Discussion paper

Estimating monthly indicators for Consumer Confidence using Structural Time Series Models

Jan van den Brakel, Sabine Krieg and Marc Smeets

October 2021

Content

Summary

The Dutch Consumer Survey measures consumer confidence in the Netherlands with a monthly frequency. In this paper a model-based inference procedure based on a multivariate structural time series model is developed for the production of monthly figures about consumer confidence. The input for the model consists of five series of direct estimates for the indices, that are used to construct the consumer confidence index. Although the standard errors of the model-based estimates are not smaller than the standard errors of the direct estimates, the model improves the accuracy of the estimates for consumer confidence, since it provides a better separation of measurement error and sampling error from estimated target parameters. The standard errors for the month-to-month changes are clearly smaller under the time series model. A second problem addressed in this paper is related to the transition to a new survey process in 2017. Structural time series models in combination with a parallel run are applied to estimate discontinuities induced by the redesign. A backcasting algorithm designed for the consumer confidence variables is developed to construct uninterrupted input series for the aforementioned structural time series model. This inference method facilitated a smooth transition to a new survey design and resulted in uninterrupted series about consumer confidence that date back to 1986.

Keywords

Official statistics, survey redesign, discontinuities, back casting, small area estimation

Reviewer

Rob Willems

1. Introduction

The Dutch Consumer Survey (CS) measures consumer confidence in the Netherlands. This paper considers two problems concerning this survey. First, the series of the monthly estimates is quite volatile. The evolution of the figures is therefore not always easy to interpret. This volatility is the result of the relatively small sample (1000 respondents every month) and because the CS measures a relative short term emotion about the respondents opinion of the financial and economic climate, despite the fact that most questions actually refer to the last or upcoming 12 months. Increasing the sample size would therefore only partially reduce the volatility of the series. This is, however, not an option, since Statistics Netherlands (as many other statistical institutes) has to reduce administration costs and response burden. The second problem is the disruption of the continuity of the time series as a result of a redesign of the survey process that took place in the beginning of 2017. Changes in the questionnaire and fieldwork strategies generally have a systematic effect on the outcomes of a survey, since these affect non-sampling errors like measurement bias and selection bias. Systematic differences in time series induced by survey redesigns are further referred to as discontinuities. In a well-designed transition process, discontinuities are quantified to avoid confounding real period-to-period change from sudden changes in measurement bias and selection effects.

Since the CS is a repeatedly conducted survey, a structural time series model is developed as a solution for both problems. By the model, sample information from the past is used to make more accurate estimates compared to the direct sample estimates. The use of time series modelling with the aim of improving the precision of survey data has been considered by many authors dating back to Blight and Scott (1973) and can be interpreted as a form of small area estimation by borrowing strength over time (Rao and Molina, 2015). When different time series are modelled at the same time, a multivariate structural time series can be applied. If multiple time series are combined in a multivariate STM, correlations between disturbance terms of the unobserved components of the different series can be modelled to further improve the precision of the estimates. This approach has been applied before in the context of official statistics, see e.g. Tam (1987), Binder and Dick (1989, 1990), Bell and Hillmer (1990), Tiller (1992), Rao and Yu (1994), Pfeffermann and Burck (1990), Pfeffermann and Bleuer (1993), Pfeffermann (1991), Pfeffermann and Tiller (2006), Harvey and Chung (2000), and Feder (2001), Van den Brakel and Krieg (2015) and Elliot and Zong (2019).

Discontinuities caused by a redesign of the survey process can be quantified in different ways, see Van den Brakel, Zhang and Tam (2020) for an overview. One approach is to collect data under both the old and the new design in parallel for some period of time, which is further referred to as a parallel run. The difference of the estimates based on both designs can be used as a direct estimate of the discontinuity. In the case of a sufficiently large parallel run this is a reliable and timely approach. The major drawback is that it requires additional data collection, which makes the approach costly. Alternatively a time series model can be applied where the discontinuities are estimated using a level intervention (Van den Brakel and Roels, 2010). In the case of a small parallel run, the information from the parallel run can be used as a-priori information in the time series model, e.g. through an exact initialization of the Kalman filter. This initial estimate is further improved with the information from the time series observed before and after the parallel run (Van den Brakel and Krieg, 2015).

In this paper multivariate structural time series models are applied to the CS as a form of small area estimation and to account for discontinuities induced by the redesign of 2017. The underlying series which together define the consumer confidence are the input series of the time series model. In this way sample information observed in previous reference periods is used to obtain more accurate estimates for CS. Moreover, correlations between the disturbances of the trend are useful to borrow strength from the variables used in the construct for consumer confidence. In this paper we also discuss the role of correlations between the measurement error terms of the series. Since the series are measurements obtained from the same sampling units, the sampling and measurement errors are correlated. Furthermore, sudden real events can influence all series simultaneously, which results in correlated population disturbance terms. In this paper it is motivated that modelling correlations between the measurement errors is necessary to achieve a more optimal separation of trend and seasonal from measurement error.

Discontinuity estimates for the CS caused by the redesign are based on a parallel run, where the old and the new design are conducted in parallel for three months, both at the regular sample size. The difference of the estimates based on both designs can be used as a direct estimate of the discontinuity. These estimates are improved with a structural time series model, as outlined above. When the estimates for discontinuities are known, it is important to communicate about them with the users of the series to avoid misinterpretation of the series. In the case of the CS the series of the past are corrected for the discontinuities. The series underlying the CS are percentages and a correction method is proposed that attempts to keep the adjusted values in the admissible range between 0% and 100%. The time series modelling approach developed in this paper has been implemented for the production of Statistics Netherlands official monthly consumer confidence figures since April 2017.

The paper is organized as follows. Section 2 provides a description of the Dutch CS. In Section 3, a structural time series model is developed for the estimation of monthly consumer confidence figures, including results for the period before the redesign in 2017. In Section 4 the change-over to the new design that took place in 2017 is described. In this section a method for estimating discontinuities that combines a parallel run with a time series modelling approach is proposed. Furthermore a correction method to adjust the series observed before the change-over to the level of the series observed under the new design is described. Section 5 summarizes how the estimation method is implemented for the production of official monthly consumer confidence figures. The paper finalizes with a conclusion in Section 6.

2. The Dutch Consumer Survey

The Consumer Survey (CS) is a monthly survey and is carried out following the joint harmonized EU Consumer Survey (European Commission, 2014). Before the redesign of 2017, each month a self-weighted sample of approximately 2,500 households was drawn by stratified two-stage sampling from a sample frame derived from the Dutch Municipal Register. Stratification is based on the cross-classification of 12 provinces and urbanization level in five classes. Primary sampling units are municipalities. Households for which a known telephone number was available were contacted by an interviewer who completes the questionnaire by computer assisted telephone interviewing (CATI) during the first ten working days of the month. On average a net sample of about 1,000 responding households was obtained, resulting in a response rate of about 40%. A major part of the nonresponse consisted of households for which no known telephone number of a land-line connection is available. The response among households for which a known telephone number was available was about 60%.

For the computation of the consumer confidence, five questions are relevant. These questions are about

1. opinion about changes of the general economic situation of the country over the last 12 months, abbreviated as Econ. L12,
2. expectations of changes of the general economic situation of the country over the next 12 months, abbreviated as Econ. N12,
3. opinion about changes of the financial situation of the household over the last 12 months, abbreviated as Fin. L12,
4. expectations of changes of the financial situation of the household over the next 12 months, abbreviated as Fin. N12,
5. whether it is the right moment for people to make major purchases, abbreviated as Major pur.

For the first four questions there are two positive and two negative answer options (“a lot better”, “a little better”, “a lot worse”, “a little worse”). Furthermore, there is the neutral option “the same” as well as “do not know”. For the fifth question there is one positive and one negative answer option (“yes, it is the right moment now”, “no, it is not the right moment now”). Furthermore, there is a neutral option (“it is neither the right moment nor the wrong moment”) as well as “do not know”. The percentages of positive, negative and neutral answers (as percentage points of the total answers) p_i^+, p_i^-, p_i^0 with $p_i^+ + p_i^0 + p_i^- = 100$ are computed for each question $i = 1, \dots, 5$. For each question $i = 1, \dots, 5$, the difference $y_i = p_i^+ - p_i^-$ of positive and negative answers as percentage points of the total answers is computed. Furthermore, the questions are combined by computing the following averages:

- $y_6 = (y_1 + y_2)/2$ which is the indicator for economic climate,
- $y_7 = (y_3 + y_4 + y_5)/3$ which is the indicator for willingness to buy,
- $y_8 = (y_1 + y_2 + y_3 + y_4 + y_5)/5$ which is the indicator for consumer confidence.

The indicators $y_1, \dots, y_8$ are the main target variables in the publication. Until the end of 2016 unweighted sample means were used as estimates for the target variables. Expressions for the variance of the eight series are given by Van den Brakel et al. (2017). The publication of monthly figures started in 1986. Both the original figures and seasonally adjusted figures of the indicator series are published. Furthermore, the underlying series of the percentages are also published.

In January 2017 the design of the CS changed. Five important changes were implemented simultaneously; 1) The sample design changed from a self-weighted stratified two-stage sample of households to a self-weighted stratified two-stage sample of persons. A sample of 2150 persons is drawn each month, resulting in generally slightly more than 1000 respondents, with around 80% of them responding via web. So the response rate is around 47%. 2) The data collection mode changed from CATI to a sequential mixed mode design, where the respondents are first asked to complete a questionnaire via web. Then, the web non-respondents are interviewed by phone (as far as phone numbers are available). 3) There are changes in the questionnaire. Most importantly is the way in which the answer categories are offered. Under the old questionnaire the respondent could first choose between the options "worse", "neutral", or "better". In the case "worse" or "better" is selected, the respondent had to specify whether it is "a lot" or "a little" better or worse. In the new questionnaire the two positive and two negative answer options for questions 1 to 4 are shown directly. 4) Another important change is that a conditional incentive is given to respondents to improve the response rate (a tablet is raffled among the respondents). 5) Finally, the sample estimates are based on the general regression estimator (Särndal et al. 1992) to correct, at least partially, for selective non-response.

A side effect of this redesign is that it causes a sudden change in selection effects, as another part of the population is willing to respond when another mode is applied, and when an incentive is offered. Furthermore, there are sudden changes in the measurement bias due to the use of another data collection mode (partially without interviewer) and changes in the questionnaire. These cause the so-called discontinuities. In order to distinguish the real period-to-period change from differences in measurement and selection bias, it is important to quantify the discontinuities that occur as a result of the redesign of the survey.

3. Inference for monthly CS figures with STM

In this section a structural time series model is developed using the series observed until December 2016. With a structural time series model, a series is decomposed in a trend component, a seasonal component, other cyclic components, a regression component and an irregular component for the unexplained variation. 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 (2012) for details about structural time series modelling.

3.1 Description model

Each month t a direct sample estimate $\hat{y}_{i,t}$ is computed for the five questions of the CS $(i = 1, \dots, 5)$. These sample estimates can be considered as the sum of the true but unknown population parameter, say $\theta_{i,t}$, and a sampling error, say $\tilde{e}_{i,t}$. This gives rise to a measurement error model

$$\hat{y}_{i,t} = \theta_{i,t} + \tilde{e}_{i,t}, (i = 1, \dots, 5). \tag{3.1}$$

For the unknown population parameter, a basic structural time series model is assumed, i.e.

$$\theta_{i,t} = L_{i,t} + S_{i,t} + I_{i,t}, (i = 1, \dots, 5), \tag{3.2}$$

with $L_{i,t}$ the level of a stochastic trend component, which models the low frequency variation, $S_{i,t}$ a stochastic component that models the seasonal fluctuation around the trend and $I_{i,t}$ the population irregular term. Inserting (3.2) into (3.1) gives the time series model for the observed series:

$$\hat{y}_{i,t} = L_{i,t} + S_{i,t} + I_{i,t} + \tilde{e}_{i,t}, (i = 1, \dots, 5).$$

The CS is a cross-sectional survey with a homoscedastic sampling error, since the net sample size is constant over time (on average 1000 respondents per month). Therefore the population irregular term and the sampling error are combined in one measurement error, i.e. $e_{i,t} = I_{i,t} + \tilde{e}_{i,t}$, resulting into the final univariate models for the five series:

$$\hat{y}_{i,t} = L_{i,t} + S_{i,t} + e_{i,t}, (i = 1, \dots, 5), \tag{3.3}$$

The five series defined in (3.3) can be combined in a vector $\hat{\mathbf{Y}}_t = (\hat{y}_{1,t}, \hat{y}_{2,t}, \hat{y}_{3,t}, \hat{y}_{4,t}, \hat{y}_{5,t})'$, which can be modelled as

$$\hat{\mathbf{Y}}_t = \mathbf{L}_t + \mathbf{S}_t + \mathbf{e}_t, \tag{3.4}$$

with $\mathbf{L}_t = (L_{1,t}, L_{2,t}, L_{3,t}, L_{4,t}, L_{5,t})'$, $\mathbf{S}_t = (S_{1,t}, S_{2,t}, S_{3,t}, S_{4,t}, S_{5,t})'$ and $\mathbf{e}_t = (e_{1,t}, e_{2,t}, e_{3,t}, e_{4,t}, e_{5,t})'$.

