Intervention analysis with state-space models to estimate discontinuities due to a survey redesign

Jan van den Brakel and Joeri Roels

Discussion paper (09039)

Statistics Netherlands

The Hague/Heerlen, 2009

Explanation of symbols

.	= data not available
*	= provisional figure
x	= publication prohibited (confidential figure)
–	= nil or less than half of unit concerned
–	= (between two figures) inclusive
0 (0,0)	= less than half of unit concerned
blank	= not applicable
2007–2008	= 2007 to 2008 inclusive
2007/2008	= average of 2007 up to and including 2008
2007/'08	= crop year, financial year, school year etc. beginning in 2007 and ending in 2008
2005/'06–2007/'08	= crop year, financial year, etc. 2005/'06 to 2007/'08 inclusive

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

Publisher
Statistics Netherlands
Henri Faasdreef 312
2492 JP The Hague

Prepress
Statistics Netherlands - Grafimedia

Cover
TelDesign, Rotterdam

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

Where to order
E-mail: verkoop@cbs.nl
Telefax +31 45 570 62 68

Internet
www.cbs.nl

ISSN: 1572-0314

6008309039 X-10

Intervention analysis with state-space models to estimate discontinuities due to a survey redesign

Jan van den Brakel and Joeri Roels

Summary: An important quality aspect of official statistics produced by national statistical institutes is comparability over time. To maintain uninterrupted time series, surveys conducted by national statistical institutes are often kept unchanged as long as possible. To improve the quality or efficiency of a survey process, however, it remains inevitable to adjust methods or redesign this process from time to time. Adjustments in the survey process generally affect survey characteristics such as response bias and therefore have a systematic effect on the parameter estimates of a sample survey. Therefore it is important that the effects of a survey redesign on the estimated series are explained and quantified. In this paper a structural time series model is applied to estimate discontinuities in series of the Dutch survey on social participation and environmental consciousness due to a redesign of the underlying survey process.

Keywords: *intervention analysis, response bias, structural time series models, survey sampling*

1. Introduction

Surveys conducted by national statistical institutes are generally conducted continuously or repeatedly in time with the purpose to produce consistent series. Quality of official statistics is based on various dimensions; see Brackstone (1999) for a discussion. One important quality aspect is comparability over time. To produce consistent series, national statistical institutes generally keep their survey processes unchanged as long as possible. It remains inevitable, however, to redesign survey processes from time to time to improve the quality or the efficiency of the underlying survey process. In an ideal survey transition process, the systematic effects of the redesign are explained and quantified in order to keep series consistent and preserve comparability of the outcomes over time. There are various possibilities to quantify the effect of a survey redesign, see Van den Brakel, Smith and Compton (2008) for an overview. If the redesign affects the data collection phase, then a parallel run is a reliable approach to avoid the confounding of real changes in the underlying phenomenon of interest with the systematic effect of the redesign. Therefore the redesign of long-standing surveys like e.g. the US Current Population Survey and the US National Crime Victimization Survey are accompanied with a parallel run, Dippo et al. (1994) and Kindermann and Lynch (1997).

Significance and power constraints necessary to establish the prespecified treatment effects, generally require large sample sizes for both the regular and the new survey in the parallel run. This is not always tenable due to budget constraints. The National Health Interview Survey (NHIS), established in 1956, is another example of a long standing survey. This survey is radically redesigned in 1997, Fowler (1996). The

absence of a parallel run obstructed the analysis of trends in different key variables of the NHIS. Akinbami et al. (2002, 2003) reported that trends in estimates of childhood asthma prevalence are disrupted due to changes in the NHIS design in 1997, which created the impression that childhood asthma prevalence declined in this period. Caban et al. (2005) used NHIS data to study trends in prevalence rates of obesity among working adults. Data were analyzed separately for NHIS periods 1986 until 1995 and 1997 until 2002 because of the major redesign of the NHIS in 1997. These examples illustrate that in situations were no parallel run is available alternative methods, which are based on explicit statistical models, should be considered to quantify the effect of a redesign. In this paper an intervention analysis using structural time series models is proposed as an alternative for conducting large scale field experiments and applied to a real life example at Statistics Netherlands. This is a direct application of the intervention approach proposed by Harvey and Durbin (1986) to estimate the effect of seat belt legislation on British road casualties.

In survey methodology, time series models are frequently applied to develop estimates for periodic surveys. Blight and Scott (1973) and Scott and Smith (1974) proposed to regard the unknown population parameters as a realization of a stochastic process that can be described with a time series model. This introduces relationships between the estimated population parameters at different time points in the case of non-overlapping as well as overlapping samples. The explicit modelling of this relationship between these survey estimates with a time series model can be used to combine sample information observed in the past to improve the precision of estimates obtained with periodic surveys. This approach is frequently applied in the context of small area estimation. Some key references to authors that applied the time series approach to repeated survey data to improve the efficiency of survey estimates are Scott et al. (1977), Tam (1987), Binder and Dick (1989, 1990), Bell and Hillmer (1990), Tiller (1992), Rao and Yu (1994), Pfeffermann and Burck (1990), Pfeffermann (1991), Pfeffermann and Bleuer (1993), Pfeffermann et al. (1998), Pfeffermann and Tiller (2006), Harvey and Chung (2000), Feder (2001), and Lind (2005).

In 1997 Statistics Netherlands started the Permanent Survey on Living Conditions (PSLC). This is a module-based integrated survey combining various themes concerning living conditions and quality of life. Two modules of the PSLC, the Module Justice and Environment and the Module Justice and Participation, are used to publish figures about justice and crime victimisation. The first module is also used to publish figures about environmental consciousness. The second module is used additionally to publish information about social participation. To realize expenditure cuts, the PSLC stopped at the end of 2004. From that moment on, figures about social participation and environmental consciousness are based on a separate survey, called the Dutch Survey on Social Participation and Environmental Consciousness (SSPEC).

In this survey transition the data collection mode, the questionnaire, the context of the survey and the fieldwork period changed, which resulted in systematic effects in the outcomes of the survey. Since the redesign mainly affects the data collection process in this application, a large scale field experiment is very appropriate to test the effect on the parameter estimates of the survey, see e.g. Van den Brakel (2008). An experimental approach might, however, be hampered due to budget and other practical constraints, which was the case for the Dutch SSPEC. Therefore, an

intervention analysis using a structural time series model is used as an alternative to quantify the effect of the redesign on the main series of the sample survey.

All target variables of the PSLC and the SSPEC have multinomial responses which are transformed to proportions of units classified in $K \geq 2$ categories. The survey estimates of these proportions are observed on a *K*-1 dimensional simplex and comprise a composition. Aitchison (1986) developed statistical methods for the analysis of compositional data, using additive logratio and central logratio transformations. Brunsdon and Smith (1998) developed VARMA models for logratio transformed compositional time series. Silva and Smith (2001) applied the structural time series modelling approach to logratio transformed compositional time series. In this paper the intervention approach proposed by Harvey and Durbin (1986) is applied to estimate the effect of a survey redesign on compositional time series obtained with periodic surveys.

In section 2 the PSLC and the SSPEC are described. The systematic effects due to the redesign are discussed in section 3. A time series model to quantify these discontinuities is developed in section 4. Results for the most important indicators for four different models are given in section 5. The performance of these models are investigated in a simulation study, which is also described in section 5. The paper concludes with a discussion in section 6.

2. Survey designs

2.1 Permanent Survey on Living Conditions

The PSLC was conducted as a repeatedly cross sectional survey, which implies that there is no sample overlap in time. The Module Justice and Environment and the Module Justice and Participation of the PSLC use persons aged 15 years or older as the target population. The PSLC was a continuously conducted survey. Each month a self-weighted stratified two-stage sample of persons was drawn from a sample frame derived from the municipal basic registration of population data. Strata are formed by geographical regions. Municipalities are considered as primary sampling units and persons as secondary sampling units. The monthly sample size averaged between 550 and 700 persons for both modules. With response rates varying around a level of 60%, this resulted in a yearly net response of about 4000 to 5000 persons for both modules.

Interviewers visited all the sampled persons at home and administered the questionnaire in a face-to-face interview. This is generally referred to as computer assisted personal interviewing (CAPI). The estimation procedure used to compile official statistics is based on the generalised regression estimator (Särndal et al. 1992, chapter 6) using a weighting scheme that is based on different sociodemographic categorical variables.

2.2 Survey on Social Participation and Environmental Consciousness

The PSLC stopped at the end of 2004. From that moment figures about social participation and environmental consciousness are based on the SSPEC. This survey

is also conducted as a repeatedly cross sectional survey and is based on a self-weighted stratified two-stage sample design of persons aged 15 years and older residing in the Netherlands. Data are collected by computer assisted telephone interviewing (CATI). As a result the subpopulation aged 15 years and older with an unlisted telephone number or cell-phone number is not observed. The data collection of the SSPEC is conducted in the months September, October and November with a monthly sample size of about 2500 persons. The estimation procedure is, like the PSLC, based on the generalised regression estimator. The response rates in the SSPEC varied around 65%. As a result about 4500 respondents are observed in the yearly samples.

Since 2005, figures about justice and crime victimisation are based on the Dutch Security Monitor. See Van den Brakel, Smith and Compton (2008) for more details about this redesign and the effects on the main series of this survey.

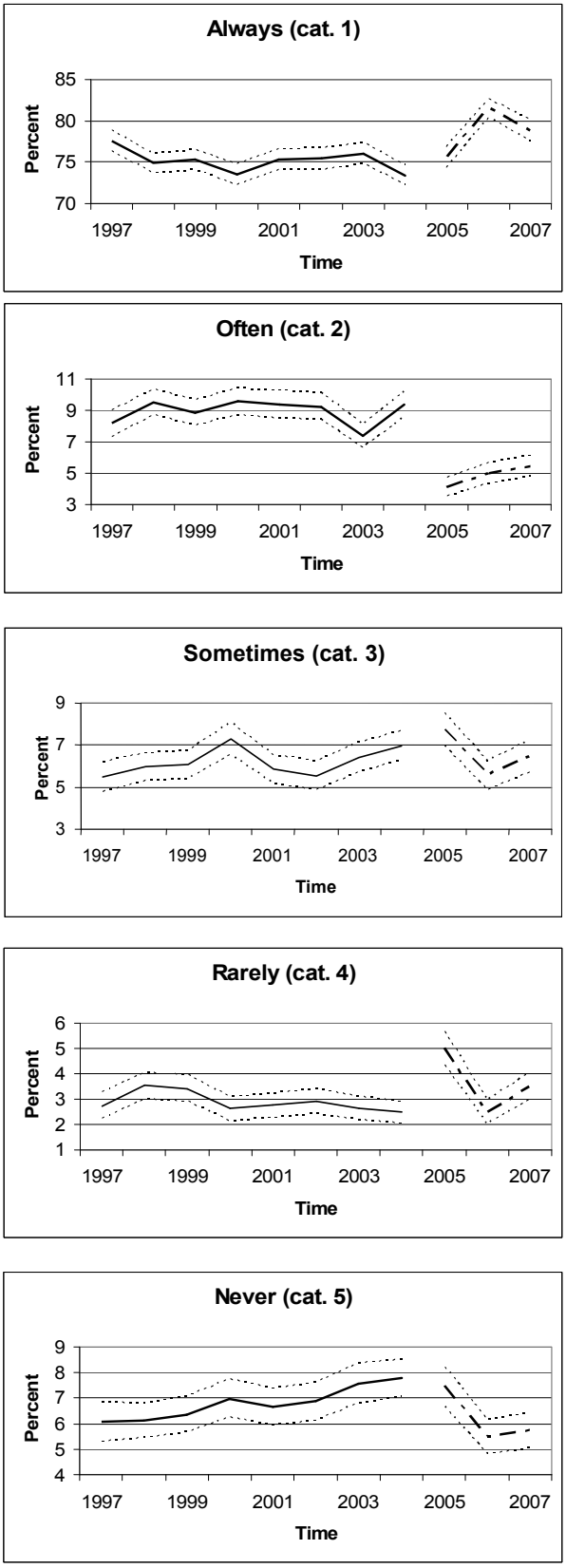
2.3 Target parameters

All target variables about environmental consciousness and social participation are based on closed questions where the respondent can choose one out of K answer categories to specify his opinion or behaviour on an ordinal scale. The target parameters are the estimated proportions that specify the distribution over these K categories for the entire population or subpopulations. In this paper the series of two variables are used for illustrative purposes. The first variable, Separating chemical waste, is an example of environmental consciousness. This variable contains five answer categories: 1) always, 2) often, 3) sometimes, 4) rarely, and 5) never. The second variable, Contact frequency with neighbours, is an example of social participation. This variable contains four answer categories: 1) at least once a week, 2) once within two weeks, 3) less than once within two weeks, 4) never.

