 (7)	79 (6)
	3	435 (13)	45 (4)	44 (4)	95 (6)	83 (5)	82 (5)	73 (4)
	4	434 (13)	45 (4)	44 (4)	95 (6)	83 (5)	82 (5)	73 (4)
	5	454 (17)	58 (4)	43 (4)	78 (8)	98 (6)	77 (4)	83 (6)
	6	433 (12)	45 (4)	45 (3)	92 (5)	83 (3)	76 (3)	74 (4)
Disc. panel 1	1	36 (12)	5 (4)	11 (4)	17 (6)	3 (5)	2 (5)	-4 (5)
	2	43 (10)	11 (3)	10 (3)	11 (5)	8 (4)	3 (4)	-2 (4)
	5	33 (12)	6 (3)	10 (4)	15 (6)	5 (5)	5 (4)	-5 (5)
	3, 4, 6	55 (17)	19 (6)	7 (6)	-1 (9)	20 (8)	6 (8)	4 (7)
Corr. slope	5	0.93	0.98	0.99	0.93	0.99	1.00	0.87
	6	0.88	0.72	1.00	0.99	1.00	1.00	0.90

7.3 Model choice

As a consequence of the strong correlation between the disturbances of the slopes, the auxiliary series has a notable effect on the level of the filtered trend. Using a model that includes this auxiliary series therefore implies that there must be great confidence in the quality of the auxiliary series. Amendments in the law with respect to unemployment benefits and social benefits, or sudden changes in the mode of operation of the employment office, may result in sudden or gradual differences in the number of people formally registered at the employment office. This would not be a problem if the ML estimates for the correlation between the disturbances of the slopes became smaller. Simulations where level breaks as well as gradual increasing disturbances are added to the auxiliary series show that the ML estimates for the correlation are adjusted with an unacceptable large delay. Therefore the auxiliary series may influence the filtered trend estimates for the monthly unemployed labour force incorrectly (results not shown). Since it is known that the evolution of the series of the number of people formally registered at the employment office is influenced by the aforementioned factors, that are unrelated to economic developments, it was decided not to choose methods 5 or 6 to produce official monthly unemployment figures.

The model diagnostics, mentioned in the second paragraph of Section 4, indicate that the innovations under model (3.9) contain more autocorrelation and slightly stronger deviations from the normality assumption than model (3.1). The model diagnostics for the four methods based on model (3.1) are very similar and do not indicate strong violations of the assumption that the innovations are normally and independently distributed. The model diagnostics are not useful for further discriminating between the different methods that rely on the same model (model (3.1) or (3.9)). This is a consequence of the interchange between the estimates for the discontinuities and the trend. As explained before, an increase in the estimated discontinuity is neutralized by an opposite effect on the filtered trend and RGB. As a result, the one-step-a-head predictions for the signals and the innovations in the different panels are more or less equal under all methods.

The main purpose of modelling discontinuities is to avoid that developments of labour force indicators are erroneously influenced by the change-over to the new survey process. The preferred method describes the development of the monthly labour force figures most accurately. The choice between methods 1 through 4 can therefore be based on the confidence in the different estimates for the discontinuities, using additional information such as knowledge from subject matter experts. Comparing the filtered trends under the different methods with the officially published figures during the parallel run is also useful for evaluating which method results in the smoothest transition during the change-over.

Recall from Section 5 that the model estimates obtained under the old data were published as the official monthly release until June 2010. The figure to be published for July 2010 must be based on one of the new methods, where the observations in the time series of the first panel changed from the old to the new method in January 2010 (see Section 5). From Figure 7.3 it follows that during the parallel run the filtered trend obtained with method 4 is, from the methods based on model (3.1), the closest to the officially published trend obtained with the old data. It can therefore be expected that this method will result in the smoothest transition in the month that the data under the new approach are used for the first time. According to labour market experts, there were no indications that the steady downward trend of the monthly unemployed labour force could change into an upward trend at that time. As follows from Figure 7.3, method 4 is the only method based on model (3.1) that resulted in a continued downward trend.