The trends $L_{i,t}$ $(i = 1, \dots, 5)$ are modelled with the so-called smooth trend model:

$$L_{i,t} = L_{i,t-1} + R_{i,t-1},$$

$$R_{i,t} = R_{i,t-1} + \eta_{R,i,t}. \tag{3.5}$$

with $R_{i,t}$ the slope parameter of series i. The disturbance terms $\eta_{\mathrm{R},i,t}$ are normally distributed with

$$\mathrm{E}(\eta_{\mathrm{R},i,t}) = 0,$$
$$\mathrm{Cov}(\eta_{\mathrm{R},i,t}, \eta_{\mathrm{R},i',t'}) = \begin{cases} \sigma^2_{\mathrm{R},i} & \text{if} \quad i = i' \text{ and } t = t' \\ \varsigma_{\mathrm{R},i,i'} & \text{if} \quad i \neq i' \text{ and } t = t' \\ 0 & \text{if} \quad t \neq t' \end{cases}. \tag{3.6}$$

This is a dynamic model for the low frequency variation, which has the flexibility to capture trend as well as economic cycles. It can therefore be interpreted as the trend plus economic cycle, which is shortly referred to as trend.

The so-called trigonometric seasonal model is used to model the seasonal component $S_{i,t}$ ($i = 1, \dots, 5$), which is defined as:

$$S_{i,t} = \sum_{l=1}^{6} S_{i,t,l}, \tag{3.7}$$

with

$$S_{i,t,l} = S_{i,t-1,l} \cos(h_l) + S^*_{i,t-1,l} \sin(h_l) + \eta_{S,i,t,l},$$
$$S^*_{i,t,l} = S^*_{i,t-1,l} \cos(h_l) - S_{i,t-1,l} \sin(h_l) + \eta^*_{S,i,t,l},\ h_l = \frac{\pi l}{6}, l = 1, \dots, 6.$$

The disturbances $\eta_{\mathrm{S},i,t,l}$ and $\eta^*_{\mathrm{S},i,t,l}$ are normally distributed with

$$\mathrm{E}(\eta_{\mathrm{S},i,t,l}) = \mathrm{E}(\eta^*_{\mathrm{S},i,t,l}) = 0,$$
$$\mathrm{Cov}(\eta_{\mathrm{S},i,t,l}, \eta_{\mathrm{S},i',t',l'}) = \mathrm{Cov}(\eta^*_{\mathrm{S},i,t,l}, \eta^*_{\mathrm{S},i',t',l'}) = \begin{cases} \sigma^2_{\mathrm{S},i} & \text{if} \quad i = i' \text{ and } t = t' \text{ and } l = l', \\ 0 & \text{otherwise} \end{cases}$$

$$\mathrm{Cov}(\eta_{\mathrm{S},i,t,l}, \eta^*_{\mathrm{S},i',t',l'}) = 0 \text{ for all } i, t, l.$$

The measurement errors are normally distributed with:

$$\mathrm{E}(e_{i,t}) = 0,$$
$$\mathrm{Cov}(e_{i,t}, e_{i',t'}) = \begin{cases} \sigma^2_{\mathrm{e},i} & \text{if} \quad i = i' \text{ and } t = t' \\ \varsigma_{\mathrm{e},i,i'} & \text{if} \quad i \neq i' \text{ and } t = t' \\ 0 & \text{if} \quad t \neq t' \end{cases}. \tag{3.8}$$

Modelling correlation between the measurement error is partly motivated by properties of the sampling error. The five questions are asked to the same persons, and it is likely that persons that are positive about one aspect of consumer confidence are more often also positive about the other questions. For another part this assumption is motivated by properties of the noise in the population parameters. It is likely that some events, e.g. good or bad news about the economy, influence the answers to all questions in a similar way.

Through the trend and seasonal components in this model, information from the past about the long-term development and seasonal fluctuations is used to improve the direct sample estimates. By modelling the covariance between the trend disturbances and the measurement errors, the precision of the estimates is further improved with information from related variables.

In (3.8) it is assumed that the variance is constant over time, which is actually the case in this application as will be shown in Section 3.2. Generally this is not the case, because the sampling errors are heteroscedastic as a result of changing sample sizes. In such cases, the variance of the measurement errors can be made proportional to the variance and covariances of the input series:

$$\mathrm{Cov}(e_{i,t}, e_{i',t'}) = \begin{cases} \widehat{\mathrm{var}}(y_{i,t})\sigma_{e,i}^2 & \text{if} \quad i = i' \text{and } t = t' \\ \widehat{\mathrm{cov}}(y_{i,t}, y_{i',t})\varsigma_{e,i,i'} & \text{if} \quad i \neq i' \text{and } t = t' \\ 0 & \text{if} \quad t \neq t' \end{cases}$$

where $\widehat{\mathrm{var}}(y_{i,t})$ and $\widehat{\mathrm{cov}}(y_{i,t}, y_{i',t})$ are estimated from the survey data and are used as a-priori available information in the time series model. In this case $\sigma_{e,i}^2$ and $\varsigma_{e,i,i'}$ are scaling factors that are estimated with the time series model but have values that should be close to one (as long as the sampling error dominates the population white noise component).

The following three versions of model (3.4) will be compared to investigate the influence of the correlations between trend and measurement error components:

- Model 1: the model as described by equations (3.4) – (3.8), i.e., with correlations between the slope disturbances and between the measurement errors of the five series.
- Model 2: similar to Model 1, but without correlation between the measurement errors of the five series, i.e., $\varsigma_{\mathrm{e},i,i'} = 0$.
- Model 3: similar to Model 1, but without correlation between the slope disturbances and without correlation between the measurement errors of the five series, i.e., $\varsigma_{\mathrm{R},i,i'} = 0$ and $\varsigma_{\mathrm{e},i,i'} = 0$.

After the model is estimated, model-based estimates for the five series can be computed. The trends $L_{i,t}$ are used instead of seasonally corrected figures. Model-based estimates with the seasonal pattern included are computed as $L_{i,t} + S_{i,t}$. This is also called the signal. The model-based estimates of the combined series economic climate, willingness to buy, and consumer confidence are computed as means of the estimates for the five series. The model estimates are therefore automatically consistent. The standard errors for the model estimates of the combined series account for the correlation between the state variables ($L_{i,t}$ and $S_{i,t}$) of the underlying series.

The general way to proceed is to put the structural time series model in state-space representation (see Harvey, 1989 and Durbin and Koopman, 2012). Then the Kalman filter can be applied to obtain optimal estimates for the state vector. The Kalman filter is a recursive procedure to obtain optimal estimates for the state vector at time t based on the data up to and including time period t, and 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. Let α_t denote the vector with unknown state variables for period t. Let $\alpha_{t|t'}$ denote the estimate for the state variables for period t, based on the observations obtained until (and including) period t'. If T denotes the length of the completely observed series, then $\alpha_{t|t}$ are the filtered estimates and $\alpha_{t|T}$ are the smoothed estimates. The hyperparameters σ_*^2 and ς_* are estimated with a maximum likelihood procedure, using a numerical optimization procedure. The maximum likelihood estimates for the hyperparameters are inserted into the Kalman filter but treated as if they are the true values, known without error. This implies that the additional uncertainty of using the maximum likelihood estimates for the hyperparameters is ignored in the standard errors for the filtered and smoothed estimates for the trend and signal of the CS parameters. This is a standard approach in state-space modelling and acceptable in this application given the long series that are available. Finally the state variables are initialized with a diffuse initialization, unless stated differently. See Harvey (1989) or Durbin and Koopman (2012) for technical details. In this paper Ssfpack 3.0 (Koopman et al., 1999b, and Koopman et al., 2008) in combination with Ox (Doornik, 1998) is used for the computations.

3.2 Results

The three models described in Section 3.1 are applied to series from January 1987 until December 2016. In the figures in this section, model estimates for 2001 – 2016 are shown. In the years before, the results are similar. To facilitate a better interpretation of the graphs, a shorter period is presented. The model evaluation is based on the entire series.

Maximum likelihood estimates for the standard deviations of slope disturbance terms, seasonal disturbance terms and the measurement error for the five baseline variables are presented in Table 3.1 for the three models. The variance of the slope disturbance terms under Model 2 are larger compared to Model 1 and Model 3. The seasonal component is time invariant for Econ. N12 and almost time invariant for Econ. L12, Fin. L12 and Fin. N12. For Major pur. the seasonal component varies over time.

Hyperparameter	Variable	Standard deviation		
		M1	M2	M3
Slope ($\sigma_{\mathrm{R},i}$)	Econ. L12	2.61	4.01	2.62
	Econ. N12	2.88	4.84	2.84
	Fin. L12	0.54	0.59	0.46
	Fin. N12	0.60	1.04	0.54
	Major pur.	0.94	1.13	0.87
Seasonal ($\sigma_{\mathrm{S},i}$)	Econ. L12	0.03	0.02	0.03
	Econ. N12	0.00	0.00	0.00
	Fin. L12	0.03	0.02	0.03
	Fin. N12	0.04	0.04	0.05
	Major pur.	0.12	0.12	0.12
Measurement error ($\sigma_{e,i}$)	Econ. L12	3.83	3.18	3.83
	Econ. N12	6.14	5.12	6.08
	Fin. L12	2.49	2.53	2.51
	Fin. N12	2.73	2.61	2.62
	Major pur.	2.68	2.66	2.69

Table 3.1: maximum likelihood estimates standard deviations slope disturbance terms, seasonal disturbance terms and the measurement error

Tables 3.2 – 3.4 show the maximum likelihood estimates of the correlations of Model 1 and 2. High correlations between slope disturbance terms are observed for the pairs Econ. L12 – Econ. N12, Fin. L12 – Econ. L12, Fin. N12 – Econ. L12, Fin. N12 – Econ. N12, Fin. N12 – Fin. L12, Major. Pur. – Econ. L12, Major. Pur. – Econ. N12, Major. Pur. – Fin. L12, and Major. Pur. – Fin. N12, which makes sense. The correlations between the slope disturbances (Table 3.2) are larger than the ones between the measurement error (Table 3.3) under Model 1. Furthermore, the correlations of the slope disturbances are larger under Model 2 than under Model 1. This is because under Model 1, part of the co-movements of the series is considered as correlations between the measurement errors. Under Model 2 this variation is interpreted as trend fluctuations. As explained in Section 3.1, there are arguments that the measurement errors are correlated. Model 1 is therefore preferred over Model 2, and the correlations under this model (Table 3.4) are probably over-estimated. From Table 3.1 it follows that the slope disturbance terms are consistently higher compared to Model 1 and Model 3, which is another indication that the trends under Model 2 tend to overfit the observed series.

	Econ. L12	Econ. N12	Fin. L12	Fin. N12	Major pur.
Econ. L12	1				
Econ. N12	0.861	1			
Fin. L12	0.595	0.360	1		
Fin. N12	0.909	0.929	0.580	1	
Major pur.	0.527	0.376	0.771	0.614	1

Table 3.2: correlations slope disturbances, model with correlations slope disturbances and correlations measurement (Model 1)

	Econ. L12	Econ. N12	Fin. L12	Fin. N12	Major pur.
Econ. L12	1				
Econ. N12	0.562	1			
Fin. L12	0.089	-0.079	1		
Fin. N12	0.199	0.340	0.203	1	
Major pur.	0.211	0.207	-0.078	0.002	1

Table 3.3: correlations measurement error, model with correlations slope disturbances and correlations noise (Model 1)

	Econ. L12	Econ. N12	Fin. L12	Fin. N12	Major pur.
Econ. L12	1				
Econ. N12	0.956	1			
Fin. L12	0.714	0.577	1		
Fin. N12	0.971	0.980	0.712	1	
Major pur.	0.740	0.668	0.830	0.756	1

Table 3.4: correlations slope disturbances, model with correlations slope disturbances and without correlations measurement error (Model 2)

Figure 3.1 – 3.3 compare the filtered trends under the three models for three target variables (Econ. L12, Fin. N12 and consumer confidence). A comparison with the direct estimates follows in Figures 3.4 – 3.6. For all three variables, the filtered trends (Figure 3.1 – 3.3) are more or less similar under the three models. For some periods, smaller differences are visible. These differences can be relevant in the publication of the figures.

In Figure 3.1, the figures under Model 1 and Model 3 are very similar in most periods. The trend under Model 2 is more volatile. As mentioned before, a part of the measurement error under Model 1 and 3 is interpreted as trend movements under Model 2. It appears that only allowing for correlation between the slope disturbance terms results in a sub-optimal separation of the variation over the trend and measurement error components.

In Figure 3.2, the trend for Fin. N12 under Model 3 deviates more from Model 1 and 2. Under Model 3, no information from the other series is used to improve the estimates. Since the trends develop similarly, the accuracy of the estimates can be improved when the correlation is taken into account, as under Model 1 and Model 2.

For the other three input series the filtered trends are similar as for the two series discussed in Figures 3.1 and 3.2. For the variable Econ. N12, the trend under Model 2 is more volatile than under the other two models, similarly as in Figure 3.1. For the variable Fin. L12, the filtered trends are similar as in Figure 3.2. For Major pur., the trend under Model 3 differs from the trend under the other models.