3. Factors responsible for discontinuities

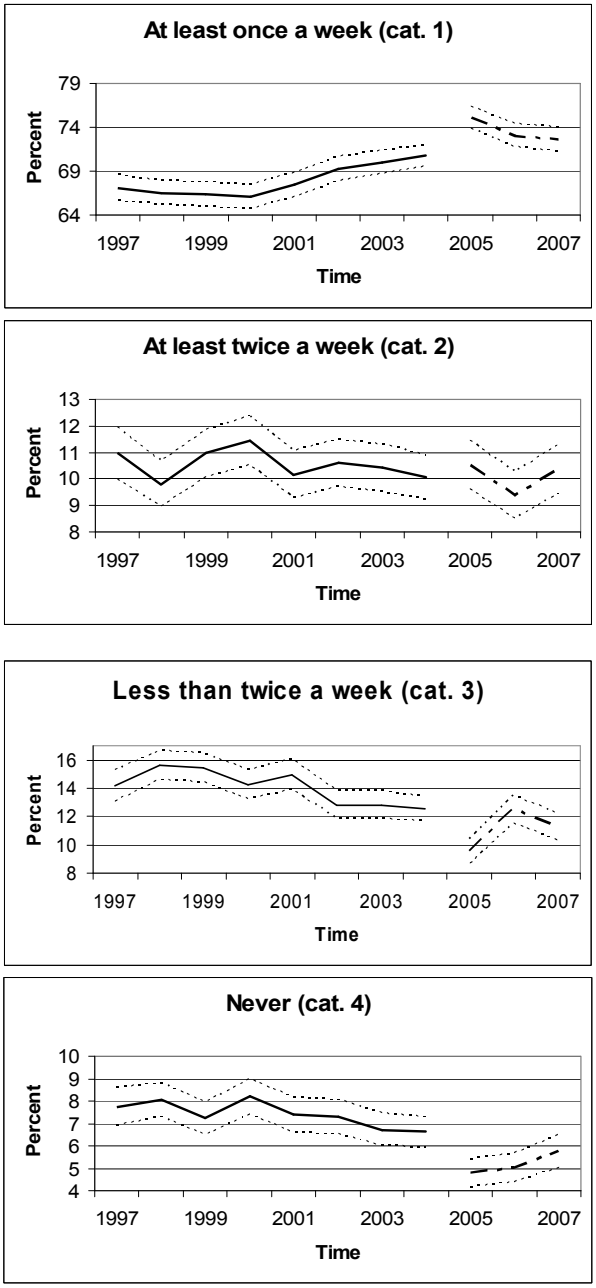
The redesign from the PSLC to the SSPEC resulted in discontinuities in most of the parameters about social participation and environmental consciousness. As an example the series with the annual figures of the parameters “Separating chemical waste” and “Contact frequency with neighbours” are shown in Figures 1 and 2 respectively. For both variables it appears that there are significant discontinuities in two or more of the underlying categories. The observed differences between the last year of the PSLC in 2004 and the first year of the SSPEC in 2005 are summarized in Table 1. The observed differences between the year before and the year after the change-over for other variables about environmental consciousness and social participation are described in Van den Brakel and Roels (2008).

Figure 1: Separating chemical waste



Solid line: observed series under the PSLC, dashed line: observed series under the SSPEC, dotted line: 95% confidence interval.

Figure 2: Contact frequency with neighbours



Solid line: observed series under the PSLC, dashed line: observed series under the SSPEC, dotted line: 95% confidence interval.

Table 1: Observed differences between the year before and the year after the change-over for "Separating chemical waste" and "Contact frequency with neighbours"

Variable	Category									
	1		2		3		4		5	
Freq. cont. neighb.	4.38**	(0.90)	0.46	(0.62)	-2.99**	(0.63	-1.84**	(0.47)		
Sep. chemical waste	2.26**	(0.89)	-5.25**	(0.50)	0.79	(0.53)	2.54**	(0.39)	-0.33	(0.54)

*Standard errors in brackets; *: p-value <0.05; **:p-value<0.01.*

The observed differences are the results of the factors that changed simultaneously in the survey redesign, real developments of the parameter and sampling errors. The most important factors that changed in the survey redesign are:

- Differences between sampled target populations. The SSPEC is based on a sample of persons aged 15 years and older with a listed telephone number or cell-phone number. The PSLC is based on a sample of persons aged 15 years and older. The SSPEC does not observe the subpopulation without a listed telephone number or cell-phone number. Additional analyses showed that this results in an under-representation of young people and ethnic minorities. This explains a substantial part of the discontinuities.
- Differences in data collection modes. The SSPEC is a telephone based survey, while in the PSLC data are collected in face-to-face interviews conducted at the respondents' homes. Many references in the literature emphasize that different collection modes have systematic effects on the responses, see for example De Leeuw (2005) and Dillman and Christian (2005). These so-called model effects arise for different reasons. Generally the interview speed in a face-to-face interview is lower compared to an interview conducted by telephone. Furthermore, respondents are more engaged with the interview and are more likely to exert the required cognitive effort to answer questions carefully in a face-to-face interview. Also fewer socially desirable answers are obtained under the CAPI mode due to the personal contact with the interviewer. As a result, fewer measurement errors are expected under the CAPI mode (Holbrook et al., 2003, and Roberts, 2007).
- Differences between data collection periods. The data collection for the SSPEC is conducted in September through November, while the PSLC is conducted continuously throughout the year. In the series of the quarterly figures observed under the PSLC, seasonal effects are observed in several parameters, which partially explain the discontinuities.
- Differences between questionnaire designs. Under the PSLC, questions about social participation and environmental consciousness were combined with questions about justice and crime victimisation in two different modules. Under the SSPEC, the questions about social participation and environmental consciousness are delineated in a new survey, which might have systematic effects on the outcomes of these surveys (Kalton and Schuman, 1982 and Dillman and Christian, 2005).
- Differences between the contexts of the surveys. The SSPEC is introduced as a survey that is focussed on topics about social participation and environmental consciousness. The PSLC is introduced as a more general survey on living conditions. Subsequently the survey focuses on topics about justice, crime victimisation, social participation or environmental consciousness. This might have a systematic selection effect on the respondents who decide to participate in the survey. Furthermore, in the SSPEC the attention of the respondent is completely focussed on one topic, contrary to the PSLC, which also may have systematic effects on the answer patterns of the respondents.

It is not immediately clear to what extent the differences summarized in Table 1 are the result of a real change in the underlying phenomenon of interest or are induced by the redesign of the survey. Even if no significant difference is observed, it is still possible that a real development could be nullified by an opposite redesign effect.

A general way to avoid confounding the autonomous development with redesign effects is to conduct an experiment embedded in the ongoing survey. If the effect of the separate factors that has changed in the survey process should be quantified, then a factorial design should be considered. Factorial designs or fractional factorial designs are generally hard to combine with the fieldwork restrictions encountered in the daily practice of survey sampling. Therefore it is generally necessary to combine the factors that changed in the redesign of the survey into one treatment and test the total effect of all factors that changed simultaneously in the redesign against the regular approach in a two-treatment experiment. See Van den Brakel (2008) and Van den Brakel, Smith and Compton (2008) for a detailed discussion and alternative approaches to quantify the effect of a survey redesign.

Since an experimental approach is not applied in this application, a time series model is developed in the next section to quantify the total effect of all factors that are modified in the survey redesign with the purpose to avoid confounding with real developments of the respective parameter. Some insight into the effect for some of the factors that have changed in the survey redesign can be obtained by conducting additional calculation on the existing data. The selection effect of surveying the subpopulation that can be contacted by telephone can be estimated with standard sampling theory for domain estimators from the data obtained with the PSLC since this survey approaches the entire population face-to-face. The effect of changing the period of data collection can also be quantified by making, for example, quarterly series for the PSLC and estimating the seasonal pattern. Due to the relatively small sample sizes and the limited length of the series, it turned out to be hard to establish significant seasonal effects.

4. Structural times series models

In this section structural time series models are developed to estimate the discontinuities in the series of a survey due to the redesign of the underlying survey process. With a structural time series model, a series is decomposed in a trend component, seasonal component, other cyclic components, regression component and an irregular component. For each component a stochastic model is assumed. This allows the trend, seasonal, and cyclic component but also the regression coefficients to be time dependent. If necessary ARMA components can be added to capture the autocorrelation in the series beyond these structural components. See Harvey (1989) or Durbin and Koopman (2001) for details about structural time series modelling.

4.1 Intervention analysis for time series obtained with periodic surveys

The variables of the PSLC and the SSPEC are defined as categorical variables measured on an ordinal scale and the population values of interest are the distributions in the population over the K categories of these variables. For each variable a K-dimensional vector $\mathbf{y}_t = (y_{t,1}, \ldots y_{t,K})$ is defined where the elements of $\mathbf{y}_t$ specify the proportions over the K categories. Based on the data observed under the PSLC and the SSPEC, direct estimates for the unknown population values are obtained with the generalized regression estimator. As a result, for each variable K series are observed that specify the estimated proportions over K categories and are collected in the K-dimensional vector $\hat{\mathbf{y}}_t = (\hat{y}_{t,1}, \ldots \hat{y}_{t,K})$, $t = 1, \ldots, T$.

Developing a time series model for survey estimates observed with a periodic survey starts with a model, which states that the survey estimate can be decomposed in the value of the population variable and a sampling error: $\hat{y}_{t,k} = y_{t,k} + e_{t,k}$, with $e_{t,k}$ the sampling error. Scott and Smith (1974) proposed to consider the true population value $y_{t,k}$ as the realisation of a stochastic process that can be properly described with a time series model. This approach is applied to the series observed with the PSLC and the SSPEC using the framework of structural time series modelling.

In classical sampling theory, it is generally assumed that the observations obtained in the sample are true fixed values observed without error, see e.g. Cochran (1977). This assumption is not tenable if systematic differences are expected due to a redesign of the survey process. Van den Brakel and Renssen (2005) proposed a measurement error model for experiments embedded in sample surveys that links systematic differences between a finite population variable observed under different survey implementations. They consider the observed population value obtained under a complete enumeration under two or more different implementations of the survey process as the sum of a true intrinsic value that is biased with a systematic effect induced by the survey design, i.e. $y_{t,k,l} = u_{t,k} + b_{k,l}$. Here $y_{t,k,l}$ is the population value of the *k*-th parameter at time *t* observed under the *l*-th survey approach, $u_{t,k}$ the true population value of this parameter and $b_{k,l}$ the measurement bias induced by the *l*-th survey process used to measure $u_{t,k}$. The systematic difference between two survey approaches is obtained by the contrast $y_{t,k,l} - y_{t,k,l'} = b_{k,l} - b_{k,l'} \equiv \beta_k$. In the case of embedded experiments, the systematic difference between two or more survey approaches is estimated as the contrast between estimates obtained from subsamples assigned to the different survey approaches. In the time series approach, these differences are estimated using an appropriate intervention variable. This allows for time dependent differences. For notational convenience, the subscript *l* will be omitted in $y_{t,k,l}$, since the survey approach will be indicated with the time period.

In the case of the PSLC and the SSPEC a relatively short series for annual data is considered. Therefore, the autonomous development of the indicator that is described by the series is modelled with a stochastic trend, a regression component and an irregular component. The regression component consists of an intervention variable with a time independent regression coefficient that describes the effect of the survey transition. This approach is initially proposed by Harvey and Durbin (1986). Seasonal, cyclic, ARMA, and other auxiliary regression components can be included in the model for example in the case of longer series or monthly or quarterly data.