Based on the aforementioned considerations, method 4 was finally chosen to produce official statistics about the monthly labour force. With method 4, the GREG estimates in the first panel were corrected back to the outcomes under the old design during the parallel run, and this resulted in the smoothest and most plausible transition to the new method.

7.4 Implementation

ML estimates for the hyperparameters based on method 4 at the national level and the six domains are presented in Table 7.2. In Figure 7.4, the five GREG series are plotted with the filtered trend based on the model, which is currently used to produce official model-based estimates for the monthly unemployed labor force figures. The detail of this figure is not important. The purpose is to illustrate how noisy the five input series of the GREG estimates are and how, with the time series model, a filtered trend from this input is obtained. Until 2010, the level of the filtered trend was equal to the level of the GREG estimates of the first panel, since the model removes the RGB by benchmarking the outcomes to the level of the series obtained in the first panel. In 2010 the change-over to the new design started. The discontinuities resulted in higher levels for the series of GREG estimates of the five panels. In this application, the time series model estimates figures that are corrected for these discontinuities. As a result, the filtered trend drops below the level of the series observed with the first panel after 2010.

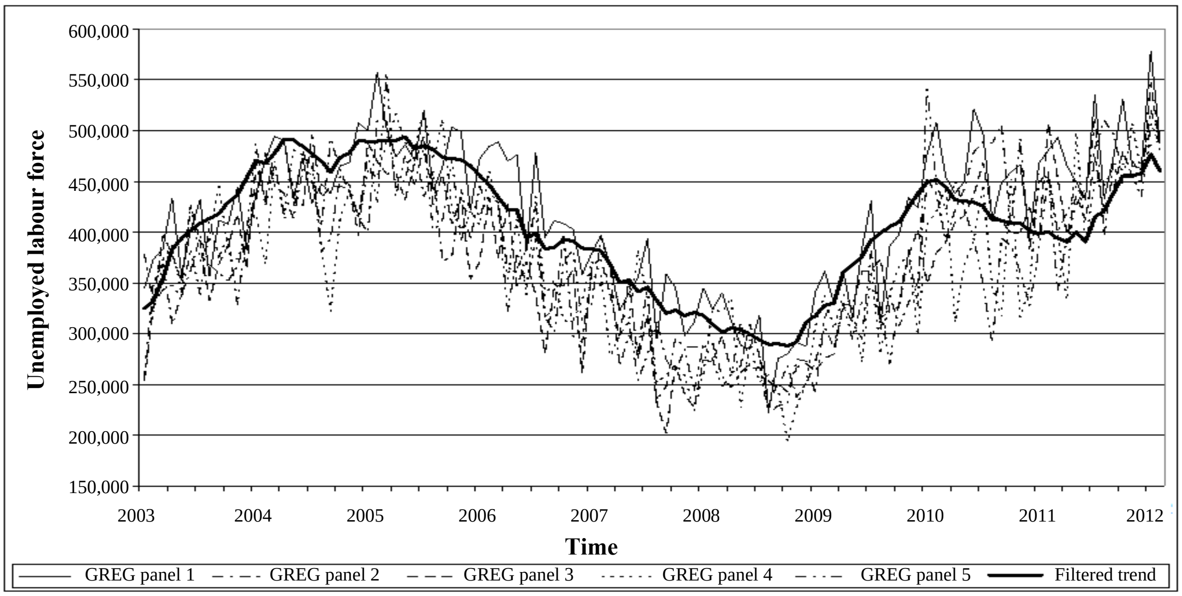


Figure 7.4 **Unemployed labour force at the national level; GREG estimates of the five panels and filtered trend based on a structural time series model.**

Table 7.2
ML estimates of hyperparameters for monthly unemployed labour force figures after the survey redesign. Values are expressed as standard deviations