The three other variables are linear combinations of these 5 series. There, the results are mixed, i.e., sometimes Model 3 deviates more from Model 1 and 2, since it does not borrow information from the other variables. Sometimes the trend under Model 2 is more volatile, as in the case of consumer confidence, which is shown as an example in Figure 3.3. A preliminary conclusion is that Model 1 is preferred. This choice will be discussed later in more detail.

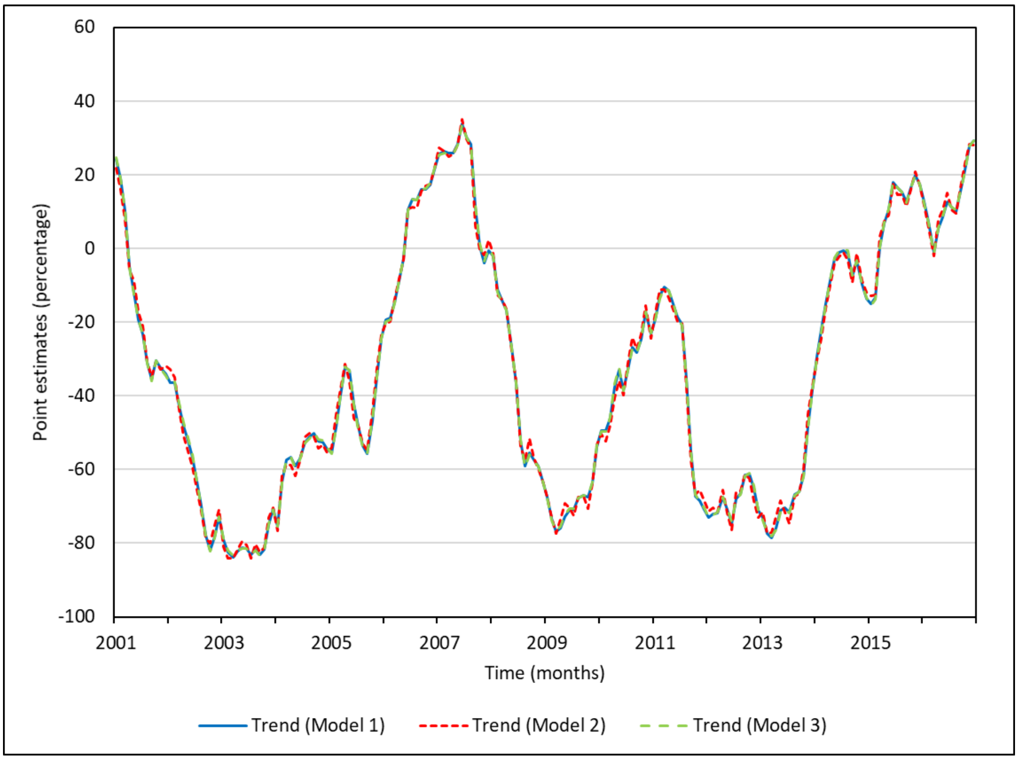

Figure 3.1: filtered trends for three models for Econ. L12

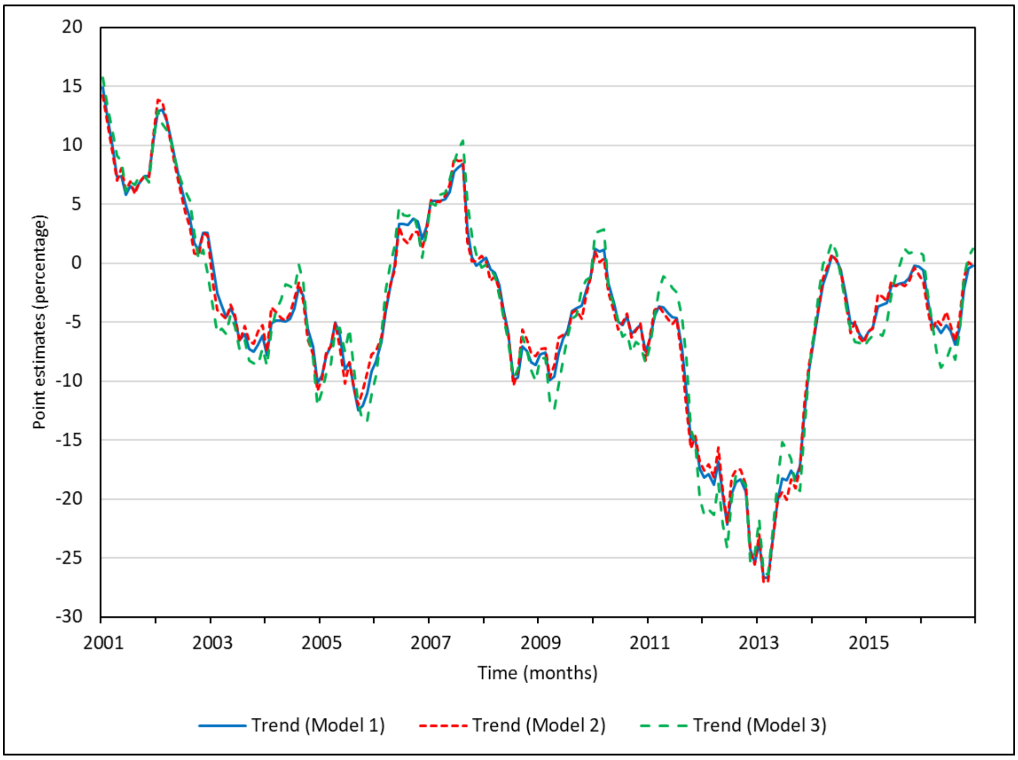

Figure 3.2: filtered trends for three models for Fin. N12

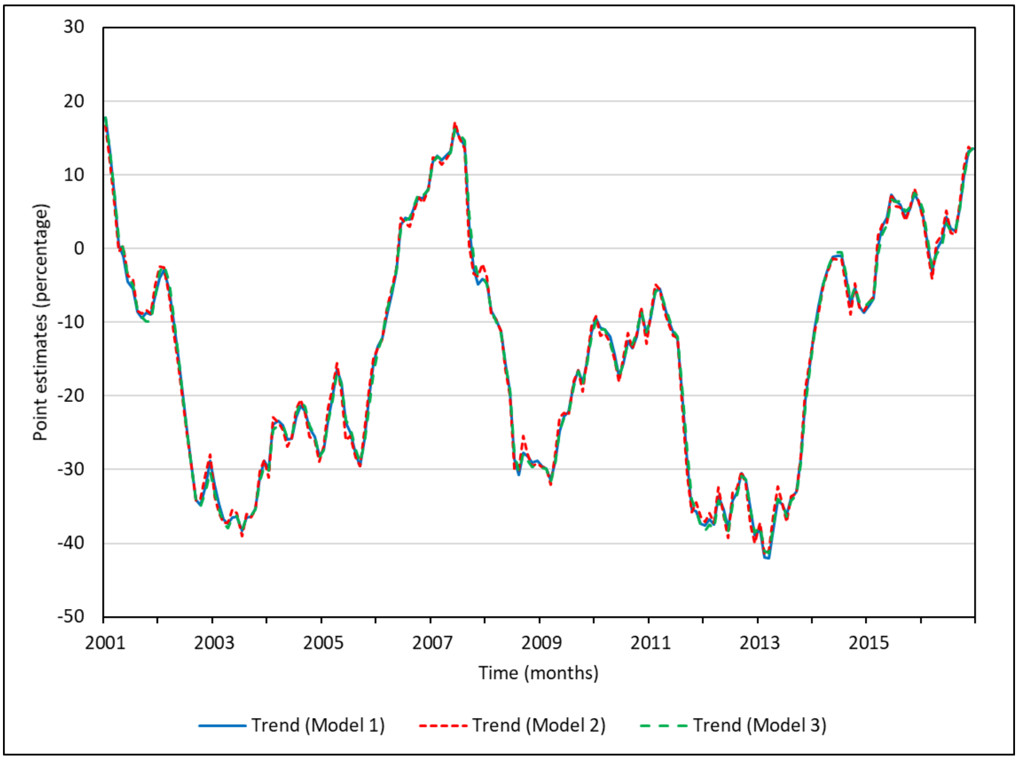

Figure 3.3: filtered trends for three models for consumer confidence

In Figures 3.4 – 3.6 the filtered signal under Model 1 is compared with the input series. Since the results under the three models are quite similar, Model 2 and 3 are omitted in these figures. Figure 3.4 (Econ. L12) illustrates that the model estimates closely follow the input series. Here the time series model hardly smooths the sample estimates. For the variable Fin. N12 in Figure 3.5, on the other hand, the model estimates are more smooth than the direct estimates. Here the model removes a substantial part of the high frequency variation from the series, which is considered an improvement of the accuracy. For consumer confidence in Figure 3.6, which is a linear combination of the five baseline series, the volatility is slightly reduced by the model. For all other variables, a reduction of the volatility is found.