Based on the preceding considerations, the univariate structural time series model for the *k*-th component of $\hat{\mathbf{y}}_t$ is defined as:

$$\hat{y}_{t,k} = L_{t,k} + \beta_k \delta_t + \nu_{t,k} + e_{t,k}, \qquad (1)$$

with $L_{t,k}$ a stochastic trend, δ_t an intervention variable that describes under which survey the observations are obtained at period *t*, β_k the time independent regression coefficient for the intervention variable, $\nu_{t,k}$ an irregular component for the time series model of the population values $y_{t,k}$ and $e_{t,k}$ the sampling error. It is assumed that the irregular component is normally and independently distributed: $\nu_{t,k} \cong N(0, \sigma_\nu^2)$.

Surveys are often based on a rotating panel design. Such designs result in partially overlapping samples with correlated sampling errors. Particularly in these cases a

separate component for the sampling error in the time series model might be required to capture this serial correlation. Through this component the estimated variances for the $\hat{y}_{t,k}$, which are generally available from the survey, can be included in the time series model as prior information. Binder and Dick (1990) proposed the following general form for the sampling error model to allow for non homogeneous variance in the sampling errors:

$$e_{t,k} = \omega_{t,k}\widetilde{e}_{t,k} \ , \tag{2}$$

where $\omega_{t,k}$ is the standard error of $\hat{y}_{t,k}$ and $\widetilde{e}_{t,k}$ an ARMA process that models the serial correlation between the sampling errors. Abraham and Vijayan (1992), and Harvey and Chung (2000) applied MA models for the serial correlation in the sampling errors. Pfeffermann (1991), Pfeffermann et al. (1998) and Van den Brakel and Krieg (2009) used AR models for the serial correlation in the sampling errors. Autocorrelations can be estimated from the survey data and can be used, like the design variances of $\hat{y}_{t,k}$, as prior information in the sampling error model. Pfeffermann et al. (1998) developed a procedure to estimate the autocorrelation in the survey errors from the separate panel estimates of a rotating panel design and used this prior information to estimate the autocorrelation coefficients of an AR model.

Generally there are systematic differences between the subsequent panels of a rotating panel design. In the literature, this phenomenon is known as rotation group bias (RGB), Bailar (1975). Pfeffermann (1991) applied a multivariate structural time series model to the series of the survey estimates of the separate panel waves that accounts for this RGB and applied an AR model for the autocorrelation of the sampling errors of the different panels. Variances and autocorrelations of the sampling errors are obtained by standard maximum likelihood estimation in this application. Van den Brakel and Krieg (2009) used a multivariate structural time series model similar to the model proposed by Pfeffermann (1991). They estimated the variances and autocorrelations of the sampling errors from the survey data and used this as prior information in the time series model.

The PSLC and the SSPEC are based on non-overlapping cross-sectional samples. The only difference between the sample designs is the yearly sample size. As a result there is no serial correlation between sampling errors and non-homogeneous variance is caused by differences in the yearly sample size. Based on these considerations it is decided to combine both terms $\nu_{t,k}$ and $\varepsilon_{t,k}$ in one irregular term, which is assumed to be normally and independently distributed with zero mean and a variance that is inversely proportional to the sample size:

$$\nu_{t,k} + e_{t,k} = \varepsilon_{t,k} \tag{3}$$

$$\varepsilon_{t,k} \cong N(0, \frac{\sigma^2_{\varepsilon,k}}{n_t}) \, .$$

Defining the variance of the irregular term inversely proportional to the sample size implies that it is implicitly assumed that the sampling error dominates the irregular term. It is also assumed that the irregular components of (3) at different time points are uncorrelated: $Cov(\varepsilon_{t,k}\varepsilon_{t',k}) = 0$ for $t \neq t'$. As a result model (1) simplifies to

$$\hat{y}_{t,k} = L_{t,k} + \beta_k \delta_t + \varepsilon_{t,k} \tag{4}$$

For the stochastic trend, the widely applied smooth trend model is assumed, see e.g. Durbin and Koopman, (2001):

$$\begin{aligned} L_{t,k} &= L_{t-1,k} + R_{t-1,k} \ , \\ R_{t,k} &= R_{t-1,k} + \eta_{t,R,k} \ , \end{aligned} \tag{5}$$

with $L_{t,k}$ the level component and $R_{t,k}$ the stochastic slope component of the trend, and $\eta_{t,R,k}$ an irregular component. The smooth trend model (5) is a special case of the local linear trend model, which also has an irregular term for $L_{t,k}$, see e.g. Durbin and Koopman, (2001) equation (3.2). The population values in this application do not change rapidly over time. Therefore a model that gives smooth trend estimates seems to be appropriate. The choice for (5) also results in a more parsimonious model, which is an additional advantage in this application where the length of the observed series is small. It is assumed that the irregular components of (5) are normally and independently distributed, i.e. $\eta_{t,R,k} \cong N(0,\sigma^2_{R,k})$, and that they are uncorrelated at different time points, i.e. $Cov(\eta_{t,R,k}\eta_{t',R,k}) = 0$ for $t \neq t'$. Furthermore, it is assumed that the irregular components of (4) and (5) are uncorrelated: $Cov(\varepsilon_{t,k}\eta_{t',R,k}) = 0$ for all t and t'.

The intervention variable models the effect of the survey redesign. Three types of interventions are discussed: a level shift, a slope intervention, and an intervention an a seasonal pattern. Let T_R denote the time period at which the survey process is redesigned. In the case of a level intervention it is assumed that the magnitude of the discontinuity due to the survey redesign is constant over time. In this case δ_t is defined as a dummy variable:

$$\delta_t = \begin{cases} 0 & \text{if} \quad t < T_R \\ 1 & \text{if} \quad t \geq T_R \end{cases} \tag{6}$$

In the case of a slope intervention it is assumed that the magnitude of the discontinuity increases over time. This is accomplished by defining δ_t as:

$$\delta_t = \begin{cases} 0 & \text{if} \quad t < T_R \\ 1 + t - T_R & \text{if} \quad t \geq T_R \end{cases} \tag{7}$$

It is also possible to define an intervention on the seasonal or cyclic pattern. Such interventions can be considered if an interaction is expected between the survey redesign and the months or the quarters of the year. In this case, a stochastic seasonal component is added to equation (1) or (4). Widely applied models are trigonometric models and the dummy variable seasonal model, see Durbin and Koopman (2001), section 3.2 for expressions. Furthermore the intervention variable δ_t has the form (6) and the regression coefficient β_k is replaced by a time independent seasonal component.

The interventions described so far assume that the redesign only affects the point estimates of the survey. A survey redesign could, however, also affect the variance of the measurement errors. An increase or decrease of the variance of the measurement errors will be reflected in the estimated variance of $\hat{y}_{t,k}$. A straightforward way to account for such effects is to incorporate the estimated variances of the survey estimates as prior information using sampling error model (2). Another possibility is to define separate model variances for the irregular term $\varepsilon_{t,k}$ in the measurement equation for the period before and after the implementation

of the survey redesign, i.e. $Var(\varepsilon_{t,k}) = \sigma^2_{\varepsilon,k,1}$ if $t < T_R$ and $Var(\varepsilon_{t,k}) = \sigma^2_{\varepsilon,k,2}$ if $t \geq T_R$. The ratio between $\sigma^2_{\varepsilon,k,1}$ and $\sigma^2_{\varepsilon,k,2}$ can be used to test hypotheses about the equivalence of both variance components. This approach, however, requires a sufficient number of observations under both surveys to test the equivalence of these variance components with sufficient power.

The discontinuity in the series is modelled with an intervention variable that describes the moment that the survey process is redesigned. This approach assumes that the other components of the time series model approximate the real development of the population variable reasonably well and that there is no structural change in e.g. the trend or the seasonal component at the moment that the new survey is implemented. If a change in the real development of the population variable exactly coincides with the implementation of the new survey, then the model will wrongly assign this effect to the intervention variable which is intended to describe the redesign effect. Information available from series of correlated variables can be used to evaluate the assumption that there is no structural change in the real evolution of the parameter. Such auxiliary series can also be added as a regression component to the model, with the purpose to reduce the risk that a structural change in the evolution of the series of the target parameter is wrongly assigned to the intervention variable. An auxiliary series can also be included as a dependent variable in a multivariate model, which accounts for the correlation between the parameters of the trend and seasonal components, Pfeffermann and Burck (1990), Pfeffermann and Bluer (1993), or allows for a common trend, Harvey and Chung (2000).

The risk that the intervention variable wrongfully absorbs a part of the development of the real population value can be reduced by applying parsimonious intervention parameters. Therefore, time dependent interventions, like an intervention on a seasonal component, must be applied carefully. These intervention parameters are more flexible and will easily absorb a part of the real evolution of the population value, particularly if only a limited number of observations after the survey change-over are available.

The intervention approach can be generalized in a straightforward way to situations were the survey process has been redesigned at two or more occasions. This is achieved by adding a separate intervention variable for each time that the survey process has been modified.

4.2 State space representation

The structural time series models developed in subsection 4.1 for the separate parameters $\hat{y}_{t,k}$ of the vector $\hat{\mathbf{y}}_t$ comprise a *K*-dimensional structural time series model. The general way to proceed is to put this model in state space representation and analyse the model with the Kalman filter. The state space representation for this *K*-dimensional structural time series model reads as:

$$\hat{\mathbf{y}}_t = \mathbf{Z}_t \boldsymbol{\alpha}_t + \boldsymbol{\varepsilon}_t \tag{8}$$

$$\boldsymbol{\alpha}_t = \mathbf{T}\boldsymbol{\alpha}_{t-1} + \boldsymbol{\eta}_t \tag{9}$$

The measurement equation (8) describes how the observed series depends on a vector of unobserved state variables $\boldsymbol{\alpha}_t$ and a vector with disturbances $\boldsymbol{\varepsilon}_t$. The state vector contains the level and slope components of the trend models and the regression coefficients of the intervention variables. The transition equation (9)

describes how these state variables evolve over time. The vector $\boldsymbol{\eta}_t$ contains the disturbances of the assumed first-order Markov processes of the state variables. The matrices in (8) and (9) are given by

$$\boldsymbol{\alpha}_t = (L_{t,1}, R_{t,1}, ..., L_{t,K}, R_{t,K}, \beta_1, ..., \beta_K)^T, \tag{10a}$$

$$\mathbf{Z}_t = \left(\mathbf{I}_{[K]} \otimes (1,0) \mid \delta_t \mathbf{I}_{[K]}\right), \tag{10b}$$

$$\mathbf{T} = Blockdiag(\mathbf{T}_{tr}, \mathbf{I}_{[K]}), \tag{10c}$$

$$\mathbf{T}_{tr} = \mathbf{I}_{[K]} \otimes \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \tag{10d}$$

with $\mathbf{0}_{[p]}$ a column vector of order p with each element equal to zero, and $\mathbf{I}_{[p]}$ the $p \times p$ identity matrix. The disturbance vectors are defined as

$$\boldsymbol{\varepsilon}_t = (\varepsilon_{t,1}, ..., \varepsilon_{t,K})^T,$$

$$\boldsymbol{\eta}_t = (0, \eta_{t,R,1}, ..., 0, \eta_{t,R,K}, \mathbf{0}_{[K]}^T)^T.$$

It is assumed that

$$E(\boldsymbol{\varepsilon}_t) = \mathbf{0}_{[K]},\ Cov(\boldsymbol{\varepsilon}_t) = \frac{1}{n_t} Diag(\sigma^2_{\varepsilon,1}, ..., \sigma^2_{\varepsilon,K}),$$

$$E(\boldsymbol{\eta}_t) = \mathbf{0}_{[3K]},\ Cov(\boldsymbol{\eta}_t) = Diag(0, \sigma^2_{R,1}, ..., 0, \sigma^2_{R,K}, \mathbf{0}^T_{[K]}).$$

In the case that each measurement equation and each transition equation has its own separate hyperparameter, then (10) is a set of K univariate structural time series models. If the measurement equations or the transition equations share common hyperparameters, then (10) is a K dimensional seemingly unrelated multivariate structural time series model. This is for example the case if $\sigma^2_{\varepsilon,1} = ... = \sigma^2_{\varepsilon,K} = \sigma^2_{\varepsilon}$.

The time independent regression coefficients of the intervention variables are also included in the state vector, 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. An alternative approach of estimating the regression coefficients is by augmentation of the Kalman filter, see Durbin and Koopman (2001), subsection 6.2.3 for details.

