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

Survey Methodology

Official Statistics based on the Dutch Health Survey during the Covid-19 Pandemic

by Jan van den Brakel and Marc Smeets

Release date: June 30, 2023

Statistics Canada Statistique Canada

Canada

How to obtain more information

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

You can also contact us by

Email at infostats@statcan.gc.ca

Telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following numbers:

• Statistical Information Service	1-800-263-1136
• National telecommunications device for the hearing impaired	1-800-363-7629
• Fax line	1-514-283-9350

Standards of service to the public

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

Note of appreciation

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

Published by authority of the Minister responsible for Statistics Canada

An HTML version is also available.

Cette publication est aussi disponible en français.

Survey Methodology, June 2023
Vol. 49, No. 1, pp. 39-68
Statistics Canada, Catalogue No. 12-001-X

Official Statistics based on the Dutch Health Survey during the Covid-19 Pandemic

Jan van den Brakel and Marc Smeets[1]

Abstract

The Dutch Health Survey (DHS), conducted by Statistics Netherlands, is designed to produce reliable direct estimates at an annual frequency. Data collection is based on a combination of web interviewing and face-to-face interviewing. Due to lockdown measures during the Covid-19 pandemic there was no or less face-to-face interviewing possible, which resulted in a sudden change in measurement and selection effects in the survey outcomes. Furthermore, the production of annual data about the effect of Covid-19 on health-related themes with a delay of about one year compromises the relevance of the survey. The sample size of the DHS does not allow the production of figures for shorter reference periods. Both issues are solved by developing a bivariate structural time series model (STM) to estimate quarterly figures for eight key health indicators. This model combines two series of direct estimates, a series based on complete response and a series based on web response only and provides model-based predictions for the indicators that are corrected for the loss of face-to-face interviews during the lockdown periods. The model is also used as a form of small area estimation and borrows sample information observed in previous reference periods. In this way timely and relevant statistics describing the effects of the corona crisis on the development of Dutch health are published. In this paper the method based on the bivariate STM is compared with two alternative methods. The first one uses a univariate STM where no correction for the lack of face-to-face observation is applied to the estimates. The second one uses a univariate STM that also contains an intervention variable that models the effect of the loss of face-to-face response during the lockdown.

Key Words: Small area estimation; Structural time series model; Corona crisis.

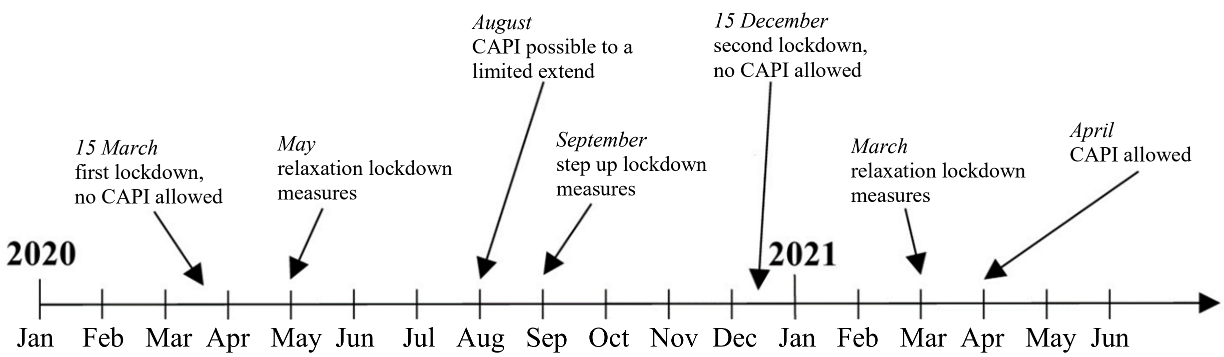
1. Introduction

The Dutch Health Survey (DHS) is a continuing survey conducted by Statistics Netherlands that measures health, healthcare use and lifestyle in the Netherlands. Data collection is based on a sequential mixed-mode design where a combination of web participation (Computer-assisted web interviewing (CAWI)) and face-to-face interviewing (Computer-assisted personal interviewing (CAPI)) is applied. Due to Dutch lockdown measures during the Covid-19 pandemic face-to-face interviewing was not allowed in parts of 2020 and 2021. Figure 1.1 displays a timeline of the lockdowns in the Netherlands and the restrictions on the CAPI mode for the DHS. In the rest of these years, there were restrictions on the normal way of data collection. This results in an abrupt change in the composition of selection effects and measurement bias and therefore results in a systematic effect on the outcomes of the DHS. A second issue is that the DHS is designed to produce reliable estimates on an annual basis, using standard direct estimators like the general regression (GREG) estimator (Särndal, Swensson and Wretman, 1992). The DHS normally publishes on an annual basis for year τ in the month of March of year $\tau+1$. The Covid-19 pandemic that started in the beginning of 2020 made clear that the release of annual data about the effect of Covid-19 on health-related themes with a delay of about one year strongly compromises the

1. Jan van den Brakel, Statistics Netherlands and Maastricht University. E-mail: ja.vandenbrakel@cbs.nl; Marc Smeets, Statistics Netherlands.

relevance of this survey. Another disadvantage of annual figures is that the period of the corona crisis is not well-delineated in the reference period of the DHS. In the second quarter of 2020, there was indeed a strong external demand for quarterly figures of the DHS, since quarterly figures are more timely and better delineate the corona period. The sample size of the DHS, however, does not allow the production of sufficiently precise direct estimates for quarterly reference periods.

Figure 1.1 Timeline of Dutch coronavirus lockdowns and restrictions on CAPI mode for DHS, January 2020 to June 2021.



To solve these issues, a bivariate structural time series model (STM) is developed for eight key variables of the DHS, defined on a quarterly frequency. This model is used to correct for the changes of measurement and selection errors due to the loss of CAPI response and is used as a form of small area estimation (Rao and Molina, 2015) since the model uses sample information observed in previous reference periods to produce sufficiently reliable model-based estimates for quarterly DHS figures. In small area estimation this is commonly called borrowing strength over time.

The models proposed in this paper can be considered as an extension of the area level model (Fay and Herriot, 1979). The extension of the area level model with a temporal component is originally proposed by Rao and Yu (1994). In this paper a time series multilevel model is applied where an AR(1) component for the domain irregular terms is assumed. Other authors who proposed time series multilevel models as an extension of the area level model are Datta, Lahiri, Maiti and Lu (1999), You, Rao and Gambino (2003), You (2008), Boonstra, van den Brakel and Das (2021) and Boonstra and van den Brakel (2022). Another class of time series models that are frequently used as a form of small area estimation are state-space models. Pfeffermann and Burck (1990), Pfeffermann and Bleuer (1993), Pfeffermann and Tiller (2006) and Krieg and van den Brakel (2012) use multivariate state-space models as a form of small area estimation to borrow strength over time and space. Pfeffermann (1991), Harvey and Chung (2000) and van den Brakel and Krieg (2016) propose multivariate time series models as a form of small area estimation for Labour Force surveys that are designed as a rotating panel. The basic difference of the state-space models with aforementioned time series multilevel models is that the population irregular terms are combined with the sampling error into one measurement error. Another difference is that these models are also applied to time series at the national level to borrow strength over time only in situations where the reference period is too short to collect sufficient data to use a direct estimator even at the national level, see e.g., Pfeffermann (1991), Tiller (1992) and Harvey and Chung (2000). This paper follows the aforementioned state space approach.

Buelens and van den Brakel (2015) proposed a weighting method for sequential mixed-mode designs to stabilize the bias in period-to-period changes that arise from fluctuations in the distributions of respondents over the data collection modes in subsequent editions of a repeated survey. This method assumes a fixed distribution of the population over the different data collection modes, which is added as an additional component to the weighting model of the GREG estimators. This method cannot be considered as an alternative to compensate for the loss of CAPI during the lockdown. The method indeed increases the weights of the CAPI respondents, but will in this case increase selection bias as well because the CAPI respondents are all observed outside the lockdown period.

The net effect of the lack of CAPI is computed based on the response of previous years. This is done by removing CAPI from the response and by reweighting the remaining response. This leads to two direct estimates for one target variable: one based on the complete response (CAWI and CAPI) and one based on only web response (CAWI). In this way quarterly time series can be constructed for DHS that start in the first quarter of 2014: the complete series based on full response and the web series based on web response only. Both series are the input for the bivariate STM. The web series is available in all quarters, also during the lockdown. In quarters without CAPI there are no estimates available for the complete series and the bivariate STM then provides nowcasts for the missing figures based on the web series.

In this paper the bivariate STM is compared with two alternative and more straightforward models. The first one is a univariate STM where no correction for the lack of CAPI is applied. This method applies a univariate STM to the series of direct estimates based on all available response in every quarter. In quarters where CAPI is available the direct estimates are based on both CAWI and CAPI, so they are equal to the estimates of the complete series. In quarters where no CAPI is available the direct estimates are based on only CAWI and are thus equal to the estimates of the web series. The second one is a univariate STM that also contains an intervention variable that models the effect of the loss of CAPI during the lockdown.

The paper is organized as follows. Section 2 gives a description of the Dutch Health Survey and both the univariate and bivariate structural time series models are developed in Section 3. Section 4 explores the results and Section 5 discusses the officially published quarterly DHS figures by Statistics Netherlands. The paper ends with a discussion in Section 6.

2. Dutch Health Survey

The Dutch Health Survey is a continuing survey that measures health, healthcare use and lifestyle in the Netherlands on a yearly basis. The target population is the Dutch population living in private households. Each month a single-stage stratified sample of approximately 1,250 persons is drawn from the Dutch Personal Records Database. The strata are defined by the municipalities.

Sampled persons are asked to participate via web interviewing (CAWI). Non-respondents are re-approached to participate in a face-to-face interview (CAPI). To reduce administration costs, the fraction of CAWI responses is increased by selecting samples from the CAWI non-respondents that are re-approached through CAPI using a target group strategy that has been used since 2018. CAWI non-respondents are first divided into so-called target groups based on age, income and migration background. From each target group only a sample is re-approached.

Until 2020 there was a yearly response of approximately 10,000 persons, of whom 6,500 responded by CAWI and 3,500 by CAPI. The response is more or less evenly divided over the months. Due to the Covid-19 pandemic that started in 2020 there was a lockdown in the Netherlands that started mid-March 2020. The first relaxations were implemented in May 2020. Due to this lockdown no face-to-face interviews were allowed from mid-March 2020 to the end of July 2020. A second lockdown started in mid-December 2020, which was gradually relaxed from March 2021. This lockdown resulted in a stop of face-to-face interviewing from mid-December 2020 until the end of March 2021. From April 2021 face-to-face interviews were possible again. In order to increase response during the pandemic, persons selected for CAPI were given the opportunity to respond via the internet. This was done by sending an invitation letter when face-to-face interviewing was not allowed and by handing over this letter otherwise. In 2020 only few people used this option and they were considered as CAWI respondents. In 2021 a substantial part of the people selected for CAPI responded via the internet. This response mode will be referred to as CAPI/CAWI response. The resulting response sizes per month and response mode are shown in Table 2.1.

Table 2.1 shows that in 2020 CAPI response is lower in the months March and December and is completely missing from April to July. The large CAWI response size in May is the result of compensation measures taken by Statistics Netherlands for the response gaps that arose due to the lockdown. In 2021 CAPI response is completely missing in the first quarter and is lower in April and May. From June CAPI response seems to recover.

Annual figures are obtained by weighting the response by means of the general regression estimator (Särndal et al., 1992). In this way it is corrected, at least partially, for selective non-response. The weighting model is given by Gender_2 × Age_16 + MaritalStatus_4 + Urbanization_5 + Region_16 + HouseholdSize_5 + Gender_2 × Age_3 × MaritalStatus_4 + Region_4 × Age_3 + Migration Background_4 + SurveySeason_4 + Income_5 + Wealth_5 + TargetGroup_12. The numbers refer to the number of categories and the times sign indicates the use of interaction terms between variables. Note that TargetGroup_12 is included since 2018.