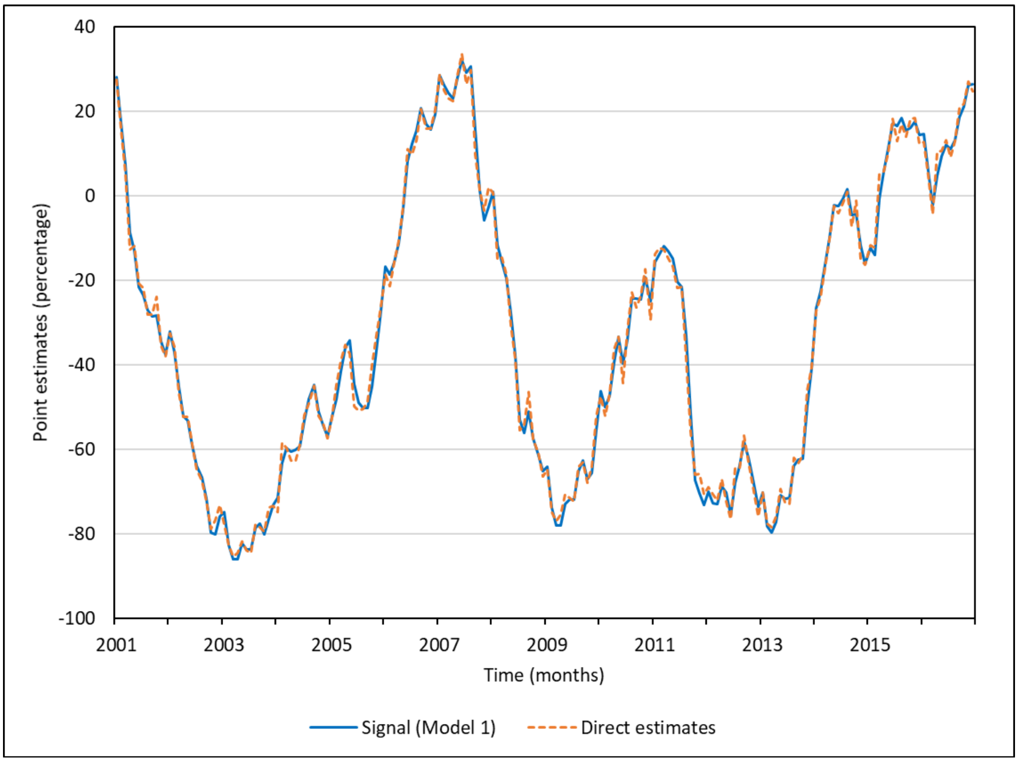

Figure 3.4: direct estimates and filtered signal under Model 1 for Econ. L12

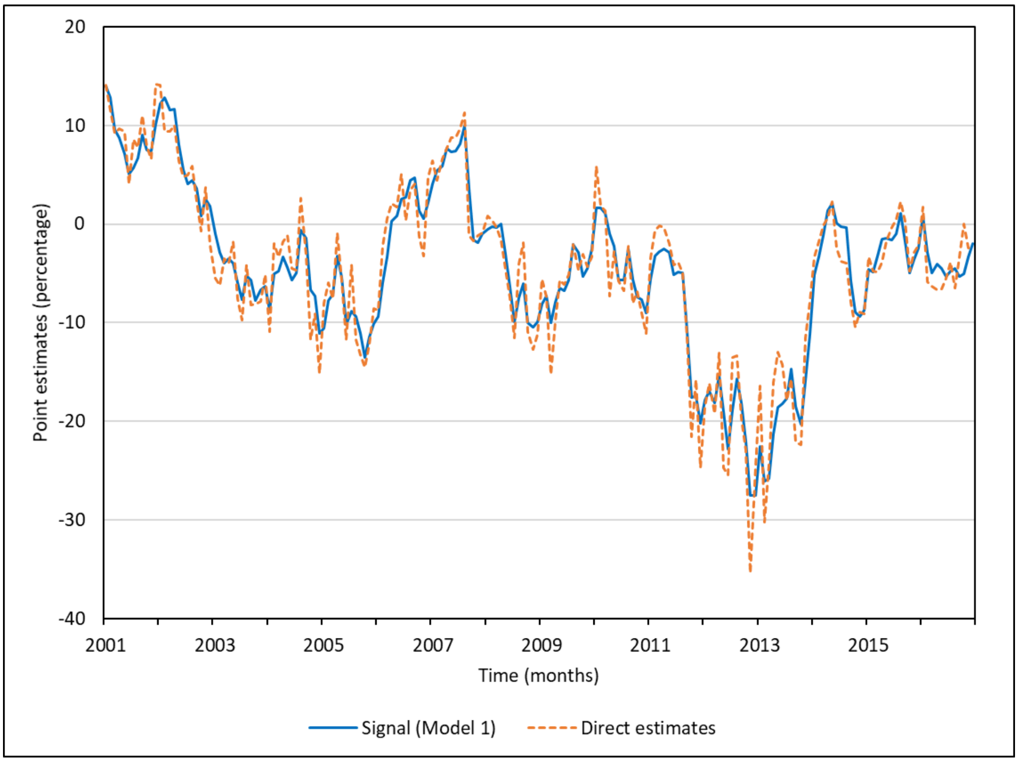

Figure 3.5: direct estimates and filtered signal under Model 1 for Fin. N12

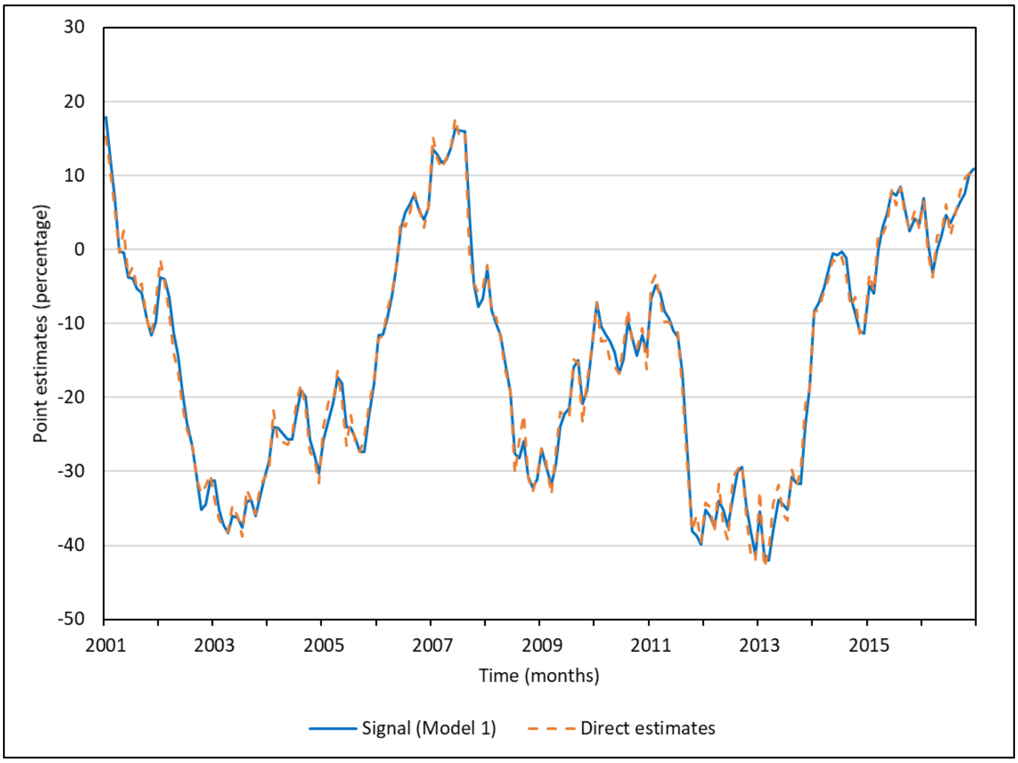

Figure 3.6: direct estimates and filtered signal under model 1 for consumer confidence

In Figure 3.7 – 3.9 the standard errors of the filtered signals under the three time series models are compared with each other and with the standard error of the direct estimates. Results are presented for the three combined series. The standard error for the filtered signals is the largest under Model 1 for all three variables. For economic climate in Figure 3.7, the standard error of the direct estimates is smaller than the standard errors of the filtered signals for all three models. For willingness to buy in Figure 3.8 the standard errors of the filtered signals under all three models are smaller than those of the direct estimates. For consumer confidence in Figure 3.9 the standard errors under Model 2 and 3 are more or less equal to the standard errors of the direct estimates, while the standard errors of Model 1 are larger. Only in Figure 3.8, the precision of the estimates is improved by all models.

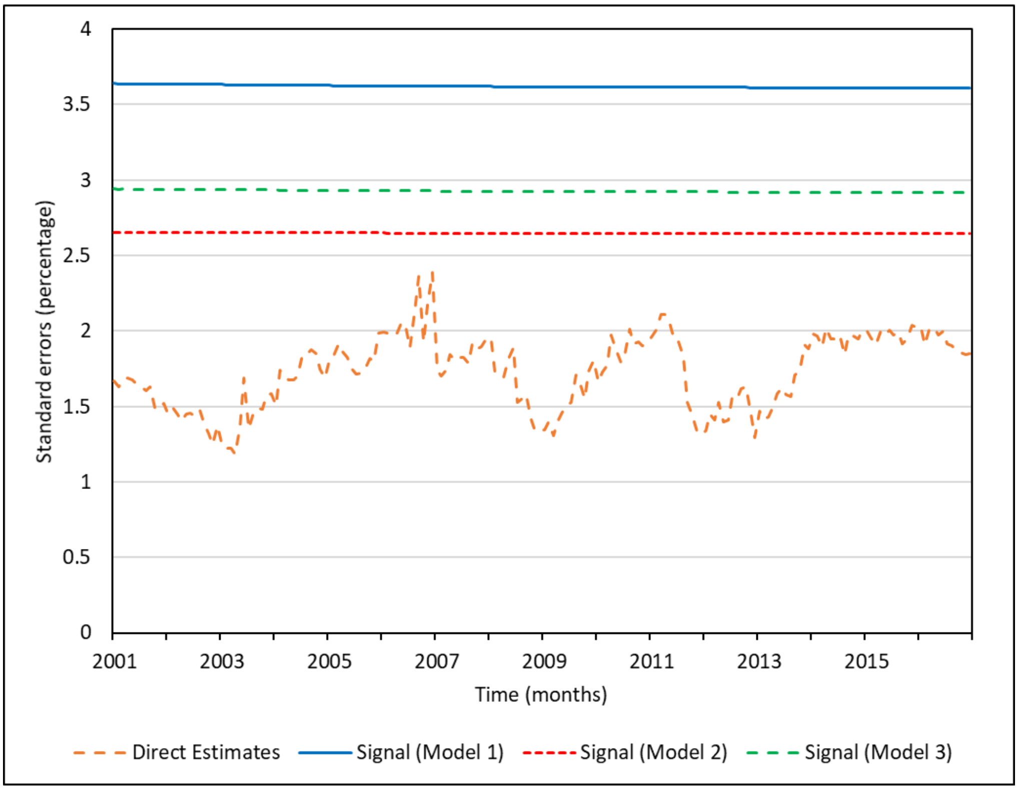

Figure 3.7: standard errors direct estimates and filtered estimates signal under three models for economic climate.

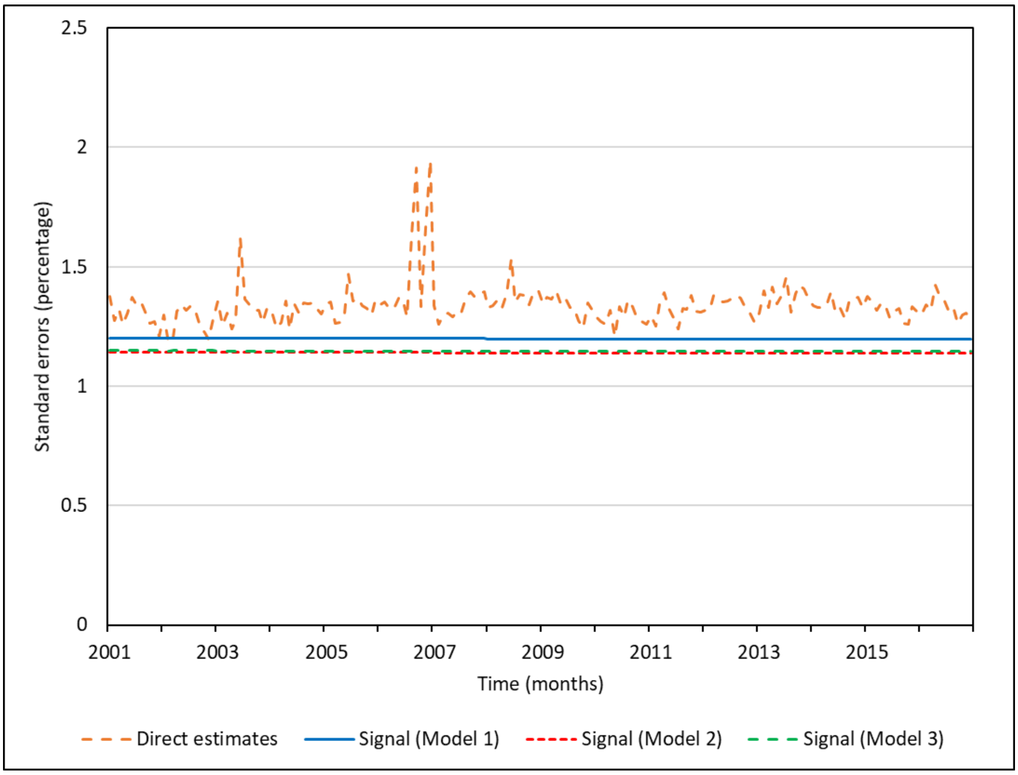

Figure 3.8: standard errors direct estimates and filtered estimates signal under three models for willingness to buy

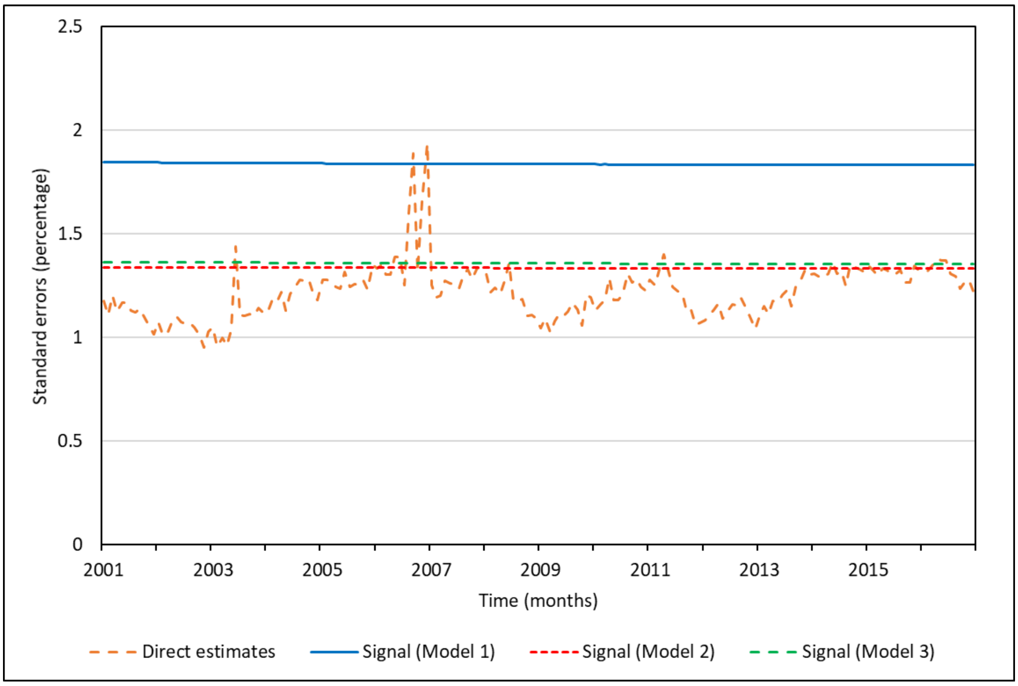

Figure 3.9: standard errors direct estimates and filtered estimates signal under three models for consumer confidence

It is a remarkable result that the standard error of the filtered signals are equal or even higher than the standard errors of the direct estimates. A general finding in the literature is that state-space models applied to series obtained with repeated surveys result in model estimates with standard errors that are substantially smaller compared to the standard errors of the direct survey estimates, see e.g. Pfeffermann and Bleuer (1993), Pfeffermann and Burck (1990), Pfeffermann and Tiler (2006), Van den Brakel and Krieg (2009, 2015, 2016), Boonstra and van den Brakel, (2019). The reason that this is not the case for the Dutch CS is as follows. For this explanation we need Table 3.5, where the standard deviation of the average measurement error in the composite series are calculated. These values are obtained by deriving the measurement error for the linear combinations of the baseline variables for economic climate, willingness to buy and consumer confidence from the standard deviations of the measurement errors in Table 3.1 and the correlations between the measurement errors in Table 3.3. Now, recall from Section 3.1 that the measurement error $e_{i,t}$ in equation (3.3) is the sum of the sampling error $\tilde{e}_{i,t}$ in equation (3.1) and the population irregular term $I_{i,t}$ in equation (3.2). The standard errors of the direct estimates, on the other hand, only contain the uncertainty of the sample design. It follows from Table 3.5 for economic climate that the average measurement error under the three models is substantially larger than the standard error of the direct estimate. This implies that the population parameter contains a white noise component that is at least as large as the sampling error. Although less extreme, this also applies to willingness to buy and consumer confidence. This uncertainty of the population white noise is reflected in the standard error of the measurement error of the time series model, but not in the standard error of the direct estimate. This is an indication that the questions of the CS measure a short-term emotion and are not interpreted by the respondents as a long term evaluation over the last and next 12 months of the economy and the financial situation. This observation is confirmed by the fact that the time series contain a seasonal pattern, which would not be present if questions are interpreted as the situation over the last 12 and next 12 months.