In this application, each variable specifies the proportions over K categories. In other words, each variable makes up a K-dimensional series, which obeys the restriction that at each point in time these series add up to one, i.e. $\sum_{k=1}^{K} \hat{y}_{t,k} = 1$, and $0 \le \hat{y}_{t,k} \le 1$. As a result, the K regression coefficients of the intervention variables must obey the restriction $\sum_{k=1}^{K} \beta_k = 0$. The multivariate structural time series model (10) can be augmented with this restriction by using the following design matrix in the transition equation (9):

$$\mathbf{T} = Blockdiag(\mathbf{T}_{tr}, \mathbf{T}_{iv}), \tag{10e}$$

where $\mathbf{T}_{tr}$ is defined by (10d), and

$$\mathbf{T}_{iv} = \begin{pmatrix} \mathbf{I}_{[K-1]} & \mathbf{0}_{[K-1]} \\ -\mathbf{1}^T_{[K-1]} & 0 \end{pmatrix}, \tag{10f}$$

with $\mathbf{1}_{[p]}$ a column vector of order p with each element equal to one. Due to $\mathbf{T}_{iv}$, defined in (10f), the regression coefficients as well as their Kalman filter estimates

obey the restriction $\sum_{k=1}^{K} \beta_k = 0$. In the case of a level intervention, the time series after the moment of the survey transition can be adjusted for the estimated discontinuities with $\tilde{y}_{t,k} = \hat{y}_{t,k} - \hat{\beta}_k$. As an alternative, the series before the survey transition can be adjusted with $\tilde{y}_{t,k} = \hat{y}_{t,k} + \hat{\beta}_k$. In the case of a slope intervention the time series is adjusted with $\tilde{y}_{t,k} = \hat{y}_{t,k} - \hat{\beta}_k \delta_t$. If the time series after the moment of the survey transition is adjusted, then δ_t is defined by (7). If the time series before the changeover is adjusted, then δ_t is defined as

$$\delta_t = \begin{cases} t - T_R & \text{if} \quad t < T_R \\ 0 & \text{if} \quad t \geq T_R \end{cases}. \tag{11}$$

Since the observed series and the estimated discontinuities obey the required consistencies, the adjusted series does too.

An intervention on a seasonal component can be implemented in a way similar to a level intervention. Let *s* denote the number of time periods of the seasonal set. The state vector $\boldsymbol{\alpha}_t$ is augmented with $K \times s$ state variables to model the seasonal pattern for each parameter $\hat{y}_{t,k}$. The *K* regression coefficients β_k are replaced by another set of $K \times s$ state variables to model the intervention on seasonal pattern for each target parameter. The design matrix of the measurement equation $\mathbf{Z}_t$ is augmented with a term $\mathbf{I}_{[K]} \otimes \mathbf{z}_{[s]}^T$, where $\mathbf{z}_{[s]}$ is an *s* dimensional vector that describes the relation between the observed series and the state variable of the trigonometric seasonal model or the dummy variable seasonal model. Furthermore $\delta_t \mathbf{I}_{[K]}$ in $\mathbf{Z}_t$ is replaced by $\delta_t \mathbf{I}_{[K]} \otimes \mathbf{z}_{[s]}^T$. The design matrix of the transition equation is augmented with a block diagonal element $\mathbf{I}_{[K]} \otimes \mathbf{T}_s$, where $\mathbf{T}_s$ denotes the transitional relation for a trigonometric model or the dummy variable seasonal model. See Durbin and Koopman (2001), subsection 3.2 for expressions of $\mathbf{z}_{[s]}$ and $\mathbf{T}_s$. To force that the sum over the seasonal intervention variables of the *K* parameters equals zero, the design matrix of the transition equation is augmented with $\mathbf{T}_{iv} \otimes \mathbf{T}_s$, where $\mathbf{T}_{iv}$ is defined by (10f). Adjusted series are obtained with the approach described for the level intervention.

4.3 Logratio transformations

The multivariate model developed for $\hat{\mathbf{y}}_t$ accounts for the restriction that $\sum_{k=1}^{K} \hat{y}_{t,k} = 1$, but ignores the restriction $0 \leq \hat{y}_{t,k} \leq 1$. Ignoring the second restriction might result in adjusted parameter estimates taking values outside the admissible range [0,1]. In fact each parameter defines a set of time series that are observed on the *K*-1-dimensional simplex. One way to account for both restrictions is to apply a logratio transformation to the original data:

$$\hat{x}_{t,k} = \ln\left(\frac{\hat{y}_{t,k}}{\hat{y}_{t,K}}\right), \; k=1, \ldots, K\text{-}1. \tag{12}$$

With (12) the original observations $\hat{\mathbf{y}}_t$ are transformed from the (*K*-1) dimensional simplex to the (*K*-1) dimensional real space, see Aitchison (1986) for details. State-space models are applied to logratio transformed compositional time series obtained from repeated surveys by Silva and Smith (2001). They also give the details how to account for serial correlation between the sampling errors in logratio transformed survey data in the case of partially overlapping surveys.

Instead of modelling the original series $\hat{\mathbf{y}}_t$ and explicitly benchmark the regression coefficients to restriction (10f), it is also possible to develop a set of *K*-1 univariate structural time series models or a set of *K*-1 seemingly unrelated structural time series for $\hat{\mathbf{x}}_t = (\hat{x}_{t,1},...,\hat{x}_{t,K-1})^t$.

This model is obtained with formulae (8) and (9) where $\hat{\mathbf{y}}_t$ is replaced by $\hat{\mathbf{x}}_t$, and taking

$$\begin{aligned}
&\boldsymbol{\alpha}_t = (L_{t,1}, R_{t,1},..., L_{t,K-1}, R_{t,K-1}, \beta_1,..., \beta_{K-1})^T, \\
&\mathbf{Z}_t = \left(\mathbf{I}_{[K-1]} \otimes (1,0) \mid \delta_t \mathbf{I}_{[K-1]}\right), \\
&\mathbf{T} = Blockdiag(\mathbf{T}_{tr}, \mathbf{T}_{iv}), \qquad \mathbf{T}_{tr} = \mathbf{I}_{[K-1]} \otimes \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \ \mathbf{T}_{iv} = \mathbf{I}_{[K-1]}, \qquad (13) \\
&\boldsymbol{\varepsilon}_t = (\varepsilon_{t,1},..., \varepsilon_{t,K-1})^T, \\
&\boldsymbol{\eta}_t = (0, \eta_{t,R,1},..., 0, \eta_{t,R,K-1}, \mathbf{0}^T_{[K-1]})^T.
\end{aligned}$$

The estimated discontinuities apply to the *K*-1 transformed series. In the case of level intervention, the series observed after the survey transition can be adjusted to the level of the series before the changeover using $\tilde{x}_{t,k} = \hat{x}_{t,k} - \hat{\beta}_k$. The series observed before the survey transition can be adjusted to level under the new situation with $\tilde{x}_{t,k} = \hat{x}_{t,k} + \hat{\beta}_k$. In the case of a slope intervention the time series is adjusted with $\tilde{x}_{t,k} = \hat{x}_{t,k} - \hat{\beta}_k \delta_t$. If the time series after the moment of the survey transition is adjusted, then δ_t defined by (7). If the time series before the changeover is adjusted, then δ_t is defined by (11). The state-space representation for a seasonal intervention follows in a straightforward way from subsection 4.2. Subsequently, the adjusted series can be transformed back to their original values that specify the proportions over *K* categories on the simplex by the inverse of (12), which is given by

$$\tilde{y}_{t,k} = \frac{\exp(\tilde{x}_{t,k})}{\sum_{k=1}^{K-1} \exp(\tilde{x}_{t,k}) + 1}, \ k=1, \ldots, K\text{-}1, \qquad (14)$$

$$\tilde{y}_{t,K} = \frac{1}{\sum_{k=1}^{K-1} \exp(\tilde{x}_{t,k}) + 1}.$$

The adjusted series meets the consistency property that the adjusted proportions add up to 1, and the values of the *K* categories take values in the range [0,1], since the logratio transformation accounts for the properties of the data observed on a simplex. Most important drawback of this approach is that the interpretation of the results is more difficult and the asymmetric treatment of the classes in the logratio transformation (12). Aitchison (1986) shows that analysis results obtained with logratio transformed compositional data are invariant for the choice of the reference category that is used as the denominator. This result is generalized to VARMA models applied to logratio transformed compositional time series by Brunsdon and Smith (1998) and state-space models by Silva and Smith (2001). The outcomes for the adjusted series, nevertheless, depend on the choice of the category that is used in the denominator of the logratio transformation, and can be attributed to the numerical optimization procedure used for maximum likelihood estimation (see subsection 4.5).

The asymmetric treatment of the K classes in logratio transformation (12) can be avoided by replacing the reference category $\hat{y}_{t,K}$ in the denominator by the geometric mean over the K categories. This results in the so called central logratio transformation, which is defined by

$$\hat{z}_{t,k} = \ln\left(\frac{\hat{y}_{t,k}}{g(\hat{y}_t)}\right), \quad k=1,\ldots, K, \tag{15}$$

with

$$g(\hat{y}_t) = \left(\prod_{k=1}^{K} \hat{y}_{t,k}\right)^{\frac{1}{K}}. \tag{16}$$

The advantage of this transformation is that the results do not depend on the choice of a reference category. With (15), however, the vector $\hat{\mathbf{y}}_t$ is transformed from the K-1 dimensional simplex to a linear subspace of the K dimensional real space that is confined by $\sum_{k=1}^{K} \hat{z}_{t,k} = 0$.

The central logratio transformed series can be modelled with a K dimensional structural time series model. Since the K regression coefficients of the intervention variables must still obey the restriction $\sum_{k=1}^{K} \beta_k = 0$, time series model (8), (9), (10a), through (10f), can be applied to model the series obtained after the central logratio transformation. The series can be adjusted for the estimated discontinuities in a similar way as described for the untransformed and logratio transformed series. Subsequently the adjusted series can be transformed back to their original values by the inverse of (15):

$$\tilde{y}_{t,k} = \frac{\exp(\tilde{z}_{t,k})}{\sum_{k=1}^{K} \exp(\tilde{z}_{t,k})}, \quad k=1,\ldots, K. \tag{17}$$

4.4 Benchmarking with series for subpopulations

In sample surveys, parameter estimates for the total population are often also itemized in different subpopulations or domains. The following relationship applies between the series at the national level and its breakdown in H subpopulations

$$\hat{y}_t = \sum_{h=1}^{H} \frac{N_h}{N} \hat{y}_t^h. \tag{18}$$

Here $\hat{y}_t^h$ and N_h denote the parameter estimate and the size of subpopulation h respectively, and $N = \sum_{h=1}^{H} N_h$ the size of the total population. Applying the time series models, described in sections 4.1, 4.2 and 4.3, separately to the series at the national level and its breakdown for these H subpopulations might result in inconsistencies between these series after adjustment for the discontinuities. These inconsistencies arise since the regression coefficients for the intervention variables do not account for the consistency requirement specified by (18).

One solution is to benchmark the adjusted series for the subpopulations to the adjusted series at the national level, for example by using the method of Lagrange

multipliers. Let $\tilde{\mathbf{y}}_t = (\tilde{\mathbf{y}}_{t,tot}^T, \tilde{\mathbf{y}}_{t,1}^T ..., \tilde{\mathbf{y}}_{t,H}^T)^T$ denote a $(H+1)K$ -vector containing the adjusted parameter estimates for period t for the total population $\tilde{\mathbf{y}}_{t,tot} = (\tilde{y}_{t,tot,1}, ..., \tilde{y}_{t,tot,K})^T$ and the H subpopulations $\tilde{\mathbf{y}}_{t,h} = (\tilde{y}_{t,h,1}, ..., \tilde{y}_{t,h,K})^T$. These parameters must obey a set of linear restrictions such that (18) is met and the unit sum constraint for the vectors $\tilde{\mathbf{y}}_{t,tot}$ and $\tilde{\mathbf{y}}_{t,h}$ for $h = 1, \ldots, H$, still applies. This gives rise to a set of $(H + K)$ linear restrictions that can be expressed as

$$\mathbf{R}\tilde{\mathbf{y}}_t^* = \mathbf{c}, \tag{19}$$

with

$$\mathbf{R} = \begin{pmatrix} (1, -\mathbf{f}_{[H]}^T) \otimes \mathbf{L} \\ \mathbf{I}_{[H+1]} \otimes \mathbf{1}_{[K]}^T \end{pmatrix}, \ \mathbf{L} = \begin{pmatrix} \mathbf{I}_{[K-1]} & \mathbf{0}_{[K-1]} \end{pmatrix}, \ \mathbf{f} = \left(\frac{N_1}{N}, ..., \frac{N_H}{N} \right)^T,$$

and $\mathbf{c} = \left(\mathbf{0}_{[K-1]}^T, \mathbf{1}_{[H+1]}^T \right)^T$.