Standard deviation		National level	Men 15-24	Women 15-24	Men 25-44	Women 25-44	Men 45-64	Women 45-64
Slope	$(\hat{\sigma}_\eta)$	2,423	292	221	703	561	451	207
Seasonal	$(\hat{\sigma}_\omega)$	0.01	0.09	0.00	0.00	0.00	0.00	0.00
RGB	$(\hat{\sigma}_\lambda)$	1,218	931	654	316	567	272	418
White noise	$(\hat{\sigma}_\varepsilon)$	7,720	1,663	3,348	4,128	4,540	4,383	3
Survey error panel 1	$(\hat{\sigma}_{e1})$	0.99	0.93	1.02	1.03	0.97	0.99	1.13
Survey error panel 2	$(\hat{\sigma}_{e2})$	1.03	0.95	1.10	1.16	1.00	1.18	1.14
Survey error panel 3	$(\hat{\sigma}_{e3})$	0.96	1.05	1.15	1.15	1.00	1.18	1.00
Survey error panel 4	$(\hat{\sigma}_{e4})$	1.12	1.05	1.17	1.16	1.03	1.13	1.07
Survey error panel 5	$(\hat{\sigma}_{e5})$	1.13	1.02	1.08	1.11	1.04	1.17	1.01
Autocorrelation	$(\hat{\rho})$	0.257	0.130	0.212	0.430	0.245	0.456	0.411

The hyperparameter estimates for the survey errors for panel 2, 3, 4 and 5 are divided by $(1-\hat{\rho}^2)$. Therefore hyperparameters for the survey errors are, as expected, around 1.

The filtered estimates, considered so far, illustrate what can be accomplished with the state-space approach to produce contemporary estimates in the production of official statistics, i.e., the optimal estimates for period t based on the sample information observed until period t. These filtered estimates, however, can be improved if new information after period t becomes available. Although Statistics Netherlands currently does not revise the contemporary estimates, it is interesting to analyze to what extent the filtered estimates are adjusted if information that becomes available after one, two or three months is used to update the filtered estimates. In Figure 7.5 the filtered trend $\left(L_{t|t}\right)$ is compared with the estimates based on the information of one $\left(L_{t|t+1}\right)$, two $\left(L_{t|t+2}\right)$ and three $\left(L_{t|t+3}\right)$ additional months after period t for the unemployed labour force at the national level. The smoothed series based on the entire series is also included in this figure.