	M1	M2	M3	direct
Economic climate	4.44	3.02	3.59	1.80
Willingness to buy	1.58	1.50	1.51	1.40
Consumer confidence	2.26	1.50	1.70	1.30

Table 3.5: standard deviation of the average measurement error in the models and the standard error for the direct estimates of the composite series

Since the standard errors of the filtered signals are stable over time, the standard errors of the five baseline series and the three combined series are shown in Table 3.6 for December 2016 (last observation before the change-over to the new design). Model 2 has the smallest standard errors for all variables. For some variables, the differences are substantial. For the five baseline series, the standard error under Model 1 is slightly smaller than the one under Model 3. For Fin. N12 the difference is even quite large. For the combined series, however, the standard errors under Model 1 are larger than the ones under Model 3. From the comparison between Model 2 and Model 3, it follows that modelling cross-sectional correlations through the trend component improves the precision of the model estimates. Modelling the correlation between measurement errors decreases the precision, because the positive correlation between the measurement error inflates the variance of the measurement error of the combined series. It is, however, necessary to account for these correlations since the five baseline series are based on the same respondents. Ignoring correlated measurement error underestimates the uncertainty of the model predictions (Model 2 and 3).

	Model 1	Model 2	Model 3
Econ. L12	3.23	2.68	3.24
Econ. N12	4.78	3.97	4.85
Economic climate	3.61	2.64	2.92
Fin. L12	1.70	1.67	1.77
Fin. N12	1.73	1.59	1.93
Major pur.	2.17	2.16	2.23
Willingness to buy	1.20	1.14	1.15
Consumer confidence	1.83	1.33	1.35

Table 3.6: standard error filtered estimates signal last period (December 2016) for 8 series

A major advantage of inference based on time series models is that the gain in precision of period-to-period changes is large, compared to the direct estimates. To illustrate this, the standard errors of the month-to-month developments for the three combined series for the three models and the direct estimates are compared in Figures 3.10-3.12. The period-to-period change and their standard errors are obtained by calculating the linear combination of $\Delta_t = L_{t|t} - L_{t-1|t} + S_{t|t} - S_{t-1|t}$ via the Kalman filter recursion. For the direct estimates, these standard errors are larger than the ones in the Figures 3.7 – 3.9, since the direct estimates of two different periods in a cross-sectional survey are independent. For the model estimates however, these standard errors are smaller than the ones in the Figures 3.7 – 3.9, mainly due to the strong positive correlation between the trend levels of two subsequent periods. As a result, the precision of the estimates of the month-to-month-development is improved by all models for two of the three combined series. For economic climate and Model 1 and 2, the standard errors of the direct estimates and the model estimates are at the same level for the last 3 years. For more details see Figure 3.10.

The standard errors for the month-to-month-development under the three models and for the 5 baseline series and the three combined series are shown in Table 3.7 for December 2016. Since the standard errors are stable over time, it suffices to show results for the last observed period. As in Table 3.6, Model 1 is (slightly) more precise than Model 3 for the baseline series, and less precise for the combined series.

Note that the standard errors of the month-to-month-development are an approximation, since L_{t-1} and S_{t-1} are kept in the state-space system for one period to evaluate the linear combination of Δ_t. As a result $L_{t-1|t-1}$ and $S_{t-1|t-1}$ are updated with the information of month t and $L_{t-1|t}$ and $S_{t-1|t}$ are used in Δ_t instead. Therefore, the standard errors shown in the Figures 3.10 – 3.12 and in Table 3.2 slightly underestimate the true standard error of the difference of two filtered estimates.

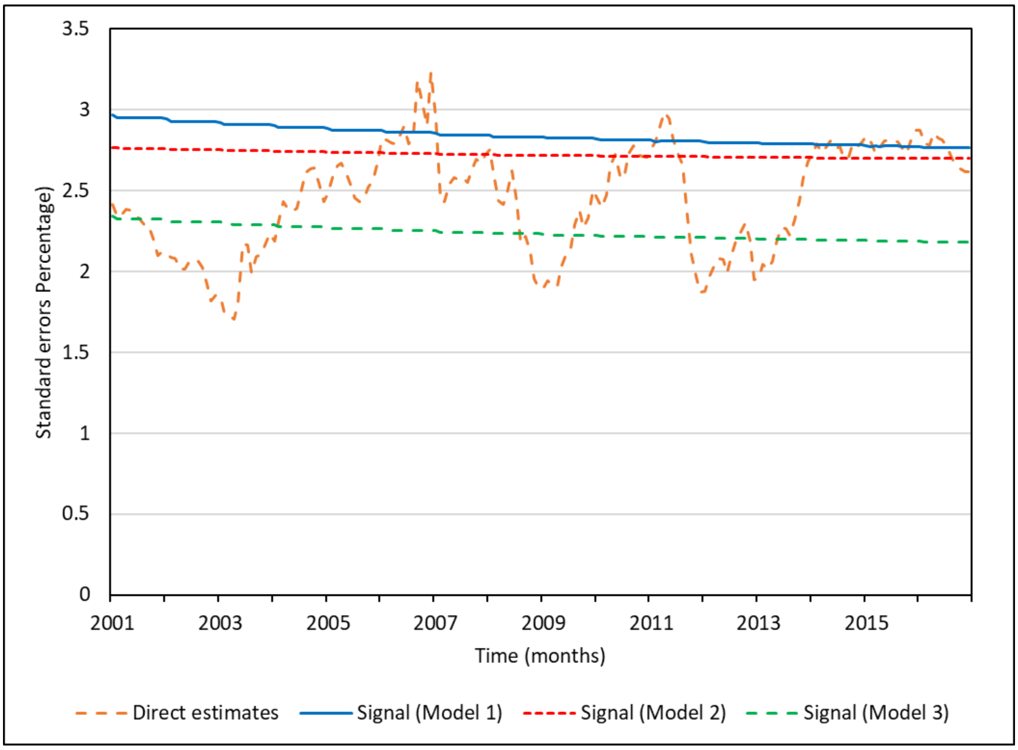

Figure 3.10: standard errors for month-to-month-development direct estimates and filtered estimates signal under three models, economic climate

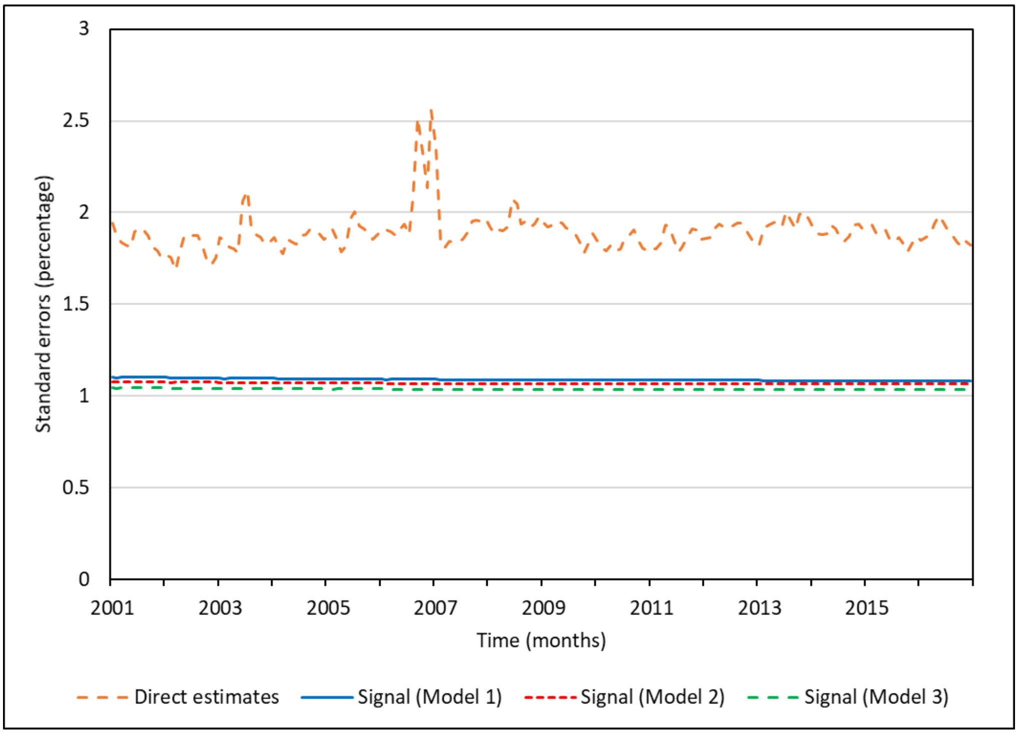

Figure 3.11: standard errors for month-to-month-development direct estimates and filtered estimates signal under three models, willingness to buy

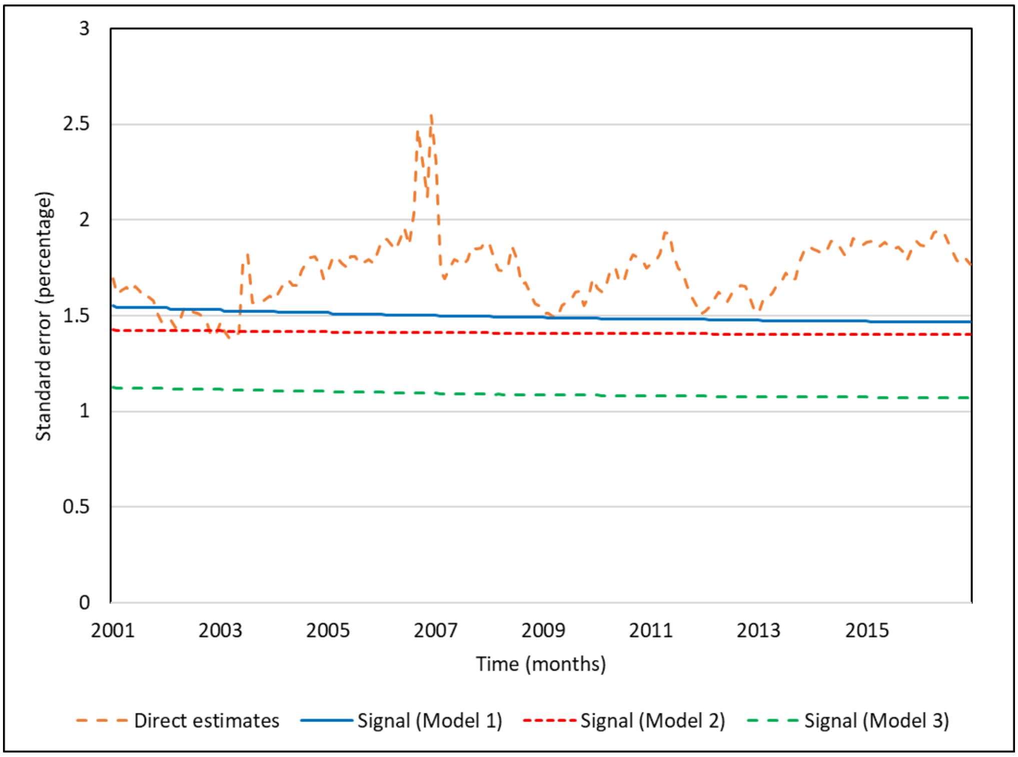

Figure 3.12: standard errors for month-to-month-development direct estimates and filtered estimates signal under three models, consumer confidence

	Model 1	Model 2	Model 3
Econ. L12	2.70	2.59	2.70
Econ. N12	3.33	3.34	3.43
Economic climate	2.77	2.70	2.18
Fin. L12	1.24	1.19	1.25
Fin. N12	1.48	1.48	1.59
Major pur.	2.31	2.30	2.35
Willingness to buy	1.08	1.06	1.03
Consumer confidence	1.47	1.40	1.07

Table 3.7: standard error for month-to-month-development filtered estimates signal last period (December 2016) for 8 series

3.3 Model evaluation

The assumptions underlying the state-space model are evaluated by testing whether the standardized innovations are standard normally and independently distributed, see Durbin and Koopman (2012), Sections 2.12 and 7.5. Different tests (Bowman-Shenton normality tests, F-tests for heteroscedasticity, QQ-plots, plots of standardized innovations and sample correlograms, Durbin Watson test) indicate some small violations of these assumptions under all models. The results on normality and heteroscedasticity are comparable under all models. The correlogram under Model 1 shows some autocorrelation of lag 1 for some of the series, which is slightly larger compared to Model 2 and 3. Since the input series are very long, even small violations of the model assumptions could be significant, but the violations we found here are acceptable.