Applying the method of Lagrange multipliers gives

$$\tilde{\mathbf{y}}_t^* = \tilde{\mathbf{y}}_t + \mathbf{V}\mathbf{R}^T(\mathbf{R}\mathbf{V}\mathbf{R}^T)^{-1}[\mathbf{c} - \mathbf{R}\tilde{\mathbf{y}}_t], \tag{20}$$

where $\mathbf{V}$ denotes the covariance matrix of $\tilde{\mathbf{y}}_t$. In (20) the discrepancies $[\mathbf{c} - \mathbf{R}\tilde{\mathbf{y}}_t]$ are distributed over the values of $\tilde{\mathbf{y}}_t$ proportional to their accuracy measure specified by $\mathbf{V}$. This implies that the parameters for the total population receive smaller adjustments than the parameters for the subpopulations, since parameters for the total population are estimated more precisely compared to domain estimates. The covariance matrix of (20) is given by

$$\mathbf{V}(\tilde{\mathbf{y}}_t^*) = \mathbf{V} - \mathbf{V}\mathbf{R}^T(\mathbf{R}\mathbf{V}\mathbf{R}^T)^{-1}\mathbf{R}\mathbf{V}.$$

The benchmarked estimates obtained with (20) have smaller or equal variances than the separately adjusted series. The interpretation of this variance reduction is that the restrictions specified by (19) add additional information to the model that is applied to adjust the series for the observed discontinuities.

Inconsistencies can also be avoided by modelling the untransformed series for the total population and its breakdown in the H subpopulations, i.e. $\tilde{\mathbf{y}}_t = (\tilde{\mathbf{y}}_{t,tot}^T, \tilde{\mathbf{y}}_{t,1}^T, ..., \tilde{\mathbf{y}}_{t,H}^T)^T$, simultaneously in one multivariate model and include the consistency requirements in the transition equation for the regression coefficient of the intervention variables. To avoid unnecessary mathematical notation, the transition equation is only given for the regression coefficients of these intervention variables. The formulation of the complete state space representation follows directly from the models defined in section 4.1.

Let $\boldsymbol{\beta} = \mathbf{T}\boldsymbol{\beta}$ denote the transition equation for the time invariant regression coefficients of the intervention variables for the series of the total population and the H subpopulations, i.e. $\boldsymbol{\beta} = \left(\boldsymbol{\beta}_{Tot}^T, \boldsymbol{\beta}_1^T, ..., \boldsymbol{\beta}_H^T \right)^T$, with $\boldsymbol{\beta}_{Tot}$ the K dimensional vector containing the intervention variables for the K categories of the parameter for the total population and $\boldsymbol{\beta}_h$ the K dimensional vector containing the intervention variables of the parameter for the subpopulations. If the transition matrix is defined as

$$\mathbf{T} = \begin{pmatrix} \mathbf{O}_{[K\times K]} & \mathbf{f}_{[H]}^{T} \otimes \mathbf{T}_{iv} \\ \mathbf{1}_{[H]} \otimes \mathbf{O}_{[K\times K]} & \mathbf{I}_{[H]} \otimes \mathbf{T}_{iv} \end{pmatrix},$$

where $\mathbf{T}_{iv}$ is defined by (10.f), then it follows that the adjusted series meet the consistencies specified by (18) as well as the unit sum constraint for the K classes of the parameter for the total population and the H subpopulations.

Both methods can be generalized to benchmark the series for the population total and two or more domain classifications simultaneously. Adding too many restrictions, however, might result in numerical problems for solving (20) or estimating the state space model.

4.5 Implementation of the Kalman filter

After having expressed the multivariate structural time series model in state space representation and under the assumption of normally distributed error terms, the Kalman filter can be applied to obtain optimal estimates for the state variables as well as the measurement equation see e.g. Durbin and Koopman, (2001). 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. So the smoothed estimate for the state vector for period t also accounts for the information made available after time period t. In this paper, point estimates and standard errors for the state variables are based on the smoothed Kalman filter estimates using the fixed interval smoother. See Harvey (1989) or Durbin and Koopman (2002) for technical details.

The non-stationary state variables are initialised with a diffuse prior, i.e. the expectations of the initial states are equal to zero and the initial covariance matrix of the states is diagonal with large diagonal elements. The time independent regression coefficients of the intervention variables are also initialised with a diffuse prior, as described by Durbin and Koopman (2001), subsection 6.2.2.

The analysis is conducted with software developed in Ox in combination with the subroutines of SsfPack 3.0, see Doornik (1998) and Koopman et al. (1999, 2008). In Ssfpack 3.0 an exact diffuse log-likelihood function is obtained with the procedure proposed by Koopman (1997). Maximum likelihood estimates for the hyperparameters, i.e. the variance components of the stochastic processes for the state variables are obtained using a numerical optimization procedure (BFGS algorithm, Doornik, 1998). To avoid negative variance estimates, the log-transformed variances are estimated.

5. Results

5.1 Results with four different time series models

The time series models developed in section 4 are applied to the series of "Separating chemical waste" and "Contact frequency with neighbors", which are plotted in Figure 1 and 2. The results obtained with four different models are

compared. These models assume that the series can be decomposed in a stochastic trend, a level intervention and an irregular term. Because the series concern annual data, it was not necessary to use a seasonal component. This allowed the selection of very parsimonious models, which was inevitable since the series are very short (11 years). Adding AR or MA components deteriorated the model fits and generally resulted in overfitting of the data.

The first model, denoted M1, is a seemingly unrelated structural time series model applied to the untransformed series. This model is defined by equations (6), (8), (9), (10a), (10b), (10c), and (10d). Note that there is no restriction for the estimated discontinuities. This is a seemingly unrelated structural time series model, since it is assumed that the variances of the irregular terms in the measurement equations are equal, i.e. $\sigma^2_{\varepsilon,1} = ... = \sigma^2_{\varepsilon,K} = \sigma^2_{\varepsilon}$. Due to the limited length of the series, this assumption is made to reduce the number of hyperparameters to be estimated.

The second model, denoted M2, is the restricted multivariate model defined by equations (6), (8), (9), (10a), (10b), (10d), (10e), and (10f). This implies that the observed series are not transformed and that the regression coefficients of the intervention variables are explicitly benchmarked by restriction $\mathbf{T}_{iv}$ defined in (10f). It is also assumed that $\sigma^2_{\varepsilon,1} = ... = \sigma^2_{\varepsilon,K} = \sigma^2_{\varepsilon}$.

The third model, denoted M3, is a seemingly unrelated structural time series model applied to the *K*-1 series obtained after applying logratio transformation (12) using the last category as the reference category in the denominator. This model is defined by (6), (8), (9) and (13). To reduce the number of hyperparameters it is assumed that $\sigma^2_{\varepsilon,1} = ... = \sigma^2_{\varepsilon,K-1} = \sigma^2_{\varepsilon}$

The fourth model, denoted M4, is the restricted multivariate model applied to the K series obtained after applying the central logratio transformation (15). This model is defined by equations (6), (8), (9), (10a), (10b), (10d), (10e), and (10f). It is assumed that $\sigma^2_{\varepsilon,1} = ... = \sigma^2_{\varepsilon,K-1} = \sigma^2_{\varepsilon}$.

For each model two analyses are conducted. One is based on the data available up to and including 2006, the other on the complete series, including 2007. This gives some intuition of the size of the revision of the estimate of the discontinuity if an additional observation under the new approach becomes available.

Estimation results for the discontinuities under the different models are given in Table 2 for the parameter "Separating chemical waste" and in Table 3 for the parameter "Contact frequency with neighbours".

As expected in advance, the estimated discontinuities under M1 do not obey the restriction $\sum_{k=1}^{K} \hat{\beta}_k = 0$. As a result, the corrected series are not consistent, since the categories for a parameter do not add up to one.

The multivariate model for the original series (M2) and the central logratio transformed series (M4) results in consistent series since the estimates for the discontinuities are forced to obey the required restriction. Augmenting the model with restriction (10f) also reduces the standard errors of the estimated discontinuities, since the restriction adds additional information to the model. This follows if the results obtained with the multivariate model (M2) are compared with the results obtained with the seemingly unrelated time series model (M1) for the original series.

Another way to preserve the consistency between the series of the *K* categories of a parameter is to apply the logratio transformation, since this transformation

eliminates the redundancy due to the unit sum constraint over the *K* categories. The estimated discontinuities for the logratio and central logratio transformation in Tables 2 and 3 are the results obtained with the transformed series.

Table 2: Estimated discontinuities for "Separating chemical waste" with different models

Model	T	Category									
		1		2		3		4		5	
M1	2006	4.29	(1.21)	-4.34	(1.21)	0.00	(1.21)	1.50	(1.21)	-1.44	(1.21)
M1	2007	1.91	(1.88)	-4.15	(0.77)	-0.07	(0.77)	1.49	(0.77)	-1.17	(0.98)
M2	2006	4.29	(1.07)	-4.35	(1.07)	-0.01	(1.07)	1.50	(1.07)	-1.44	(1.07)
M2	2007	3.07	(1.44)	-4.01	(0.75)	0.07	(0.75)	1.63	(0.75)	-0.76	(0.98)
M3*	2006	-0.06	(0.14)	-1.08	(0.20)	0.16	(0.10)	1.00	(0.20)		
M3*	2007	0.19	(0.15)	-0.77	(0.21)	0.23	(0.11)	0.68	(0.12)		
M4*	2006	-0.04	(0.26)	-1.06	(0.26)	0.22	(0.31)	1.01	(0.16)	-0.13	(0.07)
M4*	2007	-0.05	(0.25)	-1.09	(0.26)	0.17	(0.30)	1.00	(0.21)	-0.03	(0.07)

**: Results obtained for the (central) logratio transformed series. T: period of the last observation included in the analysis. Standard errors in brackets.*

Table 3: Estimated discontinuities for "Contact frequency neighbours" with different models

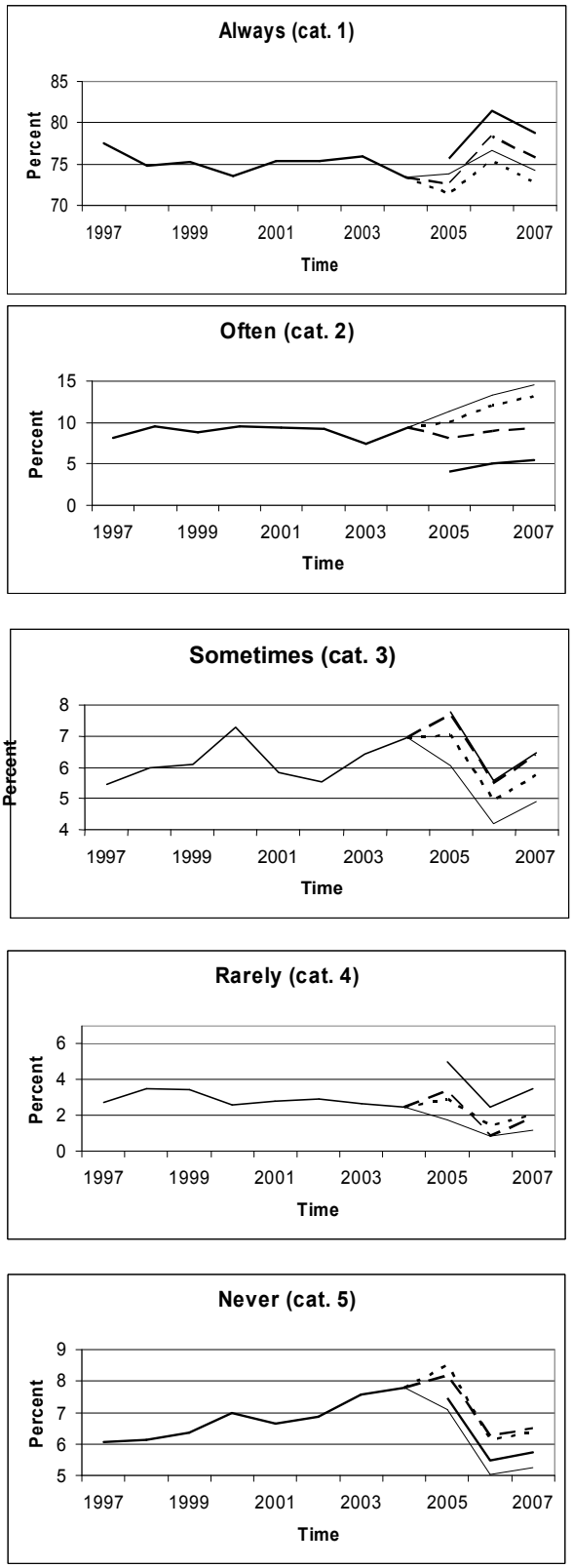
Model	T	Category							
		1		2		3		4	
M1	2006	4.79	(1.19)	0.31	(0.69)	-4.19	(1.32)	1.60	(0.51)
M1	2007	4.40	(1.20)	-0.09	(0.59)	-3.18	(1.30)	-1.36	(0.59)
M2	2006	5.02	(0.93)	0.46	(0.66)	-3.92	(0.96)	-1.56	(0.48)
M2	2007	4.44	(0.93)	-0.07	(0.56)	-3.01	(0.95)	-1.35	(0.56)
M3*	2006	0.33	(0.09)	0.27	(0.09)	0.16	(0.09)		
M3*	2007	0.38	(0.11)	0.30	(0.10)	0.14	(0.08)		
M4*	2006	0.14	(0.06)	0.08	(0.06)	-0.03	(0.06)	-0.19	(0.06)
M4*	2007	0.12	(0.05)	0.07	(0.05)	-0.03	(0.05)	-0.16	(0.05)