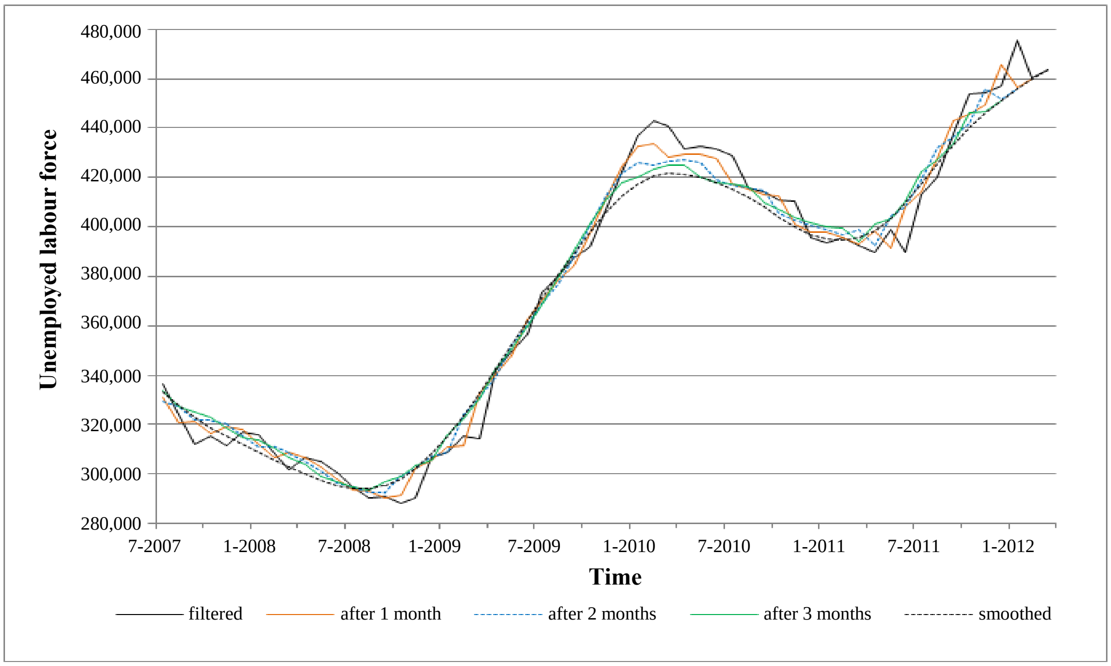


Figure 7.5 **Comparison of filtered trend, revisions after one month $\left(L_{t|t+1}\right)$, two months $\left(L_{t|t+2}\right)$, three months $\left(L_{t|t+3}\right)$, and the smooth trend for the unemployed labour force at the national level.**

The largest revisions occur if the information after the first three months are used to update the filtered estimates. The estimates based on the information observed after three months are already close to the smoothed estimates. Furthermore the revisions during the period of the change-over, starting in January 2010, are larger than in other periods. This is the result of the introduction of the intervention variables. Particularly in this period, the estimates for the intervention variables are based on a few observations under the new design, resulting in large revisions for the discontinuities. This is reflected in larger revisions for the trend during the period of the change-over. In this application, it appears that the first two or three month after period t contain substantial additional information to improve the monthly estimate for period t. It could therefore be considered to base the final estimates for period t on the information observed until $t + 2$ or $t + 3$.

8 Discussion

National statistical institutes widely apply GREG estimators to produce official statistics. The advantage of these estimators is that they are robust against model misspecification, reduce the design variance, and correct at least partially for selection bias in the case of well-specified weighting models. Furthermore, they result in domain estimates which are consistent by definition, and their use in production processes is relatively straightforward since only one set of weights is required to estimate all possible output tables in a multipurpose survey.

GREG estimators, however, have unacceptably large design variances in the case of small sample sizes and do not handle measurement bias in an effective way. The Dutch LFS is an example where these problems require additional estimation procedures. The sample size is too small to produce sufficiently precise monthly labour force figures with the GREG estimator. The rotating panel design and the major redesign of the survey process make differences in measurement bias visible and compromises comparability of outcomes over time. These problems are solved simultaneously with a multivariate structural time series model that uses the series with GREG estimates for the different panels as input. The time series method combines strong points of the GREG estimator with the advantages of a model-based approach. Since time series of GREG estimates as well as their standard errors are used as input series, the method accounts for the complexity of the sample design and corrects for unequal selection probabilities and selective non-response. The time series model accounts for small sample sizes by taking advantage of sample information observed in previous periods, the autocorrelation in the survey errors, the rotation group bias by benchmarking the estimates to the level of the first panel, and discontinuities that arise from a major survey redesign.

We discussed how the model can be extended with a strongly correlated auxiliary series, which is the number of people formally registered at the employment office in this application. Auxiliary information further decreases the standard error of the filtered trend and signal. Also the levels of the filtered estimates are affected by the auxiliary variable. Since there are strong indications that the evolution of the auxiliary series is affected by factors other than economic cycles, and that this improperly affects the monthly filtered trend of the unemployed labour force, it was decided not to use this information in the ultimately selected model. In this application, the auxiliary series hardly influences the estimated discontinuities. This conclusion, however, cannot be generalized. If e.g., the moment of the change-over coincides with a real break in the evolution of the variable of interest, then auxiliary series should contain similar breaks and can provide valuable additional information to disentangle discontinuities from real developments correctly.