Table 2.1
Response DHS 2020 per mode and month

		CAPI	CAPI/CAWI	CAWI	Total
2020	January	265		584	849
	February	261		586	847
	March	104		917	1,021
	April	0		455	455
	May	0		1,118	1,118
	June	0		708	708
	July	0		483	483
	August	193		527	720
	September	286		259	545
	October	149		763	912
	November	181		587	768
	December	53		286	339
	Total	1,492		7,273	8,765
2021	January	0	48	738	786
	February	0	36	546	582
	March	0	22	655	677
	April	38	77	460	575
	May	51	62	738	851
	June	109	62	283	454
	Total	198	307	3,420	3,925

Note: Dutch Health Survey (DHS); Computer-assisted personal interviewing (CAPI); Computer-assisted web interviewing (CAWI).

In consultation with the main data users of the DHS, i.e., the National Institute for Public Health and Environmental Protection, the Ministry of Health, Welfare and Sports and the Netherlands Institute for Social Research, eight DHS indicators were selected for which a model-based inference method is developed to produce quarterly figures that are corrected for the loss of CAPI during lockdown periods. These eight indicators are perceived health, fraction of people feeling mentally unhealthy, dental visit, GP consult, specialist consult, daily smoking, excessive alcohol consumption and overweight. These indicators cover the three main topics of the survey (perceived) health, healthcare use and lifestyle.

This paper only shows the results of perceived health, dental visit, daily smoking and excessive alcohol consumption. The results of mentally unhealthy are similar to perceived health and the results of the healthcare use variables GP consult and specialist consult are similar to dental visit. Overweight turns out to be a steady indicator and is hardly affected by the Covid-19 pandemic. *Perceived health* is measured for people of all ages. There are five possible answers: very good, good, fair, poor and very poor. Perceived health is the percentage of people that has given one of the positive answers very good or good. *Dental visit* measures the percentage of people of all ages that has visited a dentist in the past four weeks. *Daily smoking* concerns the percentage of people with a daily smoking habit and is measured for people aged 18 years or older. *Excessive alcohol consumption* is measured for the population aged 18 years or older and measures the percentage of people that report a consumption of 21 or more units per week for men or a consumption of 14 or more units per week for women.

3. Structural time series method

3.1 Univariate models

Two univariate STMs are considered. Let $\hat{y}_t^{\mathrm{A}}$ denote the GREG estimate in quarter t for the unknown population parameter based on all the available response. The first univariate STM ignores the loss of CAPI and starts with a measurement error model that states that the sample estimates is the result of the true population parameter, say ϑ_t, for quarter t and a sampling error, say $\varepsilon_t^{\mathrm{A}}$. This leads to the following measurement error model: $\hat{y}_t^{\mathrm{A}} = \vartheta_t + \varepsilon_t^{\mathrm{A}}$. In a next step the population parameter is modelled with a trend that describes the low frequency variation in the series, say L_t, a seasonal component for seasonal fluctuations, say S_t, and a population white noise for the unexplained variation of the population parameter, say I_t. This implies the following so-called basic STM for the population parameter: $\vartheta_t = L_t + S_t + I_t$. Inserting the STM for the population parameter into the measurement error model gives the first univariate STM:

$$\hat{y}_t^{\mathrm{A}} = L_t + S_t + I_t + \varepsilon_t^{\mathrm{A}} \equiv L_t + S_t + e_t^{\mathrm{A}}. \tag{3.1}$$

Note that in (3.1) the population white noise and sampling error are conveniently combined into one measurement error, i.e., $e_t^{\mathrm{A}} = I_t + \varepsilon_t^{\mathrm{A}}$. The trend L_t is modelled by a smooth trend model (Durbin and Koopman, 2012, Chapter 3), given by

$$\begin{aligned} L_t &= L_{t-1} + R_{t-1} \\ R_t &= R_{t-1} + \eta_t^{\mathrm{R}}, \end{aligned} \tag{3.2}$$

where

$$\eta_t^{\mathrm{R}} \sim N\left(0, f_t \sigma_{\mathrm{R}}^2\right), \mathrm{Cov}\left(\eta_t^{\mathrm{R}}, \eta_{t'}^{\mathrm{R}}\right) = 0, \quad \text{for} \quad t \neq t', \quad \text{and} \quad f_t \geq 1.$$

The trend model consists of a level L_t and a slope R_t with a slope disturbance term η_t^{R}. In a standard smooth trend model, the variance of the slope disturbance terms are time invariant, i.e., $f_t = 1$ for all t. The variance of the slope disturbance terms σ_{R}^2, which are estimated by maximum likelihood (see Subsection 3.4), determines the flexibility of trend model (3.2). For some variables the Covid-19 pandemic causes a sudden strong increase in the quarter-to-quarter changes of the direct estimates. Particularly at the start of the Covid-19 pandemic, the maximum likelihood estimates for σ_{R}^2 are based on the period-to-period changes observed in the past. A sudden increase in the period-to-period changes of the input series therefore results in a temporarily miss-specification of the STM. Or to phrase it differently, for some variables the assumption that the volatility of the period-to-period changes is not affected by the Covid-19 pandemic is violated. To avoid temporal miss-specification of the STM model at the start of the Covid-19 pandemic, the flexibility of the trend model is increased by defining a time-dependent variance for the slope disturbance terms. This is achieved by multiplying the maximum likelihood estimate for σ_{R}^2 with a factor $f_t \geq 1.$ As a result, the variance of the slope disturbance terms is equal to $f_t \sigma_{\mathrm{R}}^2.$ Values for f_t are determined outside the model, as explained in Section 4. This approach is initially proposed by van den Brakel, Souren and Krieg (2022) and is compared with alternative approaches to account for sudden shocks in the input series of an STM due to the Covid-19 pandemic.

Increasing the variance of the slope disturbance terms through factors f_t has the following interpretation. As the variance of the slope disturbance terms increases, the influence of more distant observations on the level of the trend becomes smaller. The proposed approach implies that the filtered estimates attach less weight to the prediction based on observations from the past and more weight to the direct estimates obtained in the last month. This seems reasonable in periods where the world suddenly changes and becomes incomparable with the past, as was the case with the COVID-19 pandemic.

The seasonal component S_t is modelled by a trigonometric seasonal model (Durbin and Koopman, 2012, Chapter 3), given by

$$S_t = \gamma_{1,t} + \cdots + \gamma_{J/2,t}, \tag{3.3}$$

where

$$\gamma_{j,t} = \gamma_{j,t-1} \cos\left(\frac{\pi j}{J/2}\right) + \gamma_{j,t-1}^{*} \sin\left(\frac{\pi j}{J/2}\right) + \omega_{j,t}$$

$$\gamma_{j,t}^{*} = \gamma_{j,t-1}^{*} \cos\left(\frac{\pi j}{J/2}\right) - \gamma_{j,t-1} \sin\left(\frac{\pi j}{J/2}\right) + \omega_{j,t}^{*} \quad \text{for} \quad j = 1, \ldots, J/2.$$

For quarters $J = 4,$ it holds that

$$S_t = \gamma_{1,t} + \gamma_{2,t}, \tag{3.4}$$

with harmonics

$$\gamma_{1,t} = \gamma_{1,t-1}^{*} + \omega_{1,t},$$

$$\gamma_{1,t}^{*} = -\gamma_{1,t-1} + \omega_{1,t}^{*},$$

$$\gamma_{2,t} = -\gamma_{2,t-1} + \omega_{2,t}.$$

Note that the last component defined by (3.3) equals $\gamma_{2,t}^{*} = \gamma_{2,t-1}^{*} + \omega_{2,t}^{*}$ and can be left out since $\gamma_{2,t}^{*}$ is not used in the previous three harmonics and also does not play a role in the measurement equation. The following assumptions for the seasonal disturbance terms,

$$\omega_{1,t} \sim N\left(0, \sigma_{\omega}^2\right), \omega_{1,t}^{*} \sim N\left(0, \sigma_{\omega}^2\right), \omega_{2,t} \sim N\left(0, \sigma_{\omega}^2\right),$$

and

$$\text{Cov}\left(\omega_{j,t}, \omega_{j,t'}\right) = 0, \quad \text{for } t \neq t' \text{ and } j = 1, 2$$

$$\text{Cov}\left(\omega_{1,t}^{*}, \omega_{1,t'}^{*}\right) = 0, \quad \text{for } t \neq t'$$

$$\text{Cov}\left(\omega_{j,t}, \omega_{1,t}^{*}\right) = 0, \quad \text{for all } t \text{ and } j = 1, 2$$

$$\text{Cov}\left(\omega_{1,t}, \omega_{2,t}\right) = 0, \quad \text{for all } t.$$

The Covid-19 pandemic may influence both the trend and the seasonal pattern. Since it is not possible to estimate a structural change in the seasonal pattern due to the Covid-19 pandemic, with less than one year of observations during the Covid-19 pandemic it is assumed that there is only an effect on the development of the trend. The seasonal component S_t is therefore modelled by a trigonometric seasonal model with a time-independent variance. In this way the seasonal pattern is modelled dynamically and therefore has the flexibility to accommodate effects of the Covid-19 pandemic on the seasonal pattern.

To accommodate heteroscedasticity caused by e.g., changes in response size and the sample design, the measurement error e_t^{A} is scaled with the standard error of the input series of $\hat{y}_t^{\mathrm{A}}$ (Binder and Dick, 1990):

$$e_t^{\mathrm{A}} = \sqrt{\hat{\mathrm{V}}\left(\hat{y}_t^{\mathrm{A}}\right)}\, \tilde{e}_t^{\mathrm{A}}, \tag{3.5}$$

$$\tilde{e}_t^{\mathrm{A}} \sim N\left(0, \sigma_{\mathrm{e,A}}^2\right),$$

$$\text{Cov}\left(\tilde{e}_t^{\mathrm{A}}, \tilde{e}_{t'}^{\mathrm{A}}\right) = 0, \quad \text{for } t \neq t',$$

and with $\hat{\mathrm{V}}\left(\hat{y}_t^{\mathrm{A}}\right)$ the variance estimate of $\hat{y}_t^{\mathrm{A}}$. It is understood that $\hat{\mathrm{V}}\left(\hat{y}_t^{\mathrm{A}}\right)$ is estimated outside the STM from the sample data and that these estimates are used as *a priori* known values in the STM. Note that in (3.5) a multiplicative model is chosen for the variance structure of the measurement error. As an alternative an additive structure of the form could be considered. Note that $\hat{\mathrm{V}}\left(\hat{y}_t^{\mathrm{A}}\right)$ in (3.5) is not the real population variance but an estimate of the variance that is subject to uncertainty and can over or under estimate the real variance. The advantage of a multiplicative model is that it scales the variance of the GREG estimator and has the flexibility to reduce the variance if $\hat{\mathrm{V}}\left(\hat{y}_t^{\mathrm{A}}\right)$ over-estimates the real variance. Similar variance structures are used by e.g., Binder and Dick (1990), van den Brakel and Krieg (2015), Elliot and Zong (2019) and Gonçalves, Hidalgo, Silva and van den Brakel (2022).