**: Results obtained for the (central) logratio transformed series. T: period of the last observation included in the analysis. Standard errors in brackets.*

The results obtained under equivalent models illustrate the size of the revision for the estimated discontinuities if the data for an additional year becomes available. Adding the estimates obtained in 2007 to the series results in a revision of the estimated discontinuities. Large revisions are observed for the first category of "Separating chemical waste" under model M1 and the fourth category of "Contact frequency with neighbours" under model M1. For the other three models the sizes of the revisions are smaller with respect to the standard errors. It can be expected that the size of the revisions decreases if the length of the series increases, particularly if the number of data points after the changeover increases.

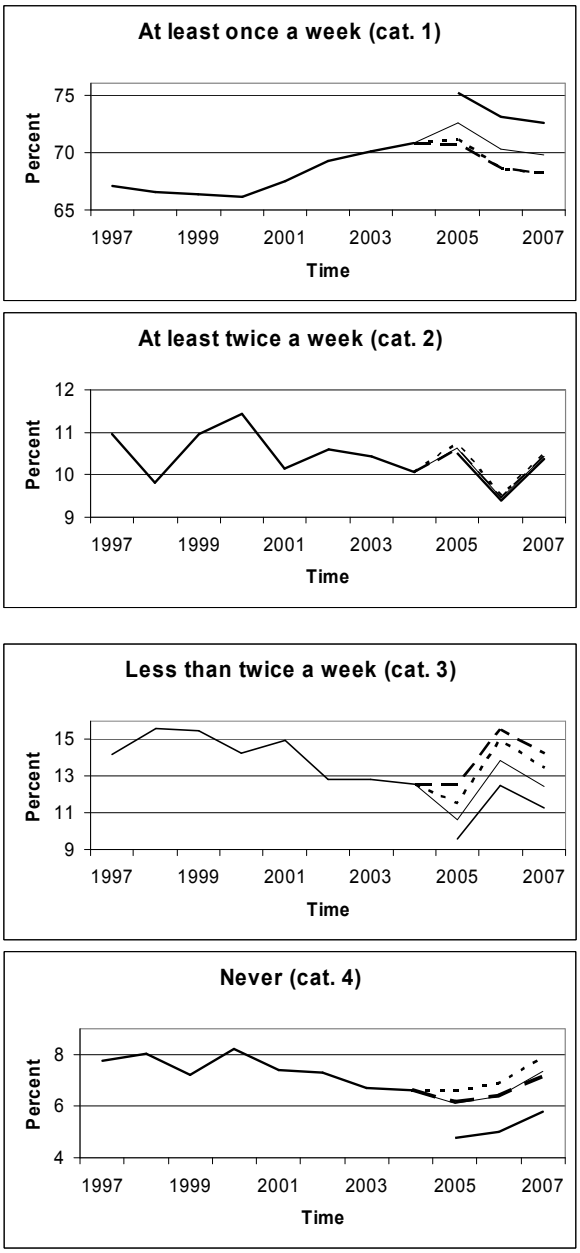
The original data, the corrected series obtained with models M2, M3, and M4, are shown in figures 3 and 4. The outcomes obtained under the SSPEC for the period 2005 through 2007 are corrected to make the series comparable with the outcomes of the PSLC, using the procedure described in section 4. In subsection 5.2 a simulation study is conducted to investigate which model is most appropriate to estimate discontinuities and produce corrected series for the variables of the PSLC and the SSPEC.

Figure 3: Separating chemical waste



Solid line 1997-2004 estimate based on the PSLC, solid line 2005-2007 estimate based on the SSPEC, dotted line corrected series based on a logratio transformation, dashed line corrected series based on untransformed data, thin solid line corrected series based on central logratio transformation.

Figure 4: Contact frequency with neighbours



Solid line 1997-2004 estimate based on the PSLC, solid line 2005-2007 estimate based on the SSPEC, dotted line corrected series based on a logratio transformation, dashed line corrected series based on untransformed data, thin solid line corrected series based on central logratio transformation.

5.2 Model evaluation

The underlying assumptions of the state-space model are that the disturbances of the measurement and system equations are normally distributed and serially independent with constant variances. There are different diagnostic tests available in the literature to test to what extent these assumptions are met, see Durbin and Koopman (2001), section 2.12.

In this application model evaluation is particularly important. The observed series are the outcome of variables that have a multinomial response at each time period. The Gaussian models M1 and M2 are applied to the untransformed data and therefore do not account for this property. Models M3 and M4 are also Gaussian, but account for the multinomial response through the logratio and a central logratio transformation. Durbin and Koopman (2000) and Durbin and Koopman (2001), chapters 10 and 11 describes simulation methods for the analysis of non-Gaussian models and can be used as an alternative.

Another point of concern is the limited length of the available series. Only 11 periods are observed, which might affect the precision of the maximum likelihood estimates for the hyperparameters and the smoothed Kalman filter estimates for the discontinuities. Furthermore, standard diagnostic tests to evaluate model assumptions will not have sufficient power to asses model-deficiencies and are therefore not very useful in this application. As an alternative two simulations are conducted.

5.2.1 Simulation with different time series lengths

In the first simulation the effect of the length of the series on the reliability of the estimates for the hyperparameters and the discontinuities is investigated. Replications of time series are generated from the unconditional distribution implied by model M3 using the maximum likelihood estimates for the hyperparameters and the smoothed estimates for the discontinuities obtained for the variable "Contact frequency with neighbours".

For each replication, states and observations are generated using the Ssfpack procedure `SsfRecursion` as described in Koopman et al. (2008), section 4.1. This procedure uses standard normal random numbers for the disturbance terms of the measurement and system equations. The maximum likelihood estimates for the hyperparameters and the smoothed estimates for the discontinuities are used to define the state-space model. Subsequently model M3 is applied to analyse the simulated time series.

Three different simulations are conducted. In the first simulation, time series with a length of 11 observations, 8 before and 3 after the survey redesign, are generated. In the second simulation, time series with a length of 22 observations, 16 before and 6 after the survey redesign, are generated. In the third simulation, time series with a length of 44 observations, 32 before and 12 after the survey redesign, are generated. The variance of the irregular terms of the measurement equation is inversely proportional to the yearly sample size of the survey. For the first simulation the actual sample sizes of the PSLC and the SSPEC are used. In the second and the third simulation additional sample sizes are generated from a uniform distribution where the minimum and maximum yearly sample size of the PSLC and the SSPEC are used as the lower and upper boundaries of the uniform distribution. For each simulation study 10,000 time series are generated.

The resample distributions of the maximum likelihood estimates for the hyperparameters and the smoothed estimates for the discontinuities are used to obtain more insight in the reliability of these model estimates in this application where only a limited number of data points are available. In Table 4 the means and standard errors of the resample distributions of the estimated hyperparameters and discontinuities are compared with the values used in the assumed distribution. The

resample distributions of the estimated hyperparameters and discontinuities are plotted in Figures 5 and 6.

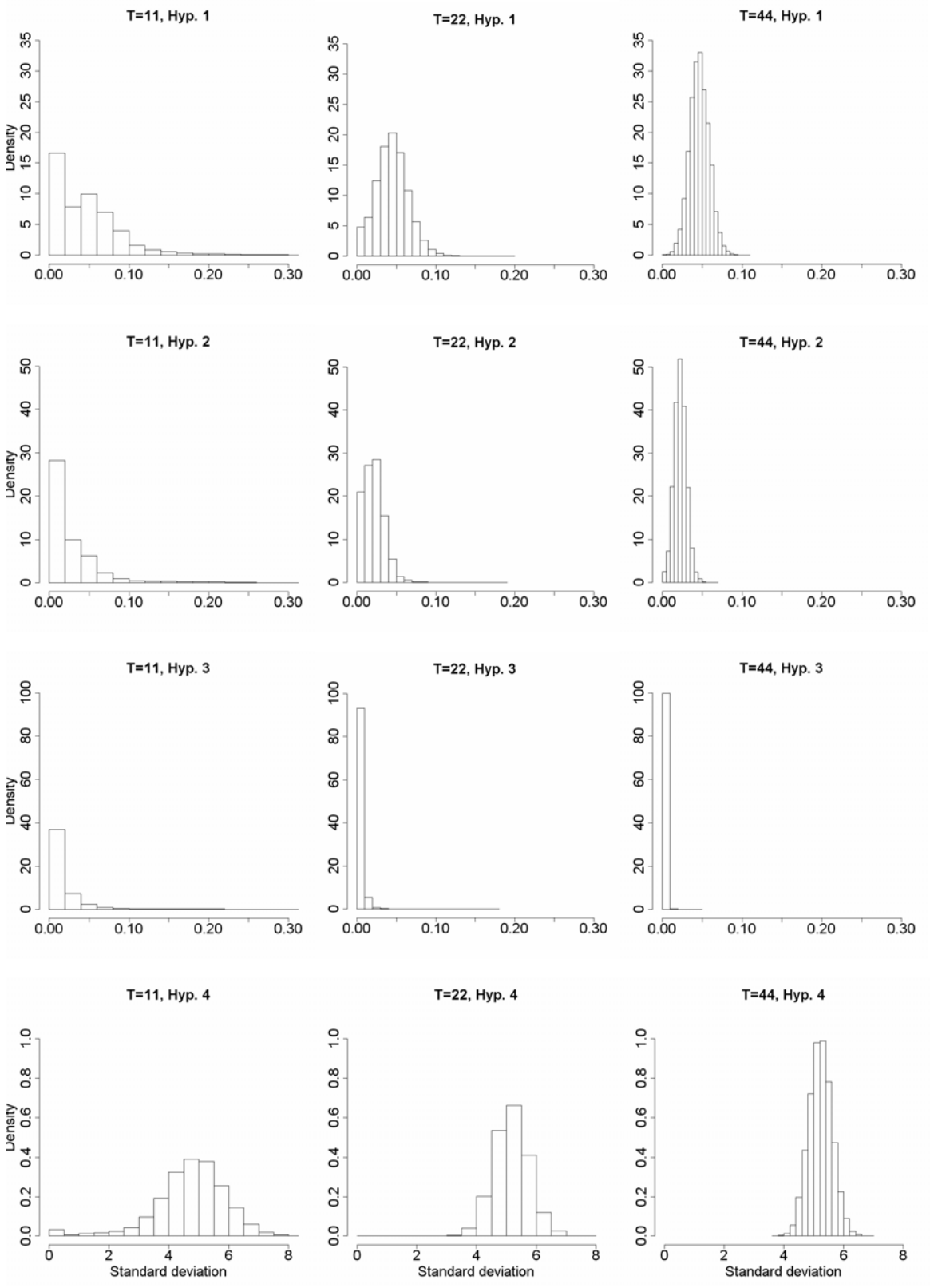
Table 4: Simulation results for the estimated hyperparameters and discontinuities with different lengths of the times series