If no parallel run is conducted, then discontinuities are estimated through an intervention variable with a regression coefficient initialized with a diffuse prior. In the case of a parallel run, direct estimates for the discontinuities provide additional information that can be used in the time series model. One possibility is to use the direct estimate with its standard error as an exact prior to initialize the regression coefficient of the intervention variable. Another approach is to assume that the regression coefficient is equal to the direct estimate. This approach treats the external information about the discontinuities as if it is observed without error. A well-conducted parallel run has the advantage that it provides a direct estimate for the discontinuities and therefore does not rely on the assumption that, at the moment of the change-over, the evolution of the variables of interest is captured by the time series components other than the intervention variable.

A consequence of modelling discontinuities is that the standard errors of the filtered trend and signal increase each time the new design enters another panel. This illustrates the importance of keeping the survey process unchanged as long as possible and of limiting the number of redesigns.

In conclusion, a time series model is proposed that simultaneously solves problems with small sample sizes, RGB in a rotating panel, and discontinuities due to a redesign. It enables Statistics Netherlands to publish real monthly figures about the labour force, instead of the rolling quarterly figures that are often used as a second best approximation. During the redesign, the model avoids distortion of real developments of the monthly labour force indicators with sudden changes in measurement bias. The method is flexible and of general interest, since most national statistical institutes apply rotating panels for labour force surveys. Furthermore, redesigns of survey processes aimed to reduce administration costs or to improve outdated methods remain inevitable, resulting in loss of comparability of the outcomes over time. Finally there is an increasing interest for small area estimates while there is always pressure to reduce sample sizes due to budget constraints and lowering the response burden.

Acknowledgements

The authors wish to thank the referees and the Associate Editor and Rita Gircour (Statistics Netherlands), for careful reading of the first draft of this paper and providing constructive comments. The views expressed in this paper are those of the authors and do not necessarily reflect the policy of Statistics Netherlands.

Appendix

With the structural time series model (3.1), monthly estimates for the employed, unemployed and the total labour force are computed for the national level and for a breakdown in the six domains. These 21 population parameters are notated by $\theta_{t,l,m}$, where $l = 1, 2, 3$ denotes respectively the employed, unemployed and total labour force, $m = 1$ the national level, and $m = 2, \ldots, 7$ the six domains. For the population parameters, the following consistency requirements hold:

$$\theta_{t,1,m} + \theta_{t,2,m} - \theta_{t,3,m} = 0, m = 1, \ldots, 7 \tag{A.1}$$

$$\sum_{m=2}^{7} \theta_{t,l,m} = \theta_{t,l,1}, l = 1, 2, 3. \tag{A.2}$$

Subscript m runs within l, which in turn runs within t. Because time series model (3.1) is applied to each population parameter separately, requirements (A.1) and (A.2) do not hold for the model estimates.

Therefore, they are restored with a Lagrange function. The model estimates for the national level are changed as little as possible, because they are based on considerably larger samples than the six domains. Therefore, the consistency is achieved in two steps. Both steps are specified for the filtered trends. Consistent filtered signals can be computed in a similar way.

Let $L_{t,l,m}$ denote the filtered trend for $\theta_{t,l,m}$. In the first step, the requirements of equation (A.1) for the national level $(m = 1)$ are considered. The consistency requirement can be written as $\mathbf{\Delta}^{[1]}\mathbf{L}_t^{[1]} = 0$ with $\mathbf{L}_t^{[1]} = \left(L_{t,1,1}, L_{t,2,1}, L_{t,3,1}\right)^T$ a vector with the model estimates for the three trends at the national level and $\mathbf{\Delta}^{[1]} = (1, 1, -1)$ a 3×1 matrix that specifies requirement (A.1). Adjusted estimates that fulfil (A.1) are computed with the Lagrange function