Model (3.1) borrows strength from the past through both the trend L_t and the seasonal pattern S_t in order to improve the accuracy of the direct estimates. Model (3.1) also accounts for a sudden increase of the volatility of the population parameter by making the trend temporarily more flexible. To account for sudden changes in measurement and selection errors due to the loss of CAPI during the lockdown, model (3.1) is extended with an intervention variable. This gives rise to the second univariate model:

$$\hat{y}_t^{\mathrm{A}} = L_t + S_t + \beta \frac{x_t}{3} + e_t^{\mathrm{A}}. \tag{3.6}$$

Here x_t is the number of months in quarter t without CAPI response and β a regression coefficient that can be interpreted as the net effect of the change in measurement and selection bias due to the loss of CAPI. In a quarter with full CAPI response, $x_t/3=0$ and β is switched off. In a quarter without any CAPI respondents, $x_t/3=1$ and β absorbs the effect of the loss of CAPI and avoids that the model estimates for the population parameter ϑ_t are affected, at least partially. If a quarter only contains one or two months without CAPI, then $x_t/3=1/3$ or $x_t/3=2/3$ respectively and the correction of β contributes proportionally to the number of months without CAPI in that quarter. The trend, seasonal component and measurement error are defined in (3.2), (3.3), and (3.5), respectively.

Compared to model (3.1) it is expected that model (3.6) better accommodates for the loss of CAPI during the lockdown. Model (3.6), however, assumes no structural change in the evolution of the population parameter ϑ_t. If the lockdown results in e.g., strong turning points in the population parameter, it can be expected that this is partially and incorrectly absorbed in the regression coefficient of the intervention variable. To accommodate for this risk, the bivariate model, proposed in the next section is developed.

3.2 Bivariate model

The input series for the bivariate model are the quarterly direct estimates based on the complete response, denoted $\hat{y}_t^{\mathrm{C}}$ (*complete series*) and the quarterly direct estimates based on the web response only, denoted $\hat{y}_t^{\mathrm{W}}$ (*web series*). The systematic difference between both series observed during the years before the start of the Covid-19 pandemic is used in a bivariate STM to make model-based estimates for the population parameter that correct for the loss of CAPI during the lockdown. The bivariate STM given by:

$$\begin{pmatrix} \hat{y}_t^{\mathrm{C}} \\ \hat{y}_t^{\mathrm{W}} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} (L_t + S_t) + \begin{pmatrix} 0 \\ \lambda_t \end{pmatrix} + \begin{pmatrix} e_t^{\mathrm{C}} \\ e_t^{\mathrm{W}} \end{pmatrix}. \tag{3.7}$$

The first component states that $\hat{y}_t^{\mathrm{C}}$ and $\hat{y}_t^{\mathrm{W}}$ are two estimates for the unknown population parameter that is decomposed in a trend and a seasonal component. The population irregular term I_t is combined with the sampling errors, similar to the univariate models. The trend L_t is modelled by the smooth trend model were the variance of the slope disturbance terms is made time varying, as defined by equation (3.2) and the seasonal component S_t by the trigonometric model given by equation (3.4). The second component of (3.7), i.e., λ_t, models the systematic difference between the regular series and the web series as a random walk, given by

$$\lambda_t = \lambda_{t-1} + \eta_{\lambda,t}, \tag{3.8}$$

where

$$\eta_{\lambda,t} \sim N\left(0, \sigma_\lambda^2\right)$$

$$\mathrm{Cov}\left(\eta_{\lambda,t}, \eta_{\lambda,t'}\right) = 0, \quad \text{for } t \neq t'.$$

Because a random walk is assumed, the model accommodates gradual changing differences between $\hat{y}_t^{\mathrm{C}}$ and $\hat{y}_t^{\mathrm{W}}$. The third component of (3.7) contains the measurement error. They contain the sampling error of $\hat{y}_t^k$ and the population irregular term, i.e., $e_t^k = I_t + \varepsilon_t^k$ for $k \in \{\mathrm{C}, \mathrm{W}\}$. The measurement error component accommodates heteroscedasticity by scaling the measurement error with the sampling error of the input series and accounts for the positive correlation between $\hat{y}_t^{\mathrm{C}}$ and $\hat{y}_t^{\mathrm{W}}$ that arises because both estimates use the same web respondents. This is achieved with the following measurement error model:

$$e_t^k = \sqrt{\hat{\mathrm{V}}\left(\hat{y}_t^k\right)}\, \tilde{e}_t^k, \quad \text{with } \hat{\mathrm{V}}\left(\hat{y}_t^k\right) \text{ the variance estimate of } \hat{y}_t^k \tag{3.9}$$

and

$$\tilde{e}_t^k \sim N\left(0, \sigma_{\mathrm{e},k}^2\right)$$

$$\mathrm{Cov}\left(\hat{y}_t^{\mathrm{C}}, \hat{y}_t^{\mathrm{W}}\right) = \frac{\sqrt{n_t^{\mathrm{W}}}}{\sqrt{n_t^{\mathrm{C}}}} \sqrt{\hat{\mathrm{V}}\left(\hat{y}_t^{\mathrm{C}}\right)} \sqrt{\hat{\mathrm{V}}\left(\hat{y}_t^{\mathrm{W}}\right)}$$

$$\mathrm{Cov}\left(\tilde{e}_t^k, \tilde{e}_{t'}^k\right) = 0, \quad \text{for } t \neq t'.$$

The covariance between the measurement errors is obtained as follows. Following Kish (1965), the correlation between two variables observed in two partial overlapping samples is given by

$$\mathrm{Cor}\left(z_1, z_2\right) = \rho \frac{n_{1\cap 2}}{\sqrt{n_1}\sqrt{n_2}},$$

where

- z_1 the variable observed in sample s_1 of size n_1,
- z_2 the variable observed in sample s_2 of size n_2,
- $n_{1\cap 2}$ the size of the sample overlap between s_1 and s_2,
- ρ the correlation between z_1 and z_2 based on the $n_{1\cap 2}$ respondents that are included in s_1 and s_2.

In this application, sample s_1 is the sample with complete response and s_2 the sample with CAWI respondents. Suppose that $z_1 = \hat{y}_t^{\mathrm{C}}$, $z_2 = \hat{y}_t^{\mathrm{W}}$, $n_1 = n_t^{\mathrm{C}}$, and $n_2 = n_t^{\mathrm{W}}$, with n_t^{C} is the size of the complete response in quarter t and n_t^{W} the size of the web response in quarter t. In this case the sample overlap is also the sample with CAWI respondents. Therefore we have $n_{1\cap 2} = n_t^{\mathrm{W}}$ and $\rho = 1$. From this it follows that

$$\mathrm{Cor}\left(\hat{y}_t^{\mathrm{C}}, \hat{y}_t^{\mathrm{W}}\right) = \rho \frac{n_t^{\mathrm{W}}}{\sqrt{n_t^{\mathrm{C}}}\sqrt{n_t^{\mathrm{W}}}} = \frac{\sqrt{n_t^{\mathrm{W}}}}{\sqrt{n_t^{\mathrm{C}}}}$$

and

$$\mathrm{Cov}\left(\hat{y}_t^{\mathrm{C}}, \hat{y}_t^{\mathrm{W}}\right) = \frac{\sqrt{n_t^{\mathrm{W}}}}{\sqrt{n_t^{\mathrm{C}}}} \sqrt{\hat{\mathrm{V}}\left(\hat{y}_t^{\mathrm{C}}\right)} \sqrt{\hat{\mathrm{V}}\left(\hat{y}_t^{\mathrm{W}}\right)}.$$

As a result, the covariance matrix for the measurement errors in (3.7) is given by

$$\begin{pmatrix} e_t^{\mathrm{C}} \\ e_t^{\mathrm{W}} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \hat{\mathrm{V}}\left(\hat{y}_t^{\mathrm{C}}\right)\sigma_{\mathrm{e,C}}^2 & \mathrm{Cov}\left(\hat{y}_t^{\mathrm{C}}, \hat{y}_t^{\mathrm{W}}\right) \\ \mathrm{Cov}\left(\hat{y}_t^{\mathrm{C}}, \hat{y}_t^{\mathrm{W}}\right) & \hat{\mathrm{V}}\left(\hat{y}_t^{\mathrm{W}}\right)\sigma_{\mathrm{e,W}}^2 \end{pmatrix}\right]. \tag{3.10}$$

Similar to the univariate models, $\hat{\mathrm{V}}\left(\hat{y}_t^{\mathrm{C}}\right)$, $\hat{\mathrm{V}}\left(\hat{y}_t^{\mathrm{W}}\right)$, and $\mathrm{Cov}\left(\hat{y}_t^{\mathrm{C}}, \hat{y}_t^{\mathrm{W}}\right)$ are estimated outside the STM from the sample data. These estimates are used as *a priori* known values in STM (3.7).

During the lockdown, $\hat{y}_t^{\mathrm{C}}$ is missing but $\hat{y}_t^{\mathrm{W}}$ is observed. With bivariate STM (3.7) it is possible to obtain estimates for the trend (L_t) and the signal $(L_t + S_t)$ of the population parameters of interest. These estimates are corrected for the bias due to the loss of CAPI, because the model accounts for the systematic difference between $\hat{y}_t^{\mathrm{C}}$ and $\hat{y}_t^{\mathrm{W}}$ through the second model component (λ_t). This correction relies on the assumption that the systematic difference between $\hat{y}_t^{\mathrm{C}}$ and $\hat{y}_t^{\mathrm{W}}$ as observed before the start of the Covid-19 pandemic does not change during the lockdown.

3.3 Direct estimates for time series models

For the DHS direct quarterly estimates can be computed starting in the first quarter of 2014. From the first quarter of 2014 up to the last quarter of 2019 these direct estimates are based on the weighted annual DHS response obtained by applying the GREG estimator. Quarterly estimates $\hat{y}_t^{\mathrm{C}}$ for the complete series are obtained by computing the domain estimator based on the GREG estimator with quarter t as domain. Quarterly estimates $\hat{y}_t^{\mathrm{W}}$ are obtainedby recalculating the GREG estimator using the CAWI response only and subsequently computing the domain estimator based on the GREG estimator with quarter t as the domain. In the quarters before 2020 there was no loss of CAPI and the direct estimates $\hat{y}_t^{\mathrm{A}}$ are equal to $\hat{y}_t^{\mathrm{C}}$. Standard errors are computed in R (R Core Team, 2015) with the package "survey" (Lumley, 2014). For the estimation of the standard errors the sample design of the DHS is taken into account, where the stratification is based on the cross-classification of months and provinces. Here provinces are used, because the subdivision into municipalities leads to strata with too little response.

Since the decision to publish quarterly figures was made in June 2020, the direct estimates for the first two quarters of 2020 are based on the weighted response based on the GREG estimator available from January to June 2020. Estimates for the third quarter of 2020 are based on the weighted response available from January to September 2020 and the fourth quarter is based on the weighted annual response of 2020. For all quarters the same weighting model and the same population totals of the covariates are used. Direct estimates for the first quarter of 2021 are computed in a similar way and are based on the response from

January to March 2021. Estimates for the second quarter of 2021 are based on the response from January to June 2021. In this way the estimates $\hat{y}_t^{\mathrm{W}}$ for the web series are obtained for all quarters of 2020 and for the first two quarters of 2021.

Adding quarterly samples during an ongoing year results in a progressively larger annual data set. The main advantage of this approach is that all available data are used for the weighting scheme of the GREG estimator. Note that this will only slightly increases the heterogeneity between the quarterly direct estimates, since the variance of the quarterly direct estimates is of the order of the quarterly sample size, not the total sample size. There might be a minor effect since the fluctuation of the GREG weights decreases if the sample used for weighting increases. This is, however, not an issue since the variance of the measurement errors is taken proportional to the variance of the GREG estimates used in the input series, as can be seen from formula (3.9) and (3.10). This approach also does not create additional dependency between the quarterly estimates, since there is no sample overlap between the quarterly estimates and the variance of the GREG estimates are based on the GREG residuals, which are assumed to be independent.