The three models are also compared using AIC and DIC defined by Durbin and Koopman (2012), Section 7.4. Results are presented in Table 3.8. Under both criteria Model 1 is preferred. Since the three models are nested, a likelihood ratio test can be used for model comparison. Table 3.9 contains the results for the likelihood ratio tests for the three possible model comparisons.

	Log likelihood	AIC	DIC
Model 1	-5305.8	30.92	31.56
Model 2	-5360.3	31.17	31.70
Model 3	-5518.1	32.02	32.45

Table 3.8: AIC and DIC values for the three models

Comparison	LR statistic	df	p-value
Model 1 versus Model 2	109.08	10	0.000
Model 2 versus Model 3	315.52	10	0.000
Model 1 versus Model 3	424.60	20	0.000

Table 3.9: Results likelihood ratio tests

The test for Model 2 versus Model 3 indicates that modelling the correlation between the slope disturbance terms significantly improves the model fit. The test for Model 1 versus Model 2 shows that modelling the correlation between the measurement errors further improves the model fit significantly. Finally the test for Model 1 versus Model 3 shows that the joint test on the inclusion of a full covariance matrix for the slope disturbance terms and the measurement errors rejects the null hypothesis that both models are equivalent.

On the one hand, Model 1 shows some (small) violations of the model assumptions, on the other hand, the model comparison statistics show that Model 1 is preferable. Furthermore, there are arguments that the measurement errors are correlated. By adding the correlations between the slope disturbances, information from other series is used to improve the accuracy of the estimates. A model that allows for correlated slope disturbance terms must also allow for correlated measurement errors. Otherwise, correlated measurement errors in all input series could be incorrectly interpreted as a true development of the trend instead of measurement errors (sampling noise or noise in the population parameter). In conclusion, Model 1 is selected for the inference of the CS.

4. Discontinuities

4.1 Estimation methods of discontinuities

The most straightforward approach to quantify discontinuities, is to collect data under the old and new survey design alongside each other for some period, i.e. a parallel run. A parallel run is preferably designed as a randomized experiment, where the sampling units from a probability sample are randomized over the old and new survey designs such that the subsamples can be considered as the treatment groups in an experiment.

In the case of a sufficiently large parallel run, contrasts between direct or design-based sample estimates under the old and new survey design can be used as estimates for the discontinuities, using design-based inference procedures for experiments embedded in probability samples (Van den Brakel et al., 2013, Van den Brakel, 2008). Alternatively, discontinuities can be estimated with a structural time series model that contains a level intervention. This means that the time series model is extended with a regression component for which the corresponding auxiliary variable switches from zero to one at the moment the survey is transferred from the old to the new design. Under the assumption that the other components of the time series model (trend and seasonal) describe the evolution of the population parameter correctly, the regression coefficient of this auxiliary variable can be interpreted as an estimate for the discontinuity. The level intervention approach with state-space models was originally proposed by Harvey and Durbin (1986) to estimate the effect of seat belt legislation on British road casualties. Van den Brakel et al. (2008) and Van den Brakel and Roels (2010) apply this approach to estimate discontinuities induced by a redesign of a sample survey process. Without a parallel run, the regression coefficient of the level intervention would be initialized diffuse in the Kalman filter. When sufficient observations are available under the new design, the Kalman filter eventually gives a stable estimate for the discontinuity. The minimum required period mainly depends on the flexibility of the trend component.

Information obtained from a parallel run can be combined with the state-space level intervention approach, by using the direct estimate for the discontinuity and its variance as an exact initialization for the regression coefficient of the level intervention in the Kalman filter. In this way, the discontinuity estimate from the parallel run is improved with the available information from the entirely observed time series (Van den Brakel and Krieg, 2015). This approach is followed in this paper. The available budget allowed a parallel run of three months in the first quarter of 2017 for the transition of the CS to the new design, where the sample sizes for both designs were equal to the normal net sample size of around 1000 persons.

Discontinuities are estimated for the percentage of positive, neutral and negative answer categories of the five questions, i.e. p_i^+, p_i^0, p_i^- for $i = 1, \ldots, 5$. These discontinuity estimates are used to compute uninterrupted series for the percentages $p_{i,t}^+$, $p_{i,t}^0$, $p_{i,t}^-$ for $i = 1, \ldots, 5$, by adjusting the series observed before the change-over to the new design to the level of the series observed under the new design. These corrected series are used in a second step to calculate uninterrupted series for $y_{1,t}, \ldots, y_{5,t}$. These backcasted series will be used as the input for model (3.4) that is used in the production of official monthly figures about consumer confidence. Discontinuities are estimated and corrected for the percentages series, since they are the variables measured through the questionnaire und used to derive the input series. As a result, discontinuities occur on the level of the estimated percentages. The size of the discontinuities in the input series depends on the size of the discontinuities of these percentages and the share of the percentages in the input series. Since the share of the percentages in the input series changes substantially over time, a discontinuity adjustment based on the percentages will be more realistic compared to an approach that directly estimates and adjusts discontinuities at the level of the input series.

Based on the parallel run there are two estimates for the variables $p_{i,t}^+$, $p_{i,t}^0$, $p_{i,t}^-$ which are denoted $\hat{p}_{i,t}^{j,\mathrm{O}}$ $(j \in \{+,0,-\})$ for the estimate of $p_{i,t}^j$ under the old design and $\hat{p}_{i,t}^{j,\mathrm{N}}$ for the estimate of $p_{i,t}^j$ under the new design, for the first three months of 2017. Direct estimates for the discontinuities are obtained as:

$$\widehat{\Delta}(p_i^j) = \frac{1}{3}\sum_{t=2017(1)}^{2017(3)}\left[\hat{p}_{i,t}^{j,\mathrm{N}} - \hat{p}_{i,t}^{j,\mathrm{O}}\right], \text{ for } j \in \{+,0,-\}. \quad (4.1)$$

From the definition of $\hat{p}_{i,t}^{j,\mathrm{O}}$ and $\hat{p}_{i,t}^{j,\mathrm{N}}$ it follows that

$$\sum_{j\in\{+,0,-\}} \widehat{\Delta}(p_i^j) = 0\,. \quad (4.2)$$

The variance of the estimates of the discontinuities can be estimated by

$$\widehat{\mathrm{Var}}[\widehat{\Delta}(p_i^j)] = \frac{1}{9}\sum_{t=2017(1)}^{2017(3)}[\widehat{\mathrm{Var}}(\hat{p}_{i,t}^{j,\mathrm{O}}) + \widehat{\mathrm{Var}}(\hat{p}_{i,t}^{j,\mathrm{N}})], \text{ with} \tag{4.3}$$

$$\widehat{\mathrm{Var}}(\hat{p}_{i,t}^{j,\mathrm{O}}) = \frac{1}{n_t^{\mathrm{R}}}\hat{p}_{i,t}^{j,\mathrm{O}}(100 - \hat{p}_{i,t}^{j,\mathrm{O}}),\ \widehat{\mathrm{Var}}(\hat{p}_{i,t}^{j,\mathrm{N}}) = \frac{1}{n_t^{\mathrm{N}}}\hat{p}_{i,t}^{j,\mathrm{N}}(100 - \hat{p}_{i,t}^{j,\mathrm{N}}),$$

and $n_t^{\mathrm{O}}, n_t^{\mathrm{N}}$ the sample size in month t under the old and new design.
In a next step a three dimensional multivariate model is applied to the three series with percentages of positive, negative and neutral answers. The model is applied to each question separately, i.e. for $i = 1, \dots, 5$. The index i is omitted in the formulas. For each of the three series the basic structural time series model is extended with a level intervention, i.e.

$$\widehat{\mathbf{p}}_t = \mathbf{L}_t + \mathbf{S}_t + \boldsymbol{\beta}'\mathbf{x}_{\mathrm{t}} + \mathbf{e}_t, \tag{4.4}$$

where $\widehat{\mathbf{p}}_t = (\hat{p}_t^+, \hat{p}_t^0, \hat{p}_t^-)'$ is the vector of direct estimates of the percentages, until 2016 based on the old design and from 2017 based on the new design, $\mathbf{L}_t = (L_t^+, L_t^0, L_t^-)'$ is the vector of the trends, $\mathbf{S}_t = (S_t^+, S_t^0, S_t^-)'$ is a vector of the seasonal patterns, $\boldsymbol{\beta} = (\beta^+, \beta^0, \beta^-)'$ the estimates for the discontinuities, $\mathbf{x}_{\mathrm{t}} = (x_t, x_t, x_t)'$ the level intervention variable, i.e. x_t switches from zero to one in January 2017 when the new design is implemented. Finally $\mathbf{e}_t = (e_t^+, e_t^0, e_t^-)'$ is a vector containing the measurement errors.

The variables L_t^j and S_t^j with $j \in \{+,0,-\}$ are smooth trend models and trigonometric seasonal models as described in Section 3.1. For the disturbance terms it is assumed that they are mutually independent, normally distributed with expectation zero and time-independent variance components. From (4.2) it follows that the coefficients for the discontinuities, β^+, β^0 and β^-, must obey the restriction that they add up to zero. This is enforced with the following transition equations in the state-space model:

$$\begin{array}{lcr} \beta_t^+ & = & \beta_{t-1}^+ \\ \beta_t^- & = & \beta_{t-1}^- \\ \beta_t^0 & = & -\beta_{t-1}^+ - \beta_{t-1}^- \end{array} . \tag{4.5}$$

The subscript t indicates the notation of the transition equations. As there is no disturbance term, $\boldsymbol{\beta}$ is still time-independent.

The measurement errors $\mathbf{e}_t$ represent the sum of the sampling errors and the white noise in the population parameter. The sampling error depends on the sample size, which is approximately constant over time, and the percentage $\hat{p}_t^j$, with $j \in \{+,0,-\}$, as the variance of the direct estimate $\hat{p}_t^j$ is $\widehat{\mathrm{Var}}(\hat{p}_t^j) = \frac{\hat{p}_t^j(100 - \hat{p}_t^j)}{n}$ with $n \approx 1000$ the sample size.

To account for heterogeneity in the sampling error, the following variance structure is assumed for the measurement errors:

$$\mathrm{E}(e_t^j) = 0,$$

$$\mathrm{Cov}\left(e_t^j, e_{t'}^j\right) = \begin{cases} \hat{p}_t^j(100 - \hat{p}_t^j)\sigma_{\mathrm{e},j}^2 & if \quad t = t' \\ 0 & if \quad t \neq t' \end{cases}, \text{ for } j \in \{+,0,-\}. \tag{4.6}$$

When the estimate of $\sigma_{\mathrm{e},j}^2$ is around $\frac{1}{1000}$, the noise in the series is explained by the sampling error. This is the case for some of the series considered here. For some other series, the estimate of $\sigma_{\mathrm{e},j}^2$ is much larger (around $\frac{1}{300}$). This means that in these series there is substantial noise in the population parameter.

As mentioned before, the information obtained with the parallel run is used in the time series model using an exact initialization of the state variables $\boldsymbol{\beta}$ in the Kalman filter. This implies that these regression coefficients are initialized with the direct estimates for the discontinuities in (4.1) and the variances of the filtered estimates are initialized with the estimates given by equation (4.3). In this way the Kalman filter improves the direct estimates for the discontinuities obtained with the parallel run with the information available from the time series observed before and after the parallel run.

4.2 Correction methods for discontinuities

As explained in Subsection 4.1, the series of the percentages observed before the redesign of January 2017 are corrected to the level of the percentage series observed under the new design. These so-called backcasted percentage series are used to compile backcasted input series for the time series model (3.4). The percentages can only have admissible values in the range [0,100]. Therefore correction methods are considered that result in backcasted series that have values in this admissible range. In that case the backcasted target variables $\hat{y}_1, \dots, \hat{y}_5$ will also have values in the admissible range [-100, +100]. Let $\tilde{p}_{i,t}^{j,\mathrm{N}}$ denote the backcasted series of $\hat{p}_{i,t}^{j,\mathrm{O}}$. The first approach to backcast percentages is based on the following correction:

$$\tilde{p}_{i,t}^{j,\mathrm{N}} = \hat{p}_{i,t}^{j,\mathrm{O}} + \hat{\beta}_i^j \frac{\hat{p}_{i,t}^{j,\mathrm{O}}(100 - \hat{p}_{i,t}^{j,\mathrm{O}})}{\hat{p}_{i,\tau}^{j,\mathrm{O}}(100 - \hat{p}_{i,\tau}^{j,\mathrm{O}})} . t = 1, \dots, T-1, \tag{4.7}$$

with T the month of the change-over to the new design, i.e. April 2017. Furthermore, $\hat{p}_{i,\tau}^{j,\mathrm{O}}$ denotes an estimate under the old design obtained during the entire period of parallel run (denoted by τ). The estimated discontinuity $\hat{\beta}_i^j$ is multiplied by a factor proportional to the variance of the percentage, estimated by $\hat{p}_{i,t}^{j,\mathrm{O}}(100 - \hat{p}_{i,t}^{j,\mathrm{O}})$. The correction is zero when $\hat{p}_{i,t}^{j,\mathrm{O}} = 0$ or $\hat{p}_{i,t}^{j,\mathrm{O}} = 100$, and it is maximal when $\hat{p}_{i,t}^{j,\mathrm{O}} = 50$. Furthermore, $\hat{p}_{i,\tau}^{j,\mathrm{O}}$ is the estimated percentage under the old design in the period of the parallel run. Dividing by the population variance of this period, i.e., the mean of the three months, makes sure that the corrected percentage estimate obtained for these three months are close to the values observed under the new design during the parallel run.