$$\mathbf{L}_{t,\text{adj}}^{[1]} = \mathbf{L}_t^{[1]} - \mathbf{V}_t^{[1]}\mathbf{\Delta}^{[1]T}\left(\mathbf{\Delta}^{[1]}\mathbf{V}_t^{[1]}\mathbf{\Delta}^{[1]T}\right)^{-1}\mathbf{\Delta}^{[1]}\mathbf{L}_t^{[1]} \tag{A.3}$$

with $\mathbf{L}_{t,\text{adj}}^{[1]} = \left(L_{t,1,1,\text{adj}}, L_{t,2,1,\text{adj}}, L_{t,3,1,\text{adj}}\right)^T$ the adjusted filtered trends. In the ideal case $\mathbf{V}_t^{[1]}$ is the variance-covariance matrix of the trend estimates $\mathbf{L}_t^{[1]}$. The covariances of the model estimates, however, are not known. Therefore the diagonal matrix of the variances is used instead.

In the second step, $\mathbf{L}_{t,\text{adj}}^{[1]}$ is not changed anymore. Now the vector of domain estimates $\mathbf{L}_t^{[2]} = \left(L_{t,1,2}, L_{t,1,3}, \ldots, L_{t,1,7}, L_{t,2,2}, \ldots, L_{t,2,7}, L_{t,3,2}, \ldots, L_{t,3,7}\right)^T$ is adjusted according to equation (A.1) for $m = 2, \ldots, 7$ and to equation (A.2) for $l = 1, 2$. Equation (A.2) for $l = 3$ is redundant and therefore left out. Again, the consistency requirements for the filtered trends of the domains are written as $\mathbf{\Delta}^{[2]}\mathbf{L}_t^{[2]} = \mathbf{C}_t^{[2]}$, with

$$\mathbf{\Delta}^{[2]} = \begin{pmatrix} \mathbf{I}_6 & \mathbf{I}_6 & -\mathbf{I}_6 \\ \mathbf{1}_6 & \mathbf{0}_6 & \mathbf{0}_6 \\ \mathbf{0}_6 & \mathbf{1}_6 & \mathbf{0}_6 \end{pmatrix},$$

$\mathbf{C}_t^{[2]} = \left(\mathbf{0}_6, L_{t,1,1,\text{adj}}, L_{t,2,1,\text{adj}}\right)^T$, $\mathbf{I}_6$ the six dimensional identity matrix, and $\mathbf{1}_6$ and $\mathbf{0}_6$ six dimensional row vectors with each element equal to one or zero respectively. Consistent domain estimates are computed with the Lagrange function

$$\mathbf{L}_{t,\text{adj}}^{[2]} = \mathbf{L}_t^{[2]} - \mathbf{V}_t^{[2]}\mathbf{\Delta}^{[2]T}\left(\mathbf{\Delta}^{[2]}, \mathbf{V}_t^{[2]}, \mathbf{\Delta}^{[2]T}\right)^{-1}\left(\mathbf{\Delta}^{[2]}\mathbf{L}_t^{[2]} - \mathbf{C}_t^{[2]}\right),$$

similarly to (A.3). In this case $\mathbf{V}_t^{[2]}$ is the diagonal matrix of the variances of the estimates of $\mathbf{L}_t^{[2]}$.

References

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

Binder, D.A., and Dick, J.P. (1990). A method for the analysis of seasonal ARIMA models. *Survey Methodology*, 16, 2, 239-253.

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

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

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

Isaki, C.T., and Fuller, W.A. (1982). Survey design under the regression superpopulation model. *Journal of the American Statistical Association*, 77, 89-96.

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

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

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

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 & Economic Statistics*, 16, 339-348.

Rao, J.N.K. (2003). *Small Area Estimation*. New York: John Wiley & Sons, Inc.

Robinson, P.M., and Särndal, C.-E. (1983). Asymptotic properties of the generalized regression estimator in probability sampling. *Sankhyā B*, 45, 240-248.

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

U.S. Census Bureau (2009). X-12-ARIMA Reference Manual. Washington DC.

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

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