For the complete series $\hat{y}_t^{\mathrm{C}}$ in 2020 the second quarter is missing and the other quarters are based on response where CAPI is partially missing (Table 2.1). In the first quarter of 2020 CAPI is only missing in the last two weeks of March and for this quarter it is assumed that sufficient CAPI response is available to obtain plausible estimates. So in the first quarter of 2020 the estimates $\hat{y}_t^{\mathrm{C}} = \hat{y}_t^{\mathrm{A}}$ are based on the available CAWI and CAPI response and in the second quarter of 2020 $\hat{y}_t^{\mathrm{C}}$ is missing and $\hat{y}_t^{\mathrm{A}} = \hat{y}_t^{\mathrm{W}}$. In the third quarter of 2020 CAPI response is only available in August and September. Here a correction is applied to $\hat{y}_t^{\mathrm{C}}$ based on the bivariate model (3.7). The direct estimate $\hat{y}_t^{\mathrm{C}}$ for the third quarter of 2020 is obtained by computing the domain estimator of the GREG applied to the available response in August and September minus $1/3$ of the difference $\hat{\lambda}_t$ estimated by model (3.7) in the second quarter. No correction is applied to the corresponding standard errors. The direct estimate $\hat{y}_t^{\mathrm{A}}$ in the third quarter is equal to the uncorrected weighted mean of the available response in August and September. In the fourth quarter of 2020 CAPI is also missing for only two weeks and it is assumed that there is enough CAPI response available to obtain plausible estimates, so $\hat{y}_t^{\mathrm{C}} = \hat{y}_t^{\mathrm{A}}$.

In 2021 there is besides CAPI and CAWI also CAPI/CAWI response (Section 2). To find out how to use the CAPI/CAWI response in the best possible way, two scenarios were elaborated. In the first scenario quarterly figures are computed where CAPI/CAWI response is considered as CAPI and in the second scenario CAPI/CAWI response is considered as CAWI. Since there were no major differences in the results of both scenarios the CAPI/CAWI response is considered as CAWI. Results of this comparison are not shown in this paper. In the first quarter of 2021 $\hat{y}_t^{\mathrm{C}}$ is missing and $\hat{y}_t^{\mathrm{A}} = \hat{y}_t^{\mathrm{W}}$ and in the second quarter of 2021 CAPI is available and so $\hat{y}_t^{\mathrm{C}} = \hat{y}_t^{\mathrm{A}}$.

In this way input series for models (3.1), (3.6) and (3.7) are obtained. The series run from the first quarter in 2014 up to the second quarter in 2021. The series $\hat{y}_t^{\mathrm{A}}$ and $\hat{y}_t^{\mathrm{W}}$ are available for all quarters and for the series $\hat{y}_t^{\mathrm{C}}$ estimates are missing in the second quarter of 2020 and in the first quarter of 2021.

3.4 Model-based estimates

Given the series of direct estimates $\hat{y}_t^{\mathrm{A}}$, $\hat{y}_t^{\mathrm{C}}$ and $\hat{y}_t^{\mathrm{W}}$, model-based estimates based on one of the models (3.1), (3.6) or (3.7) can be produced. To this end the three models are expressed in state space representation, where after the Kalman filter is applied to obtain optimal estimates for the state variables, i.e., the variables that define the trend (L_t, R_t), the seasonal component $(\gamma_{1,t}, \gamma_{1,t}^{*}, \gamma_{2,t})$, and the bias parameter (λ_t). The Kalman filter assumes that values for the hyperparameters, i.e., the variances of the measurement errors and state disturbance terms $(\sigma_{\mathrm{R}}^2, \sigma_{\omega}^2, \sigma_{\lambda}^2, \sigma_{\mathrm{e,A}}^2, \sigma_{\mathrm{e,C}}^2, \sigma_{\mathrm{e,W}}^2)$, are known. Estimates for these hyperparameters are obtained with maximum likelihood. To this end a likelihood function, obtained by the one-step-ahead error decomposition, is maximized using numerical optimization algorithm MaxBFGS. The Kalman filter is a recursive algorithm that runs from $t=1$ to the last observation of the series and gives optimal estimates with their standard errors for the state variables and the signal for each period t based on the observed series until period t. These are the so-called filtered estimates. The filtered estimates of past state vectors can be updated if new data become available. This procedure is referred to as smoothing and results in smoothed estimates that are based on the completely observed time series. In this application, interest is mainly focused on the filtered estimates, since they are based on the complete set of information that would be available in the regular production process to produce a model-based estimate for quarter t. The state variables in the Kalman filter are initialized with a diffuse initialization, which means that the starting values for the state variables are equal to zero with a very large standard error. After a few iterations, the filtered estimates for the states converge to a proper distribution. For this reason the filtered estimates for the states of the first d periods of the series are ignored in the analysis, where d equals the number of state variables with a diffuse initialization. See Durbin and Koopman (2012) for more details of the state space representation of the STMs, the Kalman filter and the maximum likelihood estimation procedure for the hyperparameters. The computations are conducted with Ssfpack 3.0 (Koopman, Shephard and Doornik, 2008) in combination with Ox (Doornik, 2009).

The Kalman filter provides optimal estimates for the state variables. For this application the trend (L_t) and the signal $(L_t + S_t)$ of the population parameter are of particular interest, since these are the variables that are published as official quarterly health indicators. Standard errors of these estimates are obtained from the Kalman filter recursion. These standard errors do not account for the additional uncertainty that arises since the values of the hyperparameters are replaced by their maximum likelihood estimates in the Kalman filter recursions. This is the standard approach in state space applications, but it will result in over-optimistic estimates for the standard errors. Note that Pfeffermann and Tiller (2005) propose a bootstrap that accounts for the additional uncertainty of the maximum likelihood estimates of the hyperparameters in the Kalman filter.

Model selection is based on likelihood-based model diagnostics such as the AIC and BIC (Durbin and Koopman, (2012, Chapter 7)). The normality assumptions of the state disturbance terms in the STMs presented in Subsections 3.1 and 3.2 imply that the standardized innovations or one-step-ahead predictions are standard normally distributed. For all three models it is evaluated whether they meet these underlying assumptions by testing to which extent the standardized innovations are standard normally and independently distributed. This is done by testing the standardized innovations on normality using Bowman-Shenton normality test, drawing QQ-plots and histograms of the standardized innovations. Sample autocorrelograms and the Durbin Watson test are applied to test for serial correlation in the standardized innovations. An F-test for heteroscedasticity is applied to test for equal variance of the standardized innovations. Finally, time series plots of the standardized innovations are drawn to check for outliers. For more details on these tests it is referred to Durbin and Koopman (2012, Chapter 2). These model diagnostics indicate that the underlying model assumptions of the finally selected models are not seriously violated.

In quarters where CAPI is missing, additional assumptions for the three STMs are required. For the univariate STM (3.1) it is assumed that there are no mode effects between CAPI and CAWI. For the univariate STM (3.6) it is assumed that the trend and the seasonal component correctly describe the evolution of the population parameter and that sudden strong changes in the true values of the population parameter, such as turning points, are not partially absorbed in the level intervention component. These assumptions are evaluated in Section 4. For the bivariate STM (3.7) it is assumed that the difference between CAWI and CAPI response does not change due to the Covid-19 pandemic. This implies that the composition of the web response does not change during the Covid-19 pandemic. It is not possible to verify whether is assumption is met. A response analysis showed that no structural change in the CAWI response and non-response distributions before and after the start of the corona crisis is observed. There were also no structural difference between the answer categories under the CAPI and the CAWI response before the first lockdown and the third and fourth quarter of 2020 where CAPI was started up again. See also the results for the bias parameter λ_t in the bottom-right panels of Figures 4.5-4.8 in Section 4.

4. Results time series models

The three models are fitted to the series of direct estimates as described in Subsection 3.3. Due to the Covid-19 pandemic some DHS variables show a strong increase in the quarter-to-quarter changes, especially at the beginning of the two lockdown periods. In these periods, the smooth trend model is not flexible enough to follow the increased period-to-period movements of the input series. This can be expected since the flexibility of the trend, which is determined by the variance of the slope disturbance terms of the trend model, is based on the quarter-to-quarter movements observed in the period before the Covid-19 crisis. A sudden increase in the dynamics of the population parameter results in temporary miss-specification of the STM, which becomes visible in large values for the standardized innovations in these periods. To accommodate in the STM for the suddenly increased volatility of the population parameters, the flexibility of the smooth trend is temporarily increased by multiplying the variance of the slope disturbance terms $\left(\sigma_{\mathrm{R}}^2\right)$ in (3.2) by a time-dependent factor $f_t \geq 1,$ as explained in Subsection 3.1.

The values for f_t are chosen in such a way that the standardized innovations in the period during the start of the Covid-19 pandemic have values within or just outside the admissible range of 1.96 in absolute terms. In this way, the value of the factor $f_t > 1$ is kept as small as possible, so that the model can still borrow strength from the past. Note that adjusting the variance σ_{R}^2 in quarter t influences the slope disturbance term from quarter $t+1$ and the trend only from quarter $t+2.$ So there is a lag of two quarters in the effect of the outcomes after adjustment of $\sigma_{\mathrm{R}}^2.$ Thus to increase the flexibility of the slope in Q2 of 2020, the value of σ_{R}^2 must be increased at the latest in Q4 of 2019. For several variables it was necessary to increase the variance already in Q3 2019. To avoid a large sudden change in the variance of the slope disturbance terms, the values of f_t are slightly increased in the quarters preceding Q3 2019. In the quarters after the first lockdown in Q2 2020, the values of f_t are reduced to 1 as soon as possible.

From the analysis of the standardized innovations it follows that for most variables it is necessary to make the slope more flexible during the pandemic. Table 4.1 shows the values of the factors $f_t > 1$ for models (3.1), (3.6) and (3.7). In quarters where $f_t = 1$ no values are shown. Variables for which it was not necessary to make the slope more flexible are not shown in the tables either. For a correct interpretation, the values for f_t must be compared with the maximum likelihood estimates for σ_{R}^2 in Table 4.2. For perceived health and dental visit a flexible slope is applied in the quarters before the first lockdown, i.e., the second quarter of 2020. For daily smoking, $f_t > 1$ only for the univariate STM without intervention (3.1) and only before the second lockdown. For excessive alcohol assumption it is not necessary to make the trend more flexible. The factors in Table 4.1 are relatively large compared to the values $\hat{\sigma}_{\mathrm{R}}$ of in Table 4.2. Because the variances of the slope disturbance terms are generally small, large values for f_t are required to give the trend component sufficient flexibility to follow the strong period-to-period changes at the start of the corona crisis. Note this is an empirical result that differs between applications.

Table 4.1
Values of flexibility parameter f_t in quarters where $f_t > 1.$ In quarters and for variables where no value is displayed, $f_t = 1.$ In the first two quarters of 2021 and in the quarters before the third quarter of 2018, $f_t = 1$ for all variables

		2018 Q4	2019 Q1	2019 Q2	2019 Q3	2019 Q4	2020 Q1	2020 Q2	2020 Q3	2020 Q4
Univariate STM without intervention	Perceived health			10	100	100	100	10		
	Dental visit				10	100	100	10		
	Daily smoking								10	50
Univariate STM with intervention	Perceived health		10	100	200	100	100	10		
	Dental visit	10	100	1,000	5,000	8,000	100	10		
Bivariate STM	Perceived health			10	100	100	100	10		
	Dental visit				10	100	100	10		

Note: Structural time series model (STM).

Figures 4.1-4.4 display the standardized innovations for perceived health estimated by the three models (3.1), (3.6) and (3.7). For all series the innovations, estimated by the model where the variance of the slope disturbance terms is not temporarily increased (black dashed line), exceed the interval of (-1.96, 1.96) implying that the model is miss-specified at the start of the first lockdown. By making the slope more flexible the standardized innovations (red solid line) get admissible values. The standardized innovations for the other variables are not shown here. After setting the values for f_t, the underlying model assumptions are evaluated by testing whether the standardized innovations are standard normally and independently distributed. For all three models the performed tests (Section 3.4) show some small violations of these assumptions for some of the variables. Alternative model formulations did not further improve the model diagnostics.