Parameter	Real values	Simulated values					
		T=11		T=22		T=44	
Hyp. 1	0.0480	0.0460	(0.0464)	0.0445	(0.0208)	0.0467	(0.0123)
Hyp. 2	0.0237	0.0261	(0.0412)	0.0210	(0.0139)	0.0227	(0.0079)
Hyp. 3	0.000	0.0170	(0.0392)	0.0027	(0.0064)	0.0006	(0.0014)
Hyp. 4	5.260	4.7182	(1.2177)	5.1664	(0.5833)	5.2223	(0.3869)
Disc. 1	0.380	0.380	(0.141)	0.378	(0.124)	0.379	(0.123)
Disc. 2	0.300	0.298	(0.122)	0.300	(0.105)	0.300	(0.101)
Disc. 3	0.140	0.142	(0.104)	0.139	(0.070)	0.140	(0.049)

Hyp.1, Hyp. 2, Hyp. 3: Standard deviations irregular terms of the slope from the trend model for three series obtained after logratio transformation, i.e. $\sigma_{R,1}$, $\sigma_{R,2}$, $\sigma_{R,3}$. *Hyp. 4: Standard deviation irregular terms of the measurement equations, i.e.* σ_{ε}. *Disc. 1, Disc. 2, Disc. 3: Discontinuity for three series obtained after logratio transformation, i.e.* β_1, β_2, β_3. *Standard errors in brackets.*

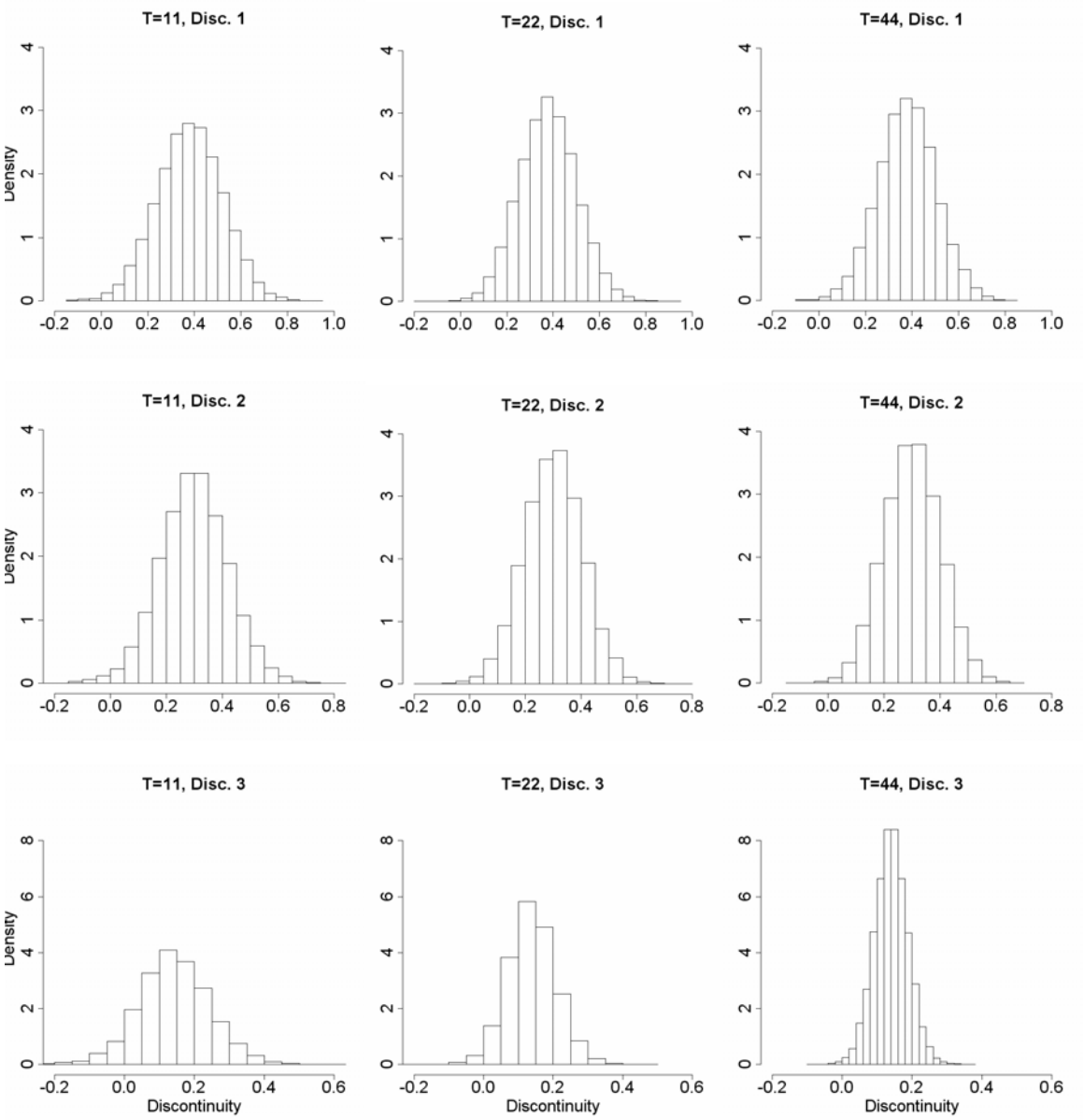
The mean of the resample estimates for the hyperparameters and the discontinuities are not significantly different from the real values used in the assumed distribution in each of the three simulations with different time series lengths. This implies that there are no indications that a limited number of observations results in biased parameter estimates. The precision of the maximum likelihood estimates of the hyperparameters clearly improves with the length of the time series. It follows from Table 4 that the size of the standard errors decreases with the length of the series. The same conclusion follows from figure 5. Short series result in wide and skewed resample distributions around the true values. The resample distributions centre on the true value and become more symmetrically if the length of the series increases. The precision of the smoothed estimates of the discontinuities, on the other hand, is much better in the case of the shortest time series. It can be seen from Table 4 that the decrease of the standard errors if the length of the series increases is much smaller compared to the hyperparameters. The same conclusion follows from Figure 6. The effect of the length of the series on the dispersion of the resample distribution around the true values is much smaller. The sample distributions are also allocated more symmetrically around the true values, even in the case of the shortest time series.

Figure 5: Resample distributions estimated hyperparameters for different time series lengths.



Hyp.1, Hyp. 2, Hyp. 3: Standard deviations irregular terms of the slope from the trend model for three series obtained after logratio transformation, i.e. $\sigma_{R,1}$, $\sigma_{R,2}$, $\sigma_{R,3}$. *Hyp. 4: Standard deviation irregular terms of the measurement equations, i.e.* σ_{ε}.

Figure 6: Resample distributions estimated discontinuities for different time series lengths.



Disc. 1, Disc. 2, Disc. 3: Discontinuity for three series obtained after logratio transformation, i.e. β_1, β_2, β_3.

5.2.2 Simulation with different models under multinomial response

In the second simulation the performance of the four models, used in subsection 5.1, under a multinomial response with different discontinuities is studied. In this simulation, time series with a length of 11 time points are generated as follows. For each time point n_t independent trials are drawn from a multinomial distribution with parameters n_t and $\mathbf{p}_t = (p_{t,1}, p_{t,2}, p_{t,3}, p_{t,4})$, with n_t the yearly sample size and $\mathbf{p}_t$ the observed distribution over the four categories of "Contact frequency with neighbours" observed with the PSLC in the first 8 years and the SSPEC in the last 3 years. The distributions observed with the SSPEC are corrected for the estimated discontinuities obtained with model M2. Thus $\mathbf{p}_t = \hat{\mathbf{y}}_t$ if $t \leq 2004$ and $\mathbf{p}_t = \hat{\mathbf{y}}_t - \hat{\boldsymbol{\beta}}$ if $t > 2004$. According to this approach uninterrupted time series $\mathbf{p}_t^{r*}$ are generated.

Subsequently two different types of discontinuities are added to the last three time points of the series, i.e. $\mathbf{p}_t^r = \mathbf{p}_t^{r*} + \boldsymbol{\Delta}_t$. The first set of discontinuities are chosen constant over time by taking $\boldsymbol{\Delta}_t = (4.5, -0.1, -3.0, -1.4)^t$ for t = 2005, 2006 and 2007. These discontinuities are approximately equal to the estimated discontinuities under model M2, see Table 3. The second set of discontinuities is derived from the estimation results obtained with model M3. Time varying discontinuities are obtained by taking $\boldsymbol{\Delta}_t = \hat{\mathbf{y}}_t - \tilde{\mathbf{y}}_t$ for t = 2005, 2006 and 2007. Here $\hat{\mathbf{y}}_t$ are the originally observed series under the SSPEC and $\tilde{\mathbf{y}}_t$ the adjusted series obtained with the inverse of the logratio transformation (14). Although M3 assumes a time independent regression coefficient for the intervention variable, the discontinuities become time dependent since the adjusted series is mapped from the real space back to the simplex with the inverse of the logratio transformation (14).

In each simulation 10,000 series are generated and analysed with the four models proposed in subsection 5.1. Let $\hat{\boldsymbol{\Delta}}_t^r$ denote the estimated discontinuities for time periods t = 2005, 2006 and 2007 for the r-th replicate. For models M1 and M2 the estimated discontinuities are equal to the estimated regression coefficients of the intervention variable, i.e. $\hat{\boldsymbol{\Delta}}_t^r = \hat{\boldsymbol{\beta}}^r$ and thus constant in time. For models M3 and M4 the simulated series are transformed using the logratio and the central logratio transformation respectively. Time varying discontinuities for the r-th replicate are estimated as the difference between the original and adjusted series, i.e. $\hat{\boldsymbol{\Delta}}_t^r = \mathbf{p}_t^r - \tilde{\mathbf{p}}_t^r$ for t = 2005, 2006 and 2007. Here $\tilde{\mathbf{p}}_t^r$ denotes the adjusted series for the r-th replicate obtained with the inverse of the logratio transformation (14) or the inverse central logratio transformation (17).

In Table 5, the mean and standard errors of the estimated discontinuities $\hat{\boldsymbol{\Delta}}_t^r$ are summarised for the simulation with constant discontinuities. Standard errors are obtained with the resample standard deviation. In Table 6 the same analysis results are specified for the simulations with time dependent discontinuities.

For each model it follows that the real values of the discontinuities are included in the 95% confidence intervals of the simulated means. Nevertheless it can be concluded that the simulated means of the discontinuities of model M1 and M2 are closer to the real values of the discontinuities than models M3 and M4. This is the case for the simulation with constant discontinuities (Table 5) but also for the time varying discontinuities (Table 6). Furthermore the simulated standard errors under models M1 and M2 are smaller than the simulated standard errors obtained with models M3 and M4.