When all three percentages $\hat{p}_t^j$ for $j \in \{+,0,-\}$ are corrected with (4.7) the sum of the corrected percentages is no longer 100. Since the neutral percentage $\hat{p}_t^0$ is not used in the computation of the indicators y_i , this percentage is corrected as $\tilde{p}_{i,t}^{0,\mathrm{N}} = 100 - \tilde{p}_{i,t}^{+,\mathrm{O}} - \tilde{p}_{i,t}^{-,\mathrm{O}}$. The correction in (4.7) diminishes when the percentage $\hat{p}_{i,t}^{j,\mathrm{O}}$ is close to 0 or 100. It is nevertheless not guaranteed that the values of $\tilde{p}_{i,t}^{0,\mathrm{N}}$ are in the admissible range of [0,100]. They can take values outside this range when the percentages during the parallel run, represented by $\hat{p}_{i,\tau}^{j,\mathrm{O}}$, are close to 0 or 100 and the discontinuity β_i^j is large. Note that the underlying assumption of the proportional correction method is that the discontinuity is small when the variance is small. This assumption cannot be true in this case.

The second approach is a log-ratio transformation, which forces that the adjusted values will always take values in the range [0,100] and that the sum of the three categories is exactly 100. Since the neutral percentages $\hat{p}_{i,t}^{0,d}$, for $d \in \{\mathrm{O},\mathrm{N}\}$, are not used in the computation of the indicators y_i, the log-ratio transformation for the CS is chosen to be

$$z_{i,t}^{+,d} = \ln\left(\frac{\hat{p}_{i,t}^{+,d}}{\hat{p}_{i,t}^{0,d}}\right),\ z_{i,t}^{-,d} = \ln\left(\frac{\hat{p}_{i,t}^{-,d}}{\hat{p}_{i,t}^{0,d}}\right), \text{ for } d \in \{\mathrm{O},\mathrm{N}\}. \tag{4.8}$$

The discontinuities are now defined by

$$\widehat{\Delta}(z_i^j) = \frac{1}{3}\sum_{t=2017(1)}^{2017(3)}\left[\hat{z}_{i,t}^{j,\mathrm{N}} - \hat{z}_{i,t}^{j,\mathrm{O}}\right], \text{ for } j \in \{+,-\} \tag{4.9}$$

and the variance of $\widehat{\Delta}(z_i^j)$ can be computed as

$$\mathrm{Var}[\widehat{\Delta}(z_i^j)] = \frac{1}{9}\sum_{t=2017(1)}^{2017(3)}[\mathrm{Var}(\hat{z}_{i,t}^{j,\mathrm{N}}) + \mathrm{Var}(\hat{z}_{i,t}^{j,O})]. \tag{4.10}$$

By Taylor linearization it can be shown that

$$\widehat{\mathrm{Var}}\left[\hat{z}_{i,t}^{j,d}\right] = \widehat{\mathrm{Var}}\left[\ln\left(\frac{\hat{p}_{i,t}^{j,d}}{\hat{p}_{i,t}^{0,d}}\right)\right] \approx \frac{\left(100-\hat{p}_{i,t}^{j,d}\right)}{n_t^d\hat{p}_{i,t}^{j,d}} + \frac{\left(100-\hat{p}_{i,t}^{0,d}\right)}{n_t^d\hat{p}_{i,t}^{0,d}} + \frac{2}{n_t^d}, \text{ for } j \in \{+,-\} \text{ and } d \in \{\mathrm{O},\mathrm{N}\}. \tag{4.11}$$

Final estimates for the discontinuities are obtained by applying a two dimensional version of time series model (4.4) to the transformed series $\hat{\mathbf{z}}_{i,t} = (\hat{z}_{i,t}^{+,d}, \hat{z}_{i,t}^{-,d})^t$. The regression coefficients for the level interventions are initialized in the Kalman filter with (4.9) and (4.10). Restriction (4.5) for the discontinuity estimates does not apply under this transformation.

The transformed series under the old design are corrected additively for the discontinuity:

$$\tilde{z}_{i,t}^{+,\mathrm{N}} = \hat{z}_{i,t}^{+,\mathrm{O}} + \hat{\beta}_i^+ \text{ and } \tilde{z}_{i,t}^{-,\mathrm{N}} = \hat{z}_{i,t}^{-,\mathrm{O}} + \hat{\beta}_i^-. \tag{4.12}$$

With the following transformations the corrected percentages can be computed:

$$\tilde{p}_{i,t}^{j,\mathrm{N}} = 100\frac{e^{\tilde{z}_{i,t}^{j,\mathrm{N}}}}{e^{\tilde{z}_{i,t}^{+,\mathrm{N}}} + e^{\tilde{z}_{i,t}^{-,\mathrm{N}}} + 1}, j \in \{+,-\} \tag{4.13}$$
$$\tilde{p}_{i,t}^{0,\mathrm{N}} = 100\frac{1}{e^{\tilde{z}_{i,t}^{+,\mathrm{N}}} + e^{\tilde{z}_{i,t}^{-,\mathrm{N}}} + 1}.$$

These corrected series of percentages can be used to compute the corrected indicators. A drawback of the log-ratio transformation is that the effect of the correction can become very large when the numerator of the ratio, i.e., $\hat{p}_{i,t}^{+,\mathrm{O}}$ or $\hat{p}_{i,t}^{-,\mathrm{O}}$, is smaller than the denominator $\hat{p}_{i,t}^{0,\mathrm{O}}$. This will be demonstrated in Section 4.3.

4.3 Results estimation of discontinuities

For the Dutch Consumer Survey a parallel run was performed in the first three months of 2017. Table 4.1 shows the results of the parallel run for the variable Econ. L12. For every answer option the percentages are given under the old and new design together with their differences. In this period most of the respondents were positive about the economic situation. The percentage of "a little better" turns out to substantially increase under the new design. Also the percentage of "a little worse" increases, but to a smaller extent. The percentages of the other answer options all decrease.

We assume that part of these changes can be explained by the changes in the questionnaire. Under the old design the respondent could, in addition to the neutral options, only choose between "better" and "worse". When, according to the respondent, the situation was changed only a little, the options "better" or "worse " did probably not feel appropriate, not knowing that "a little worse" or "a little better" are also possible answers. The respondent then probably chose one of the neutral answers. Under the new design the respondent chooses "a little worse" or "a little better" and during the parallel run mostly "a little better" more frequently. We do not have an explanation for the fact that the options "a lot better" and "a lot worse" are chosen less often under the new design.

	January			February			March			
	reg.	new	diff.	reg.	new	diff.	reg.	new	diff.	mean diff.
a lot better	14.1	7.0	-7.1	14.9	7.8	-7.1	16.4	9.9	-6.5	-6.9
a little better	31.3	51.1	19.8	31.1	50.0	18.9	32.4	48.6	16.2	18.3
the same	35.7	25.6	-10.1	35.0	25.8	-9.2	34.0	25.7	-8.3	-9.2
a litte worse	7.1	10.0	2.9	6.0	10.5	4.5	6.6	8.4	1.8	3.0
a lot worse	6.1	3.4	-2.7	6.9	3.1	-3.8	5.8	3.2	-2.6	-3.1
do not know	5.8	2.9	-2.9	6.1	2.9	-3.2	4.8	4.2	-0.6	-2.2

Table 4.1: results parallel run for economic situation last 12 months

Table 4.2 shows the mean differences of the percentages for three other variables. The results are similar to the economic situation in the last 12 months (Table 4.1). The percentages of "a little better" and "a little worse" increase, while the percentages of the other answer options decrease. For these three variables the decrease for “a lot better” and “a lot worse” is smaller than for the economic situation in the last 12 months.

	Econ. N12	Fin. L12	Fin. N12
a lot better	-0.8	-2.2	0.0
a little better	17.6	10.9	11.1
the same	-7.7	-11.6	-11.3
a litte worse	1.8	7.6	4.8
a lot worse	-1.7	-4.4	-1.7
do not know	-9.2	-0.3	-2.8

Table 4.2: results parallel run, mean differences estimates old and new design for three variables

Table 4.3 displays the direct estimates of the discontinuities in the percentages for all questions based on the parallel run. We see that the question major purchases, the only question where the questionnaire is not changed, is the only question where the discontinuity for the positive answers is negative, and smaller than for the other questions.

	positive answers		negative answers		difference
Econ. L12	11.4	(1.3)	-0.1	(0.9)	11.5
Econ. N12	16.8	(1.2)	0.1	(0.8)	16.7
Fin. L12	8.7	(1.0)	3.2	(1.1)	5.5
Fin. N12	11.1	(1.0)	3.1	(0.9)	8.0
Major pur.	-4.7	(1.2)	0.1	(0.8)	-4.8

Table 4.3: results parallel run, estimates discontinuities, standard errors in brackets

Now model (4.4) is applied to further improve the estimates of the discontinuities obtained with the parallel run. The series start in April 1986 and run on up to February 2020. Up to and including March 2017 the estimates are based on the old design and starting from April 2017 they are based on the new design. The estimates of the discontinuities based on this model are shown in the Tables 4.4 and 4.5. In Table 4.4 a diffuse initialization of the regression coefficients of the level intervention in the Kalman filter is used, which implies that the information from the parallel run is ignored. The results in Table 4.5 are obtained with an exact initialization of the regression coefficients of the level intervention in the Kalman filter using the direct estimates and their standard errors obtained in the parallel run from Table 4.3.

	positive answers		negative answers	
Econ. L12	10.3	(3.0)	0.1	(2.4)
Econ. N12	20.2	(3.5)	0.8	(3.1)
Fin. L12	9.7	(1.2)	2.2	(1.5)
Fin. N12	12.2	(1.2)	4.6	(1.5)
Major pur.	-6.5	(1.8)	1.3	(1.5)

Table 4.4: estimates discontinuities based on STM with diffuse Kalman filter initialization, standard errors in brackets

	positive answers		negative answers	
Econ. L12	11.2	(1.1)	-0.1	(0.8)
Econ. N12	17.3	(1.2)	0.2	(0.7)
Fin. L12	9.1	(0.8)	3.0	(0.9)
Fin. N12	11.7	(0.7)	3.5	(0.8)
Major pur.	-5.1	(1.0)	0.3	(0.7)

Table 4.5: estimates discontinuities based on STM with exact Kalman filter initialization, standard errors in brackets

The point estimates in the Table 4.4 and 4.5 are comparable to the direct estimates based on the parallel run in Table 4.3. In the case of the exact prior the point estimates are closer to the direct estimates than in the case of a diffuse prior. When a diffuse initialization is used (Table 4.4) the standard errors are substantially larger than the standard errors of the direct estimates (Table 4.3). These large standard errors are caused by the dynamics of the CS series, which are quite volatile indeed. This results in particular in a flexible trend. The influence of observations on the estimates of the discontinuities depends on the flexibility of the trend. As the flexibility of the trend increases, the estimates for the discontinuities are based more on the local observations obtained directly before and after the transition to the new design. In this application, model estimates for the discontinuities are therefore not very accurate without a parallel run. The direct estimates for the discontinuities based on the parallel run are more precise compared to the time series estimates without the information of a parallel run, which follows from Table 4.3 and 4.4. When the results of the parallel run are combined with the time series modelling approach through an exact initialization of the Kalman filter, then the most precise estimates for the discontinuities are obtained, since all available information from the observed time series before and after the change-over and the parallel run are combined. The improvement of the precision with respect to the time series model with a diffuse initialization is substantial. With respect to the direct estimates of the parallel run there is only a slight improvement of the precision of the discontinuity estimates. This might, however, be different in other applications where the series are less volatile.

The estimates of the discontinuities obtained with the time series model are instable if only a few observation after the change-over are available and improve when more date under the new design become available. Tables 4.4 and 4.5 show the most accurate estimates that can be computed using the data observed until February 2020. Figures 4.1 and 4.2 show the real time estimates for the discontinuities based on the data from April 2017 until February 2020 for the percentages of respectively the positive and negative answers for Econ. L12. The figures show how the estimates of the discontinuities evolve when more date become available. It can be seen that the estimates are in the right order of magnitude from the beginning. Nevertheless, there are some visible changes in the first six months. In the case of the exact initialization the changes are much smaller. In the first six months the standard errors of the estimates decrease, where the decrease for the diffuse prior is substantial. After this period both the point estimates and the standard errors are stable.