Figure 4.1 Standardized innovations for perceived health estimated by univariate STM without intervention (3.1).

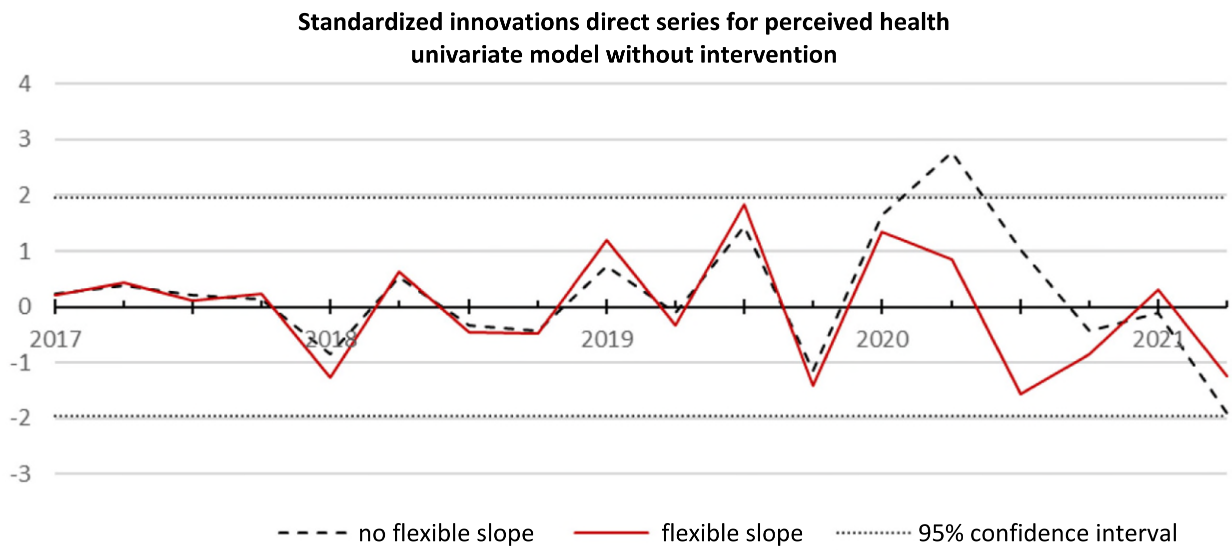


Figure 4.2 Standardized innovations for perceived health estimated by univariate STM with intervention (3.6).

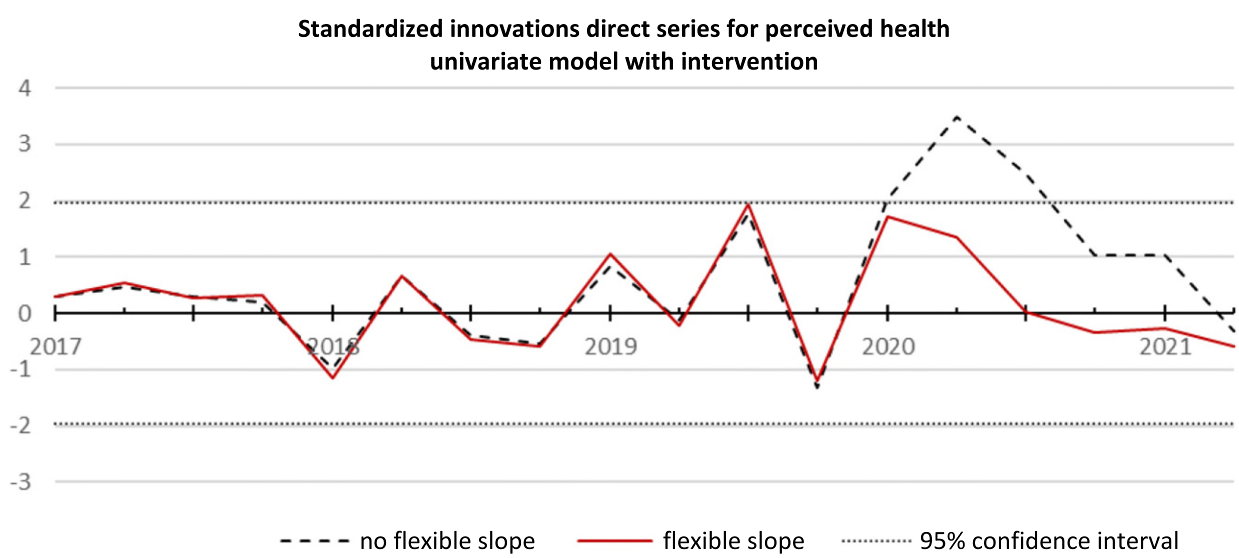


Figure 4.3 Standardized innovations complete series for perceived health estimated by bivariate STM, given by (3.7).

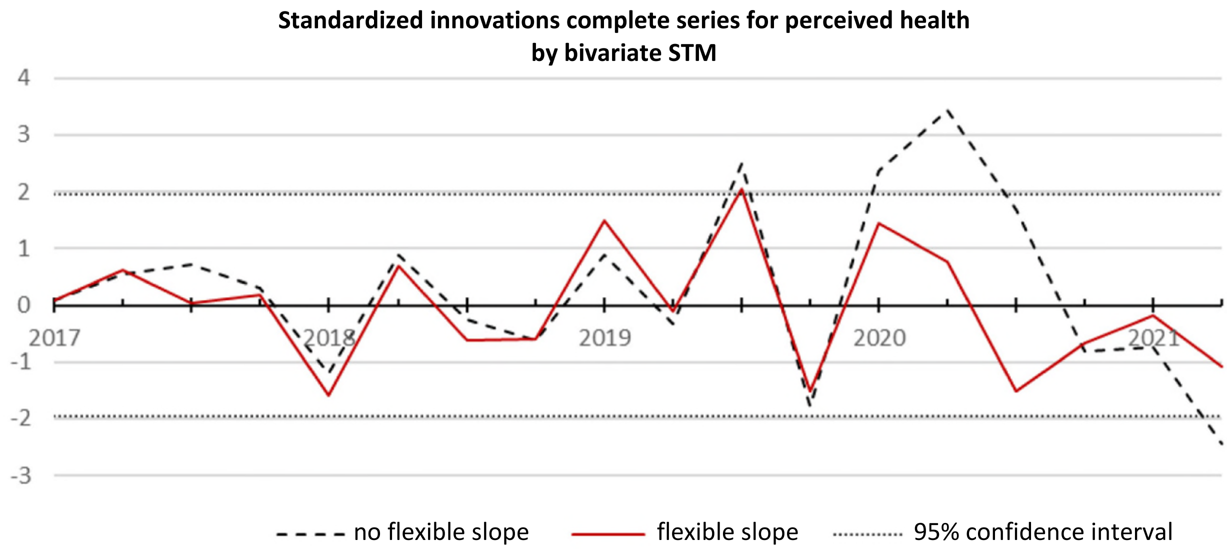


Figure 4.4 Standardized innovations web series for perceived health estimated by bivariate STM, given by (3.7).

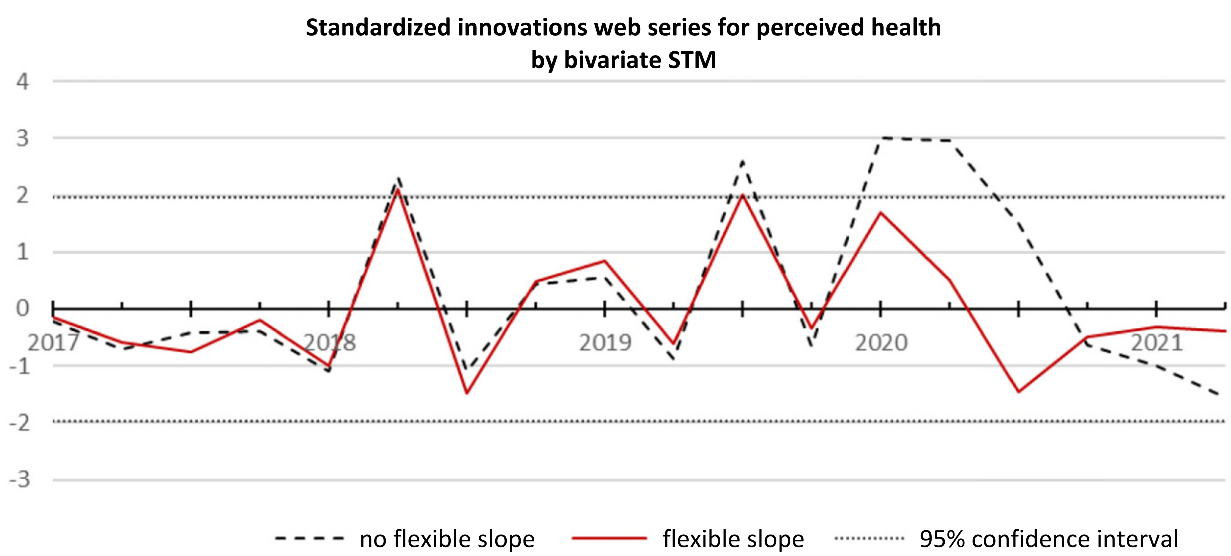


Table 4.2 gives the real-time or concurrent maximum likelihood estimates of the hyperparameters of the three STMs. This means that the maximum likelihood estimates are based on the series observed until the particular quarter in the table. In order to show the values of the hyperparameters before the pandemic, the estimates are also displayed for the second quarter of 2019. Even though the variance σ_{R}^2 is multiplied by a factor f_t in the model, it can be seen that in many cases the (square root of the) variance estimate $\hat{\sigma}_{\mathrm{R}}$

increases. The largest increases occur for dental visit before the first lockdown. For daily smoking the variance estimate $\hat{\sigma}_{\mathrm{R}}$ increases in the second quarter of 2021 for the univariate model.

The estimates of some of the variance components in Table 4.2 are very small. This is the case with $\hat{\sigma}_{\mathrm{R}}$ for perceived health, daily smoking and excessive alcohol consumption and $\hat{\sigma}_{\omega}$ for daily smoking. These hyperparameters could, on the one hand, be removed from the model and it can therefore be assumed that the trend and seasonal components are time invariant. The slope disturbance terms, however, cannot be removed from the model because the flexibility of the trends needed to be increased during the corona crisis by increasing the variance of the slope disturbance terms. Also the variance of the seasonal disturbance terms are kept to make the models more robust for changes in the seasonal pattern during the corona crisis. In a similar way the $\hat{\sigma}_{\lambda}$ for dental visits could be set to zero, but that would make the assumption that the difference between CAPI and CAWI after the start of the corona crisis did not change even stronger.

Figures 4.5-4.8 show the results of the estimates for the variables under the three models. The displayed series start in the first quarter of 2017. Since a diffuse initialisation of the Kalman filter is applied, the model predictions for the first three years obtained with the STM are ignored. For all variables, four graphs are displayed. The first one compares the direct estimates $\hat{y}_t^{\mathrm{C}}$ (*dir compl*) and $\hat{y}_t^{\mathrm{W}}$ (*dir web*) with the model-based estimates $\hat{L}_t + \hat{S}_t$ based on the bivariate STM (*STM biv*), the univariate model without intervention (*STM univ*) and the univariate model with intervention (*STM univ with int*). The second graph shows the estimated standard errors of the quarterly estimates of the point estimates presented in the first graph. The graphs in the bottom-left panel shows the intervention coefficient β of the univariate model (*intervention STM univ*) of STM (3.6). The graph in the bottom-right panel shows the systematic difference λ_t (syst. diff. web and compl. resp.) of STM (3.7) together with the 95% confidence intervals.