Table 5: Real and simulated values time independent discontinuities

	Discontinuity							
	Cat. 1		Cat. 2		Cat. 3		Cat. 4	
Real value	4.5		-0.1		-3.0		-1.4	
M1	4.400	(1.232)	0.037	(0.631)	-2.672	(1.248)	-1.529	(0.489)
M2	4.266	(1.209)	-0.001	(0.650)	-2.694	(1.125)	-1.572	(0.497)
M3-2005	3.489	(1.430)	0.042	(0.759)	-1.818	(1.118)	-1.713	(0.578)
M3-2006	3.946	(1.682)	0.100	(0.685)	-2.274	(1.437)	-1.773	(0.696)
M3-2007	3.976	(1.677)	0.108	(0.745)	-2.038	(1.230)	-2.046	(0.850)
Mean value M3*	3.804		0.083		-2.043		-1.844	
M4-2005	3.353	(1.336)	0.191	(0.864)	-1.935	(1.443)	-1.609	(0.577)
M4-2006	3.852	(1.658)	0.230	(0.775)	-2.426	(1.853)	-1.657	(0.680)
M4-2007	3.847	(1.591)	0.256	(0.852)	-2.192	(1.707)	-1.911	(0.825)
Mean value M4*	3.684		0.226		-2.184		-1.725	

**: Mean over the three years. Standard errors between brackets.*

Table 6: Real and simulated values time dependent discontinuities

	Discontinuity							
	Cat. 1		Cat. 2		Cat. 3		Cat. 4	
Real value 2005	4		-0.21		-1.96		-1.83	
Real value 2006	4.45		-0.11		-2.46		-1.88	
Real value 2007	4.47		-0.12		-2.20		-2.15	
M1	3.788	(1.207)	-0.035	(0.614)	-1.562	(1.134)	-1.975	(0.446)
M2	3.665	(1.153)	-0.072	(0.629)	-1.582	(1.052)	-2.011	(0.459)
M3-2005	2.997	(1.245)	-0.041	(0.710)	-0.845	(0.932)	-2.111	(0.538)
M3-2006	3.207	(1.422)	-0.010	(0.645)	-0.993	(1.123)	-2.204	(0.681)
M3-2007	3.331	(1.461)	0.001	(0.703)	-0.896	(1.041)	-2.437	(0.830)
M4-2005	2.910	(1.153)	0.064	(0.781)	-0.925	(1.184)	-2.048	(0.548)
M4-2006	3.146	(1.361)	0.083	(0.705)	-1.095	(1.445)	-2.134	(0.679)
M4-2007	3.246	(1.348)	0.107	(0.774)	-0.996	(1.348)	-2.357	(0.813)

Standard errors between brackets.

5.3 Implementation

The simulations indicate that time series models applied to the untransformed series result in more accurate estimates for the discontinuities than the models applied to the logratio or central logratio transformed series. The main advantage of the logratio and central logratio transformation is that the adjusted values add up to one and always take values within the admissible range of [0,1] by definition. The major drawback of both transformations is that the interpretation of the results is complex. The estimated discontinuities as well as the corrected series for a particular class are influenced by the discontinuity of the reference class in the case of the logratio transformation. In the case of the central logratio transformation the estimated discontinuities as well as the corrected series for each particular class is influenced by the discontinuities of all other classes, via the geometric mean over all classes in the denominator of this transformation. An additional disadvantage of the logratio transformation is that the results depend on the choice of the reference category to be used in the denominator of the logratio transformation.

The advantage of the multivariate model applied to the untransformed data is that the interpretation of the results is straightforward and that the estimated discontinuities for the separated categories are only affected by the other categories through the zero sum constraint. The major drawback is that the corrected values might take values outside the admissible range of [0,1]. This, however, did not occur in this application.

Based on these considerations, the multivariate model M2 applied to the untransformed data is finally used in this application to estimate discontinuities and calculate corrected time series for all other parameters about environmental consciousness and social participation. The common picture of the effect of the redesign is an increase of the proportion of respondents in the first categories compensated by a decrease in the last categories after the change-over. See Van den Brakel and Roels (2009) for a more detailed discussion about the results.

In this application, the series for the two domains of gender were also analyzed and adjusted for the observed discontinuities. For a few parameters, the method of Lagrange multipliers, describes in section 4.4, was applied to restore the consistency with the series for the total population. In this case the covariance matrix in (20) was taken diagonal with the variances of the smoothed Kalman filter estimates for the regression coefficients of the intervention variables as elements. This benchmark resulted in small modifications of the adjusted series.

Consistent time series can be obtained by correcting the observed series for the estimated discontinuity. Depending on the anticipated impact of the redesign on the quality of the estimates, the series observed in the past can be adjusted to make it comparable with the outcomes obtained under the new design. It is also possible to adjust the outcomes obtained under the new approach to make them comparable with the series under the old survey design. In this application the data collection mode changed from CAPI under the PSLC to CATI under the SSPEC. Therefore it is anticipated that the series observed in the past are more accurate than the outcomes obtained under the SSPEC. Indeed, with the CAPI mode the entire target population is reached while the CATI mode only surveys the subpopulation with a listed telephone number. Furthermore less measurement errors and social desirable answers are expected under the CAPI mode due to the personal contact with an interviewer and the lower interview speed, see e.g. Holbrook et al. 2003, Roberts 2007. Based on these considerations it was decided that the outcomes obtained under the SSPEC are corrected to make the series comparable with the outcomes of the PSLC. Under the assumption that the development observed with the CATI data is representative for the entire target population, consistent time series are obtained.

6. Discussion

The relevance of official statistics, produced by national statistical institutes, strongly depends on the comparability of the outcomes over time. A redesign of the survey process generally results in discontinuities in time series obtained with repeatedly conducted sample surveys. To avoid the confounding of real developments with the systematic effect induced by the redesign, structural time series models with an intervention variable are developed to estimate the size of the discontinuities. This approach relies on the assumption that there is no structural change in the evolution of the series of the population value at the moment that the survey is redesigned. Additional auxiliary information and subject matter expert knowledge can e used to asses whether the assumption that there is no structural change in the real evolution of the population variable is tenable. Auxiliary time series can be incorporated in the model to improve the estimates for the discontinuities. If this assumption is questionable, experiments where both surveys are run in parallel for some period of time should be considered as an alternative.

The transition of the PSLC to the SSPEC resulted in systematic differences in the estimates for parameters about environmental consciousness and social participation. In this application, Gaussian state-space models are applied to compositional time series which are derived from variables with a multinomial response at each time period. In a simulation study the performance of multivariate models applied to untransformed, logratio transformed and central logratio transformed series are compared. In this application the most accurate estimates for the discontinuities are obtained with a multivariate model applied to the untransformed series that accounts for the unit sum constraint. This is a remarkable result, since the logratio and central logratio transformations were considered to account for the multinomial response. It is worthwhile to investigate to what extent simulation methods for the analysis of non-Gaussian models further improve the accuracy of the estimated discontinuities.

Another point of concern is the limited length of the available series. Simulations indicate that the dispersion of the resample distribution of the maximum likelihood estimates for the hyperparameters narrows rapidly if the length of the available series increases. The dispersion of the resample distribution of the smoothed estimates of the discontinuities, on the other hand, remains more stable if the length of the series in the simulations increases. Therefore it appears that although the maximum likelihood estimates of the hyperparameters of the state-space models can be far from the true values under the available series, the models already produce useful estimates for the discontinuities. This is a plausible result. Most information about the size of the discontinuity comes from the observations close to the moment of the survey redesign. This also depends on the flexibility of the other model components. The discontinuities are increasingly based on local observations close to the moment of the survey redesign, as the trend and other model components are more flexible.

One aspect of the time series approach is that more observations under the new approach become available when time proceeds. The advantage is that the discontinuities can be quantified more accurately if this additional information becomes available. A concomitant drawback is that the estimated discontinuities three years after redesigning the survey are still subject to revisions. A publication policy is required to deal with these revisions in practice. For this application it was decided to base the final estimates for the discontinuities on the information available up until 2007 in this application.

References

Abraham, B., and Vijayan, K. (1992). Time series analysis for repeated surveys. *Communications in Statistics: Simulation Computation*, 21, 893-908.

Aitchison, J. (1986). *The statistical analysis of compositional data*. London: Chapman and Hall.

Akinbami, L.J., and Schoendorf, K.C. (2002). Trends in childhood asthma: Prevalence, health, care utilization and mortality. *Pediatrics,* 110, 315-322.

Akinbami, L.J., Schoendorf, K.C., and Parker, J. (2003). US childhood asthma prevalence estimates: The impact of the 1997 National Health Interview Survey Redesign. *American Journal of Epidemiology*, 158, 99-104.

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

Bell, W.R., and Hillmer, S.C. (1990). The time series approach to estimation of periodic surveys. *Survey Methodology*, 16, pp. 195-215.

Binder, D.A., and Dick, J.P. (1989). Modeling and estimation for repeated surveys. *Survey Methodology*, 15, pp. 29-45.

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

Caban, A.J., Lee, D.J., Fleming, L.E., Gómez-Marin, O., LeBlanc, W., and Pitman, T. (2005). Obesity in US workers: The National Health Interview Survey, 1986-2002. *American Journal of Public Health*, 95, 1614-1622.

Cochran, W.G. (1977). *Sampling techniques*. Third edition. New York: John Wiley & Sons.

Blight, B.J.N., and Scott, A.J. (1973). A stochastic model for repeated surveys. *Journal of the Royal Statistical Society*, *Series B*, 35, 61-66.

Brackstone, G. (1999). Managing data quality in a statistical agency. *Survey Methodology*, 25, 139-149.

Brunsdon, T.M., and Smith, T.M.F. (1998). The time series analysis of compositional data. *Journal of Official Statistics*, 14, 237–253.

De Leeuw, E. (2005). To mix or not to mix data collection modes in surveys. *Journal of Official Statistics*, 21, 233–255.

Dillman, D.A., and Christian, L.M. (2005). Survey mode as a source of instability in responses across surveys. *Field Methods*, 17, 30-52.

Dippo, C.S., Kostanich, D.L., and Polivka, A.E. (1994). Effects of methodological change in the Current Population Survey. *Proceedings of the section on Survey Research Methods*, American Statistical Association, pp. 260-262.

Doornik, J.A. (1998). *Object-oriented matrix programming using Ox 2.0*. London: Timberlake Consultants Press.

Durbin, J., and Koopman, S.J. (2000). Time series analysis of non-Gaussian observations based on state-space models from both classical and Bayesian perspectives (with discussion). *Journal of the Royal Statistical Society, Series B*, 62, 3-56.

Durbin, J., and Koopman, S.J. (2001). *Time series analysis by state space methods*. Oxford: Oxford University Press.

Feder, M. (2001). Time series analysis of repeated surveys: the state-space approach. *Statistica Neerlandica*, 55, pp. 182-199.

Fowler, F.J. (1996). The redesign of the National Health Interview Survey. Public Health Reports, 111, 508-511.

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

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

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

Holbrook, A.L., Green, M.C., and Krosnick, J.A. (2003). Telephone versus Face-to-Face Interviewing of National Probability Samples with Long Questionnaires. *Public Opinion Quarterly*, 67, pp. 79-125.

Kalton, G., and Schuman, H. (1982). The effect of the question on survey responses: a review. *Journal of the Royal Statistical Society*, *Series A*, 145, 42-73.

Kindermann, C., and Lynch, J. (1997). Effects of the redesign on vistimization estimates. Technical report, US department of Justice, Bureau of Justice Statistics,
http://www.ojp.usdoj.gov/bjs/abstract/erve.htm

Koopman, S.J., Shephard, N., and Doornik, J.A. (1999). Statistical algorithms for models in state space using SsfPack 2.2. *Econometrics Journal*, 2, 113-166.

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

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

Lind, J.T. (2005). Repeated surveys and the Kalman filter. *Econometrics Journal,* 8, 418-427.

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

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

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

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, pp. 339-348.

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

Roberts, C. (2007). Mixing modes of data collection in surveys: A methodological review. Review paper, NCRM/008, National Centre for Research Methods, City University London.

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

Scott, A.J., and Smith, T.M.F. (1974). Analysis of repeated surveys using time series methods. *Journal of* the *American Statistical Association*, 69, 674-678.

Scott, A.J., T.M.F. Smith, and Jones, R.G. (1977). The application of time series methods to the analysis of repeated surveys. *International Statistical Review*, 45, 13-28.

Silva, D.B.N., and Smith, T.M.F. (2001). Modelling compositional time series from repeated surveys. *Survey* Methodology, 27, pp. 205-215.

Tam, S.M. (1989). Analysis of repeated surveys using a dynamic linear model. *International Statistical Review*, 55, 63-73.

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

Van den Brakel, J.A. (2008). Design-based analysis of embedded experiments with applications in the Dutch Labour Force Survey. *Journal of the Royal Statistical Society*, *Series A*, 171, 581-613.

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*, in press.

Van den Brakel, J.A. and Roels, J. (2009). A time series approach to estimate discontinuities due to a survey redesign. Discussion paper 09003, Statistics Netherlands, Heerlen (http://www.cbs.nl/en-GB/menu/methoden/research/discussionpapers/archief/2009/ default.htm?Languageswitch=on).

Van den Brakel, J.A. & Renssen, R.H. (2005). Analysis of experiments embedded in complex sampling designs. *Survey Methodology,* 31, 23-40.

Van den Brakel, J.A., Smith, P.A., and Compton, S. (2008). Quality procedures for survey transitions, experiments, time series and discontinuities. *Journal for Survey Research Methods*. 2, 123-141.