By comparing the series of the direct estimates based on the complete response and the web response and by analysing the estimates of the systematic difference (λ_t) it follows for most variables that there is a clear mode effect between the CAPI and CAWI response. This is picked up by the λ_t parameter of the bivariate model. For perceived health the differences between the series with and without CAPI are relatively small (Figure 4.5, top panel). For dental visit, CAWI respondents score higher than CAPI respondents and the systematic difference λ_t varies between 1.5% and 2% (Figure 4.6, top panel). For daily smoking and excessive alcohol consumption it is just the other way around (Figures 4.7 and 4.8, top panel). For these variables CAPI scores are higher than CAWI and for daily smoking the difference is the largest with a systematic difference, measured by λ_t, of around -4%. This illustrates that ignoring the effect of the loss of CAPI during the lockdown, results in a substantial bias in the direct estimates. Combining direct quarterly estimates that are based on CAWI only for the lockdown periods with estimates based the complete response obtained in forgoing or preceding periods of the lockdown in one time series, would result in misleading period-to-period changes during the Covid-19 period. See e.g., the top panel of Figure 4.7 for daily smoking.

Table 4.2
Concurrent maximum likelihood estimates hyperparameters STM

	Perceived health						**Daily smoking**					
	2019 Q2	**2020 Q2**	**2020 Q3**	**2020 Q4**	**2021 Q1**	**2021 Q2**	**2019 Q2**	**2020 Q2**	**2020 Q3**	**2020 Q4**	**2021 Q1**	**2021 Q2**
	Univariate STM without intervention						Univariate STM without intervention					
$\hat{\sigma}_{\mathrm{R}}$	<0.001	0.001	<0.001	0.001	0.001	0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.002
$\hat{\sigma}_{\omega}$	<0.001	0.002	0.001	0.002	0.002	0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
$\hat{\sigma}_{\mathrm{e,\,A}}$	0.007	0.007	0.008	0.007	0.007	0.007	0.010	0.011	0.010	0.01	0.010	0.010
	Univariate STM with intervention						Univariate STM with intervention					
$\hat{\sigma}_{\mathrm{R}}$	<0.001	<0.001	<0.001	<0.001	0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
$\hat{\sigma}_{\omega}$	<0.001	0.002	0.002	0.002	0.002	0.001	<0.001	<0.001	<0.001	0.001	<0.001	0.002
$\hat{\sigma}_{\mathrm{e,\,A}}$	0.007	0.006	0.006	0.007	0.008	0.008	0.010	0.009	0.010	0.009	0.009	0.009
	Bivariate STM						Bivariate STM					
$\hat{\sigma}_{\mathrm{R}}$	<0.001	0.001	<0.001	<0.001	<0.001	0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
$\hat{\sigma}_{\omega}$	<0.001	<0.001	0.002	0.002	0.002	0.002	<0.001	<0.001	<0.001	0.001	0.003	0.003
$\hat{\sigma}_{\lambda}$	<0.001	<0.001	0.002	<0.001	<0.001	0.004	0.002	0.003	0.007	0.01	0.002	0.002
$\hat{\sigma}_{\mathrm{e,\,C}}$	0.957	1.120	0.979	0.934	0.928	0.862	1.310	1.220	1.200	1.250	1.710	1.710
$\hat{\sigma}_{\mathrm{e,\,W}}$	1.310	1.340	1.180	1.240	1.240	1.040	1.280	1.210	1.100	0.653	0.718	0.760
	Dental visit						**Excessive alcohol consumption**					
	Univariate STM without intervention						Univariate STM without intervention					
$\hat{\sigma}_{\mathrm{R}}$	<0.001	0.003	0.005	<0.001	0.002	0.003	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
$\hat{\sigma}_{\omega}$	<0.001	<0.001	<0.001	0.004	0.003	0.002	0.002	0.001	0.001	0.001	0.001	<0.001
$\hat{\sigma}_{\mathrm{e,\,A}}$	0.009	0.008	0.008	0.011	0.009	0.010	0.008	0.010	0.009	0.009	0.009	0.010
	Univariate STM with intervention						Univariate STM with intervention					
$\hat{\sigma}_{\mathrm{R}}$	<0.001	0.003	0.005	<0.001	0.002	0.003	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
$\hat{\sigma}_{\omega}$	<0.001	<0.001	<0.001	0.004	0.003	0.002	0.002	<0.001	<0.001	0.001	0.001	<0.001
$\hat{\sigma}_{\mathrm{e,\,A}}$	0.009	0.008	0.008	0.011	0.009	0.010	0.008	0.010	0.010	0.010	0.009	0.009
	Bivariate STM						Bivariate STM					
$\hat{\sigma}_{\mathrm{R}}$	<0.001	0.003	0.006	0.004	0.003	0.006	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
$\hat{\sigma}_{\omega}$	<0.001	<0.001	<0.001	0.004	0.004	0.001	0.002	0.004	<0.001	<0.001	<0.001	<0.001
$\hat{\sigma}_{\lambda}$	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.003	0.002	0.002	0.002	0.003	0.002
$\hat{\sigma}_{\mathrm{e,\,C}}$	1.080	1.130	1.100	0.713	0.624	0.954	1.360	1.460	1.450	1.480	1.490	1.470
$\hat{\sigma}_{\mathrm{e,\,W}}$	1.070	1.080	1.140	1.050	1.010	1.190	0.822	1.070	1.050	1.040	1.040	1.050

Note: Structural time series model (STM).

Figure 4.5 Results STM for perceived health.

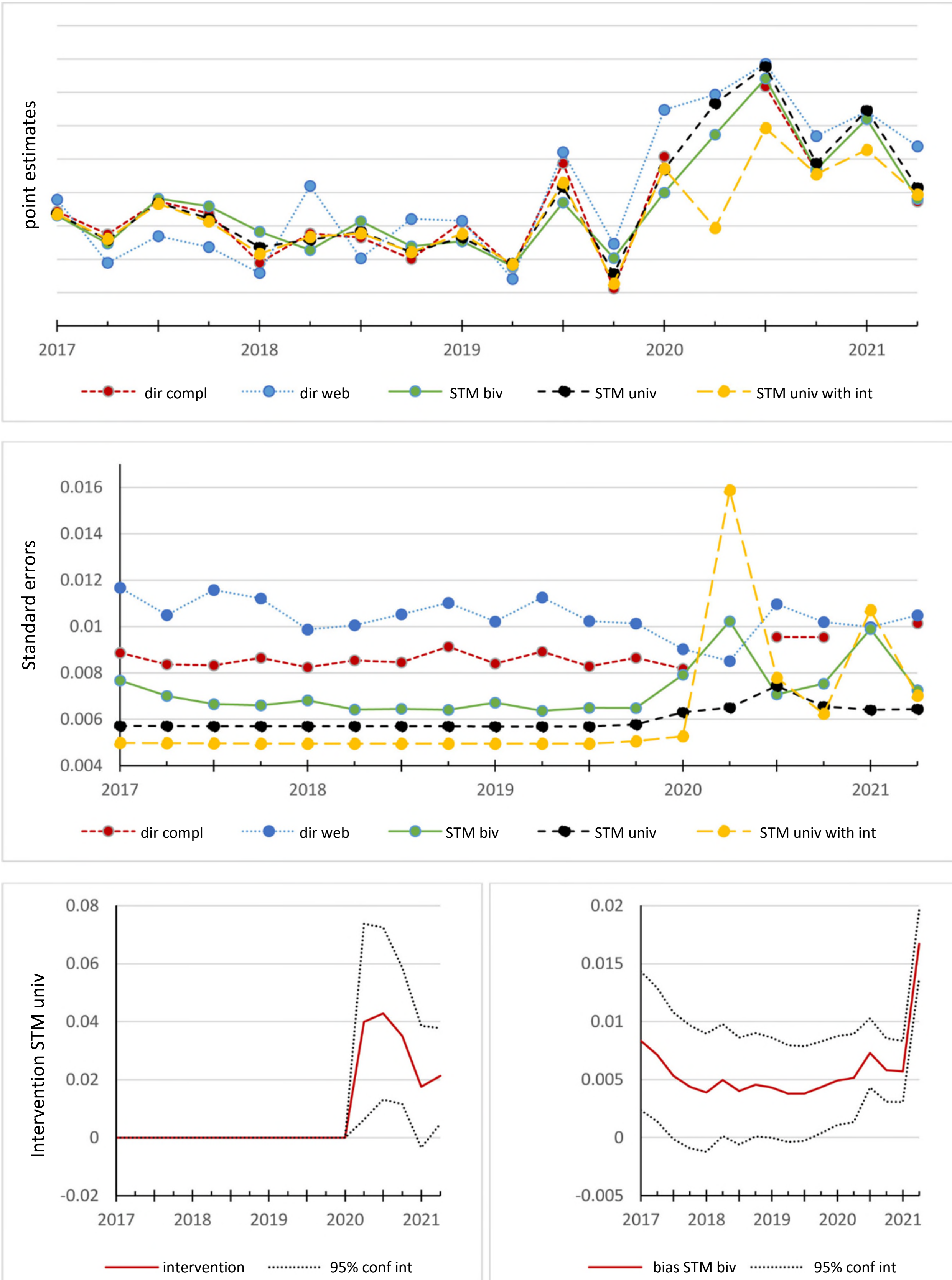


Figure 4.6 Results STM for dental visit.

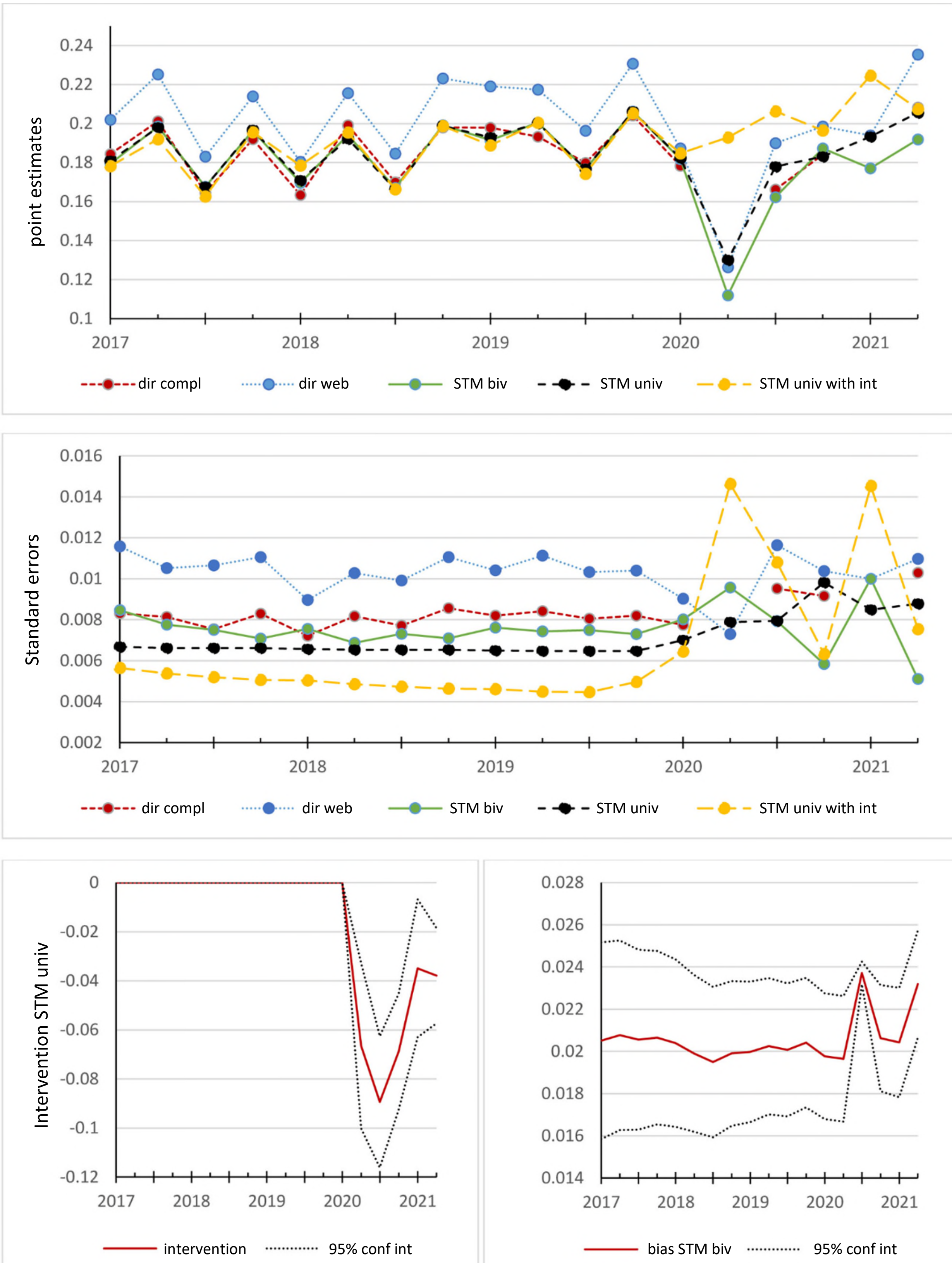


Figure 4.7 Results STM for daily smoking.

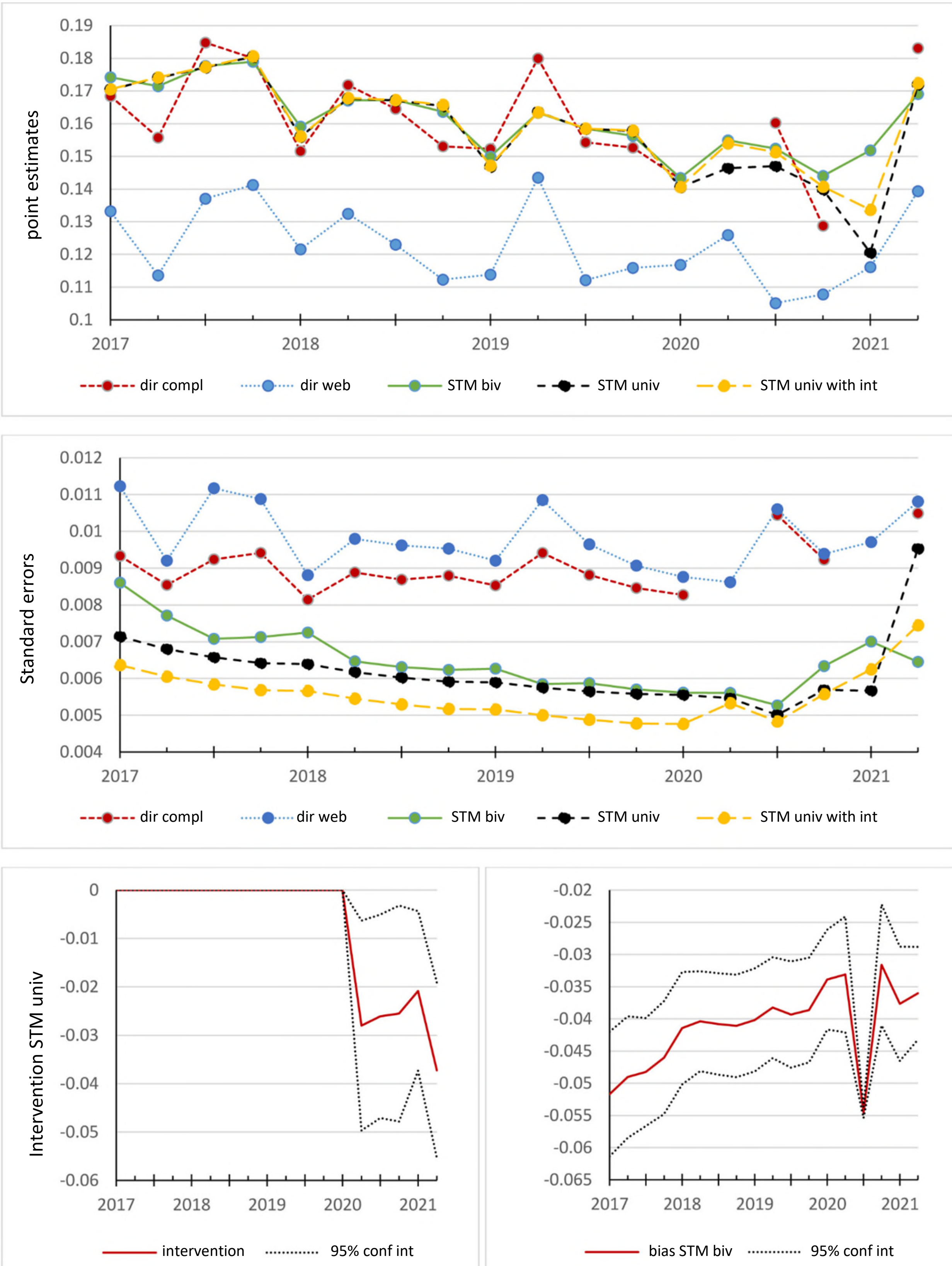
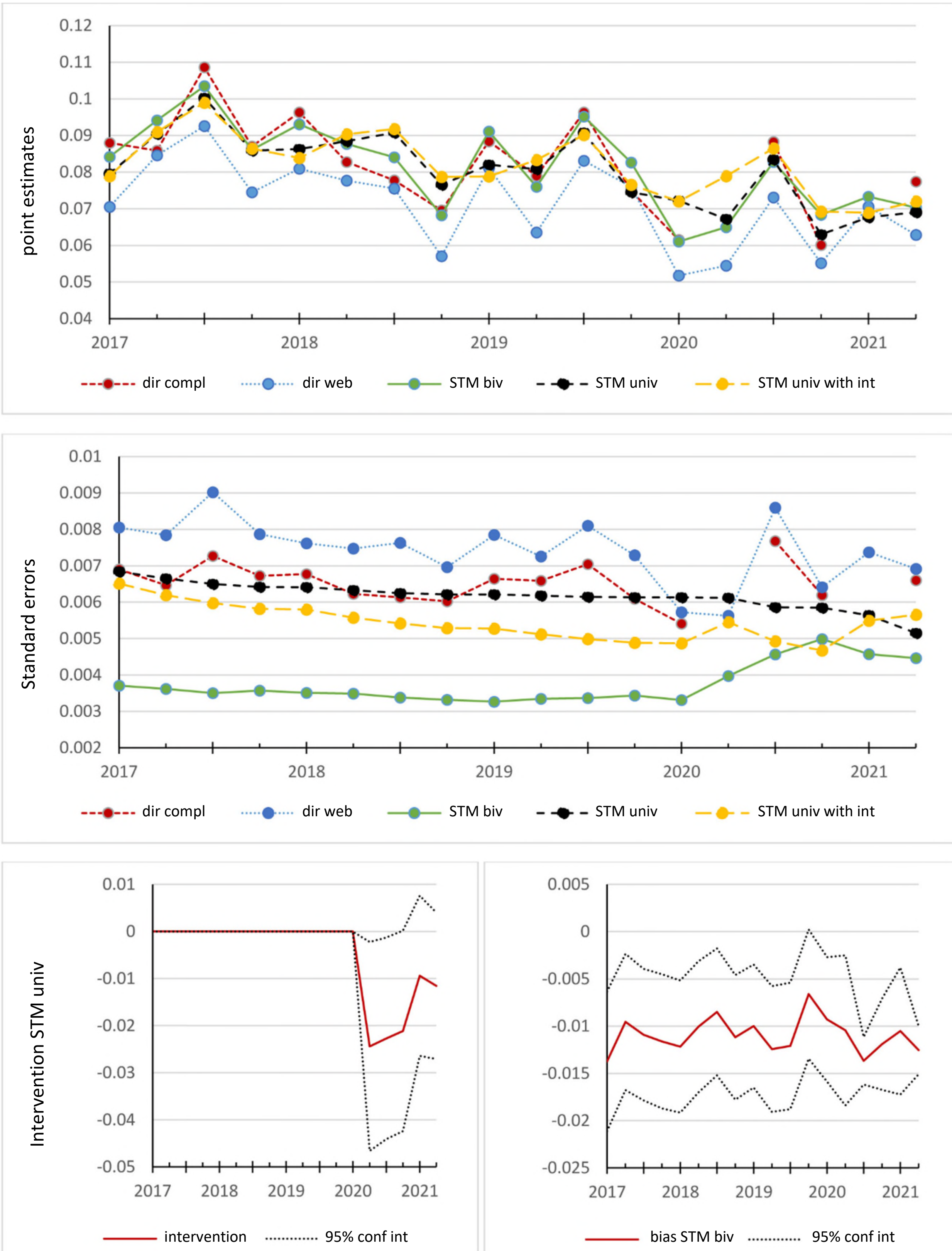


Figure 4.8 Results STM for excessive alcohol consumption.



Until 2020 there was no loss of CAPI and the STM estimates based on the univariate and bivariate models are very similar. During the Covid-19 pandemic that started in 2020 there are more clear differences between the STM estimates. Especially in quarters where CAPI is missing and for variables with a clear mode effect the univariate STM without intervention produces estimates at the level of the web series while the STM estimates by the bivariate model are at the level of the complete series. That is for example the case in the first quarters of 2020 for perceived health (Figure 4.5, top panel) and dental visit (Figure 4.6, top panel) and in the first quarter of 2021 for daily smoking (Figure 4.7, top panel). For excessive alcohol consumption similar effects are found in 2020 and 2021 (Figure 4.8, top panel), but to a lesser extent. The univariate STM without intervention produces, as expected, biased estimates during the Covid-19 pandemic in quarters where CAPI is partially or completely missing.

The univariate STM with intervention also leads to biased estimates in quarters where CAPI is partially or completely missing during one of the lockdowns. This is because the model incorrectly interprets a part of the sudden changes in the real quarterly developments as differences in measurement bias and selection effects. This can result in a large estimate for the intervention coefficient β. The effect can be seen for all variables, but is the largest for dental visit (Figure 4.6, bottom-left panel). For dental visit the resulting bias is the largest in the second quarter of 2020, when dentists in the Netherlands were only open for emergency treatments.

The bivariate STM avoids that sudden changes in the developments of the population parameter are interpreted as differences in measurement and selection bias, because nowcasts are obtained for the missing estimates based on the complete response by means of the systematic difference λ_t in the model observed in the period before the lockdown. Estimates based on the bivariate STM are at the level of the complete series and are therefore used as the official quarterly DHS figures, since they provide the most plausible correction for the loss of the CAPI respondents.

For most variables the standard errors of the STM estimators are smaller than those of the direct estimators and the standard errors of the estimates based on the univariate models are generally smaller than those based on the bivariate model. At first sight it might come as a surprise that the standard errors under the bivariate model are larger than those of the univariate models. It should be understood that the series based on CAWI is based on the same respondents that are also used in the series of the complete response. Therefore the CAWI series does not provide new sample information to the time series model. This is reflected in the covariance structure of the measurement errors (3.10). From that perspective the univariate models are more parsimonious resulting in smaller standard errors for the parameter estimates of interest. In quarters where the flexibility parameter $f_t > 1,$ the models assign more weight to the direct estimates and less strength is borrowed from the past. This results in larger standard errors that sometimes exceed the standard errors of the direct estimates. For the univariate STM with intervention this effect is large in the second quarter of 2020 (see e.g., the middle panel of Figure 4.5).

5. Official publications based on the DHS

Official quarterly figures have been published for the eight selected DHS variables (Section 2) based on the bivariate STM (3.7). The first quarterly series were published in August 2020. These series ran from the first quarter in 2017 up to the second quarter of 2020. Subsequently, new estimates were published every quarter. The quarterly figures are computed in real time and will not be revised after publication. Based on the quarterly figures also quarterly and annual developments are published. Quarterly developments are defined as the difference between two consecutive quarters and the annual developments as the difference between the same quarters in two consecutive years. The developments can be directly derived from the published quarterly figures. Standard errors for the quarterly developments are obtained by calculating the linear combination $\Delta_t^{\mathrm{Q}} = L_{t|t} - L_{t-1|t} + S_{t|t} - S_{t-1|t}$ via the Kalman filter recursion in (3.10) and (3.11). For the annual developments the standard errors are computed by calculating the linear combination of trends $\Delta_t^{\mathrm{A}} = L_{t|t} - L_{t-4|t}$. Here the linear combination of signals $L_{t|t} - L_{t-4|t} + S_{t|t} - S_{t-4|t}$ has not been used, because in that case many extra state variables should be kept in the state vector in order to compute the seasonal components $S_{t-4|t}$. This may lead to unstable estimates.

The annual DHS figures for 2020 and 2021 have been benchmarked with the quarterly figures by extending the regular weighting model described in Section 2 with the quarterly STM estimates for the eight variables for which STMs are developed. For each variable a component is constructed with eight categories that is added to the weighting model. Each target variable specifies the distribution over two categories, i.e., the fraction of people that meet the characteristic of that variable (e.g., daily smoker) and a rest category (e.g., not being a daily smoker). The components in the weighting model specify the distribution of the population over these two categories on a quarterly basis. The numbers per quarter are divided by four, such that the sum over the eight categories is equal to the size of the target population. In this way numerical consistency is achieved between the annual and quarterly publications. There is also a correction for the loss of CAPI for more detailed breakdowns of the eight variables. And finally a best possible correction is realized for the loss of CAPI for other related variables for which no model-based quarterly estimates are developed. Quarterly and annual publications for 2017, 2018 and 2019 have not been made consistent with each other, since revisions are undesired and since the size of the revision is small because there was no loss of CAPI response during this period.

The extension of the weighting model with the quarterly STM estimates resulted in a slight increase of the dispersion of the regression weights. Table 5.1 shows some results of the annual DHS figures for 2020, including the variables cancer (ever had) and bronchitis (past 12 months). The estimates in the table are percentages and the corresponding standard errors are given in parentheses. The corrections to the annual figures for the variables for which quarterly figures have been estimated are in line with the previous results discussed in Section 4. For perceived health and dental visit there is a negative correction for the loss of CAPI in 2020, while CAWI respondents score higher than CAPI respondents (Section 4). For daily smoking, and excessive alcohol consumption the correction is positive, while CAWI scores lower than CAPI. One would expect that cancer is related to the lifestyle variables, but this variable is negatively corrected from 6.47 (regular weighting) to 6.44 (extended weighting). At first sight it appears that for this variable the correction by means of the model-based quarterly figures does not work very well. On the other hand, this variable concerns all types of cancer and the relationship may be less strong and it can be anticipated that the majority of people that faced cancer in the past gave up smoking afterwards. For bronchitis, where a strong relation is expected with daily smoking, the correction is indeed in the same direction as for daily smoking.

Table 5.1
Results annual figures DHS 2020. Estimates are in percentages and standard errors in parentheses

Variable	Regular weighting	Extended weighting
Perceived health	81.70 (0.45)	81.46 (0.46)
Dental visit	16.83 (0.42)	16.08 (0.42)
Daily smoking	13.61 (0.45)	14.87 (0.49)
Excessive alcohol consumption	6.43 (0.30)	6.93 (0.33)
Cancer	6.47 (0.26)	6.44 (0.26)
Bronchitis	4.28 (0.23)	4.33 (0.23)

Note: Dutch Health Survey (DHS).

6. Discussion

Based on the Dutch Health Survey (DHS), until 2020 only annual figures on health, healthcare use and lifestyle were published by Statistics Netherlands. As a result of the Covid-19 pandemic and the associated lockdown it was decided in June 2020 to publish a series of quarterly figures based on a structural time series model (STM) for a selection of eight DHS key variables. This serves multiple purposes. Firstly, with quarterly figures the period of the corona crisis can be better delineated, so that possible effects of the crisis on the health figures is portrayed more clearly. Secondly, quarterly figures are more timely available, namely already during the statistical year and not only after the end of the reference year. This clearly increases the relevance of the health figures. Because the sample size of the DHS is too small to produce sufficiently precise quarterly figures with a direct estimator, structural time series models are used as a form of small area estimation to improve the precision of the quarterly figures with sample information from preceding reference periods. And finally, the bivariate time series model corrects for the bias that is a result of the loss of face-to-face observation during the lockdown.

The bivariate STM combines two series of direct estimates, a series based on complete response and a series based on web response. The differences between the complete series and the series based on web response are modelled dynamically in a separate component as a random walk. In quarters where face-to-face response is missing, there are no estimates available based on the complete response. For these periods, the bivariate model provides nowcasts for the population parameter of interest that are not affected by the sudden change in measurement and selection effects that are the result of the loss of CAPI because the model accommodates this difference in the aforementioned component. This approach is based on the assumption that the observed differences between the two input series in the period before the lockdown, do not change during the lockdown. The validity of this assumption is difficult to evaluate, but it has been established through a response analysis that the composition of the web response did not change during the corona crisis.

Two univariate STMs are considered as an alternative. The univariate model without an intervention component to model the shock in the input series that is the result of the loss of CAPI response, assumes that there are no mode effects between web response and face-to-face response. For the selected DHS variables there are clearly mode effects implying that this univariate STM produces biased estimates in quarters during the lockdown when there is no or less face-to-face observation possible. The second univariate STM attempts to model the change in measurement and selection bias with a level intervention variable. This is also a less optimal solution, since the lockdown also has a strong effect on the population parameters. A part of the real evolution of the population parameters is incorrectly absorbed in the level intervention, resulting in biased model predictions for the population parameters of interest. For these reasons the univariate models are unsuitable for estimating quarterly figures during the Covid-19 pandemic. Based on the bivariate STM official quarterly figures are published for the eight selected DHS variables.

The corrections for the loss of face-to-face interviewing have been incorporated in the annual figures of 2020 and 2021 by including in the weighting model of the annual response a table with the corrected model-based quarterly figures for the eight selected DHS variables. This provides numerical consistency between quarterly and annual figures. In this way a correction is also realized for the loss of face-to-face response for more detailed breakdowns of the annual figures of these eight variables and to some extent also for other related variables for which no model-based quarterly estimates are developed.

An essential advantage of using the STM is that model-based estimates are more accurate than direct estimates. In particular, period-by-period developments can be estimated much more accurately thanks to the positive correlation between trend estimates and consecutive periods.

For some variables the pandemic has had a major effect on the development. In order to account for the sudden increase in the dynamics of these figures in the time series model, it is necessary to make the trend component more flexible during the pandemic. This has been done by increasing the variance of the disturbance terms of the trend component during the pandemic. A consequence is that the standard errors of the model-based estimates increase for these quarters and are in some cases larger than the standard errors of the direct estimates.

The Covid-19 crisis increased the awareness that variance is not the only quality concept for official statistics, but that other quality dimensions such as timeliness and comparability over time are at least as important. As a result of this, Statistics Netherlands extended the traditional design-based inference approach for the annual publications of the DHS, with a model-based inference method as a form of small area estimation to produce more timely figures. At the same time, the proposed method compensates for the bias that occurs as a result of the temporal loss of CAPI responses to maintain comparability over time and avoid a sudden increased MSE.

Acknowledgements

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

References

Binder, D.A., and Dick, J.P. (1990). A method for the analysis of seasonal ARIMA models. *Survey Methodology*, 16, 2, 239-253. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/1990002/article/14533-eng.pdf.

Boonstra, H.J., and van den Brakel, J.A. (2022). Multilevel time series models for small area estimation at different frequencies and domain levels. *Annals of Applied Statistics*. Accepted for publication.

Boonstra, H.J, van den Brakel, J.A. and Das, S. (2021). Multilevel time series modeling of mobility trends. *Journal of the Royal Statistical Society A series*, 184, 985-1007.

Buelens, B., and van den Brakel, J.A. (2015). Measurement error calibration in mixed-mode surveys. *Sociological Methods & Research*, 44, 391-426.

Datta, G., Lahiri, P., Maiti, T. and Lu, K. (1999). Hierarchical bayes estimation of unemployment rates for the states of the U.S. *Journal of the American Statistical Association*, 94, 1074-1082.

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

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

Elliot, D., and Zong, P. (2019). Improving timeliness and accuracy of estimates from the UK labour force survey. *Statistical Theory and Related Fields*, 3, 186-198.

Fay, R., and Herriot, R. (1979). Estimates of income for small places: An application of James-Stein procedures to census data. *Journal of the American Statistical Association*, 74, 269-277.

Gonçalves, C., Hidalgo, L., Silva, D. and van den Brakel, J.A. (2022). Model-based single-month unemployment rate estimates for the Brazilian Labour Force Survey. *Journal of the Royal Statistical Society*, 185, 1707-1732.

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.

Kish, L. (1965). *Survey Sampling*. New York: John Wiley & Sons, Inc.

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

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

Lumley, T. (2014). 'survey': Analysis of Complex Survey Samples. R package version 3.30.

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

Pfeffermann, D., and Bleuer, S.R. (1993). Robust joint modelling of labour force series of small areas. *Survey Methodology*, 19, 2, 149-163. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/1993002/article/14458-eng.pdf.

Pfeffermann, D., and Burck, L. (1990). Robust small area estimation combining time series and cross-sectional data. *Survey Methodology*, 16, 2, 217-237. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/1990002/article/14534-eng.pdf.

Pfeffermann, D., and Tiller, R. (2005). Bootstrap approximation to prediction MSE for state-space models with estimated parameters. *Journal of Time Series Analysis*, 26(6), 893-916.

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

R Core Team (2015). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/.

Rao, J.N.K., and Molina, I. (2015). *Small Area Estimation*, 2nd edition, New York: John Wiley & Sons, Inc.

Rao, J.N.K., and Yu, M. (1994). Small area estimation by combining time series and cross-sectional data. *The Canadian Journal of Statistics/La revue canadienne de statistique*, 22(4), 511-528.

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

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., and Krieg, S. (2015). Dealing with small sample sizes, rotation group bias and discontinuities in a rotating panel design. *Survey Methodology*, 41, 2, 267-296. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2015002/article/14231-eng.pdf.

van den Brakel, J.A., and Krieg, S. (2016). Small area estimation with statespace common factor models for rotating panels. *Journal of the Royal Statistical Society A series*, 179, 763-791.

van den Brakel, J.A., Souren, M. and Krieg, S. (2022). Estimating monthly Labour Force Figures during the COVID-19 pandemic in the Netherlands. *Journal of the Royal Statistical Society, Series A*, 185, 1560-1583.

You, Y. (2008). Small area estimation using area level models with model checking and applications. SSC annual meeting, Proceedings of the Survey Methods Section.

You, Y., Rao, J.N.K. and Gambino, J. (2003). Model-based unemployment rate estimation for the Canadian Labour Force Survey: A hierarchical Bayes approach. *Survey Methodology*, 29, 1, 25-32. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2003001/article/6602-eng.pdf.