It is difficult to conclude in general how many observations under the new design are required before a stable estimate for the discontinuities is obtained, since this depends on the volatility of the series and the flexibility of the trend. The influence of observations further away from the moment of the change-over increases as the trends become less volatile. Therefore, the minimum number of observations under the new design increases as the flexibility of the trend decreases.

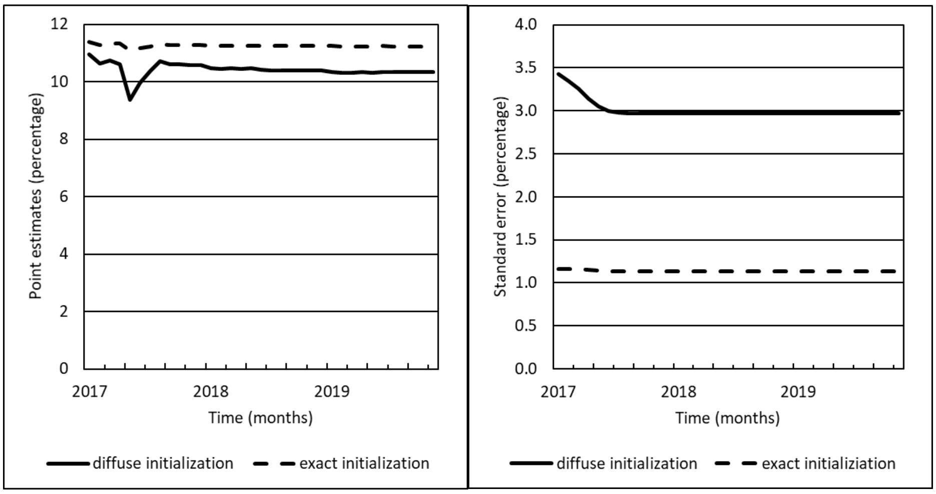

Figure 4.1: development of point estimates (left panel) and standard errors (right panel) of discontinuities percentages of positive answers, Econ. L12.

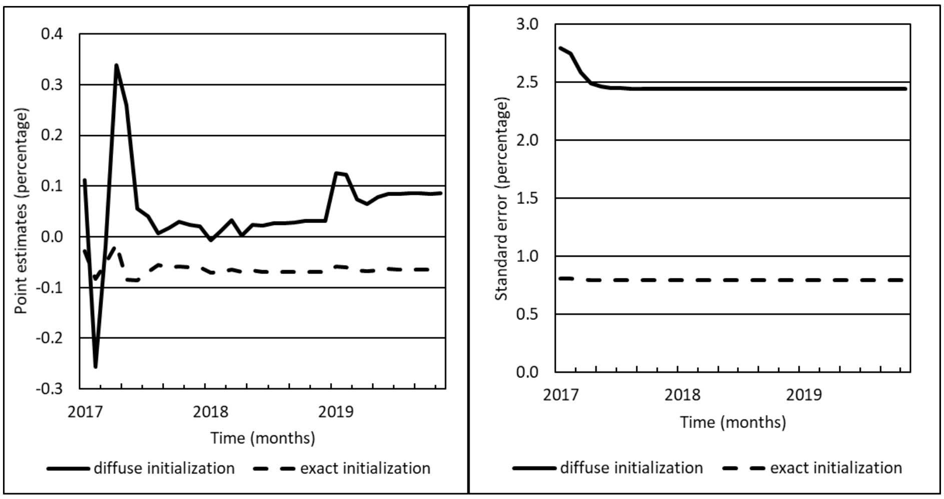

Figure 4.2: development of point estimates (left panel) and standard errors (right panel) of discontinuities of percentages negative answers, Econ. L12

4.4 Results correction for discontinuities

In this section the effects of the correction methods described in Section 4.2 on the indicator series of the CS are investigated for the variables Econ. L12 and Major pur. The discontinuity estimates are based on the time series model (Model 1) with an exact initialization of the Kalman filter, since this approach makes optimal use of all available information from the parallel run and the observed time series.

Figure 4.3 shows the indicator series under the old design for Econ. L12 from January 2001 up to March 2017 together with the series corrected by the proportional method defined by (4.7) and the log-ratio transformation. The discontinuities for this variable are $\widehat{\Delta}(p_1^+) = 11.2$ and $\widehat{\Delta}(p_1^-) = -0.1$ (Table 4.5) in a period where the consumer confidence is positive. Figure 4.3 shows that in positive periods both methods correct in more or less the same way, i.e., the adjusted series becomes more positive and the corrections are equal. In negative periods the log-ratio transformation makes the adjusted series more negative, while the proportionally corrected series stays close to the original series. For this variable the log-ratio correction has a larger effect than the proportional correction, particularly in periods where the economic situation over the last 12 months is negative, which seems to be less plausible.

Figure 4.4 shows the original and corrected series for the Major pur. Also for this variable the discontinuities are rather small, but compared to Econ L12 they have opposite signs: $\widehat{\Delta}(p_5^+) = -5.1$ and $\widehat{\Delta}(p_5^+) = 0.3$. For Major pur. both correction methods give similar results: the adjusted series are smaller than the original series and the effect of both corrections is the same. For this variable there is no preference for one of the two correction methods.

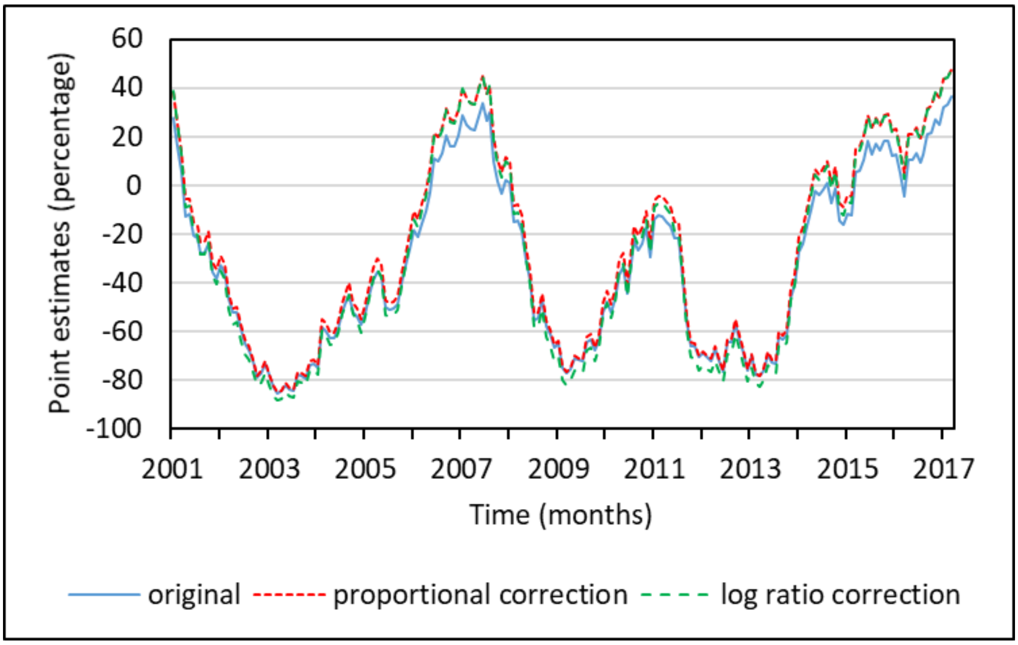

Figure 4.3: comparison of backcasting methods for indicator Econ. L12. Estimates of discontinuities are based on STM with an exact initialization.

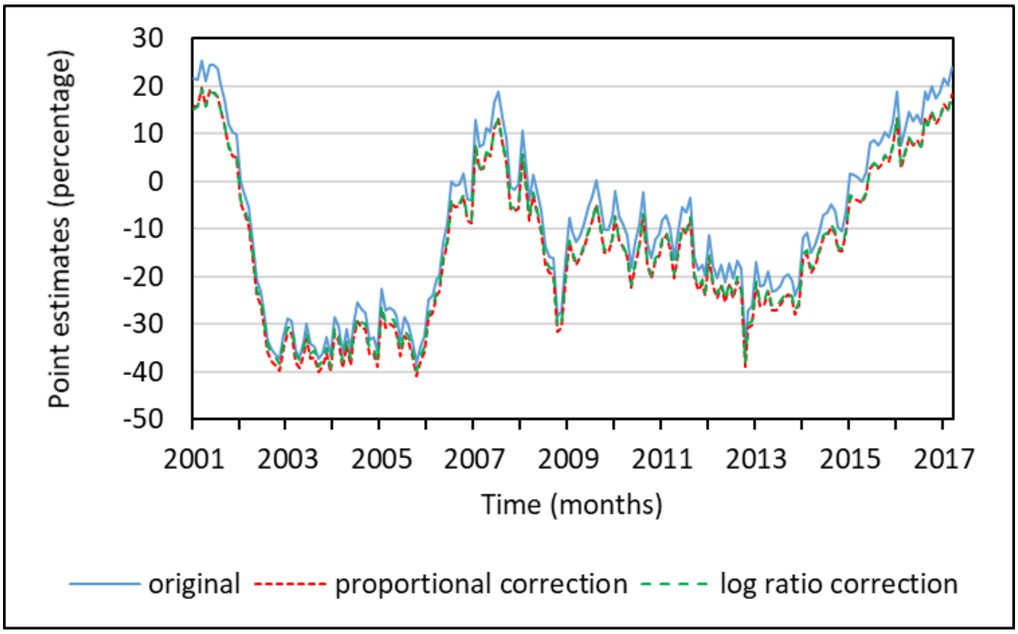

Figure 4.4: comparison of backcasting methods for indicator of Major pur. Estimates of discontinuities are based on STM with exact prior

From these results there is a slight preference for the proportional correction, since the log-ratio transformation sometimes leads to larger corrections. A disadvantage of the proportional correction is that the corrected values could fall outside the range of $[-100, +100]$. However, this only happens for unrealistically large values of the discontinuities as is shown in Figure 4.5. Here the corrected series are computed for Econ. L12, where the fictive discontinuities are given by $\widehat{\Delta}(p_1^+) = -10$ and $\widehat{\Delta}(p_1^+) = 20$. In this situation the proportional correction leads to outcomes smaller than -100 in the most negative periods. In this case the log-ratio correction would be preferred.

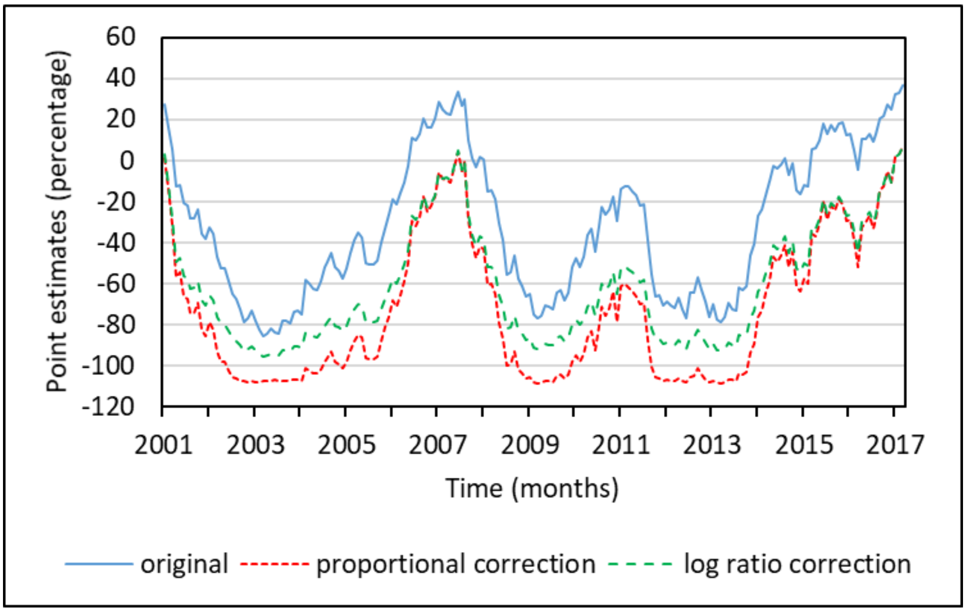

Figure 4.5: comparison of backcasting methods for indicator of Econ. L12 and fictive discontinuities: $\widehat{\boldsymbol{\Delta}}(\mathbf{p}_1^+) = -\mathbf{10}$ ***and*** $\widehat{\boldsymbol{\Delta}}(\mathbf{p}_1^-) = \mathbf{20}$

Figure 4.6 shows that also with realistic values of the discontinuities the log-ratio transformation could result in large corrections. Here the correction methods are applied to the indicator series of Major pur. where the discontinuities have fictive but realistic values of $\widehat{\Delta}(p_5^+) = -5$ and $\widehat{\Delta}(p_5^-) = -10$. In some periods the log-ratio correction is small (for example around 2007), but is extremely large in other periods (for example in 2003-2006 and 2009-2015). Furthermore, the log-ratio corrected series is always positive, even in periods where the original series is negative. This shows the disadvantage of the log-ratio correction, namely that correction can become very large when the ratio of the original figure is much smaller than 1. This effect is also visible for the proportional correction, but to a lesser extent. From these results we conclude to use the proportional correction for backcasting the input series $\hat{y}_1, \dots, \hat{y}_5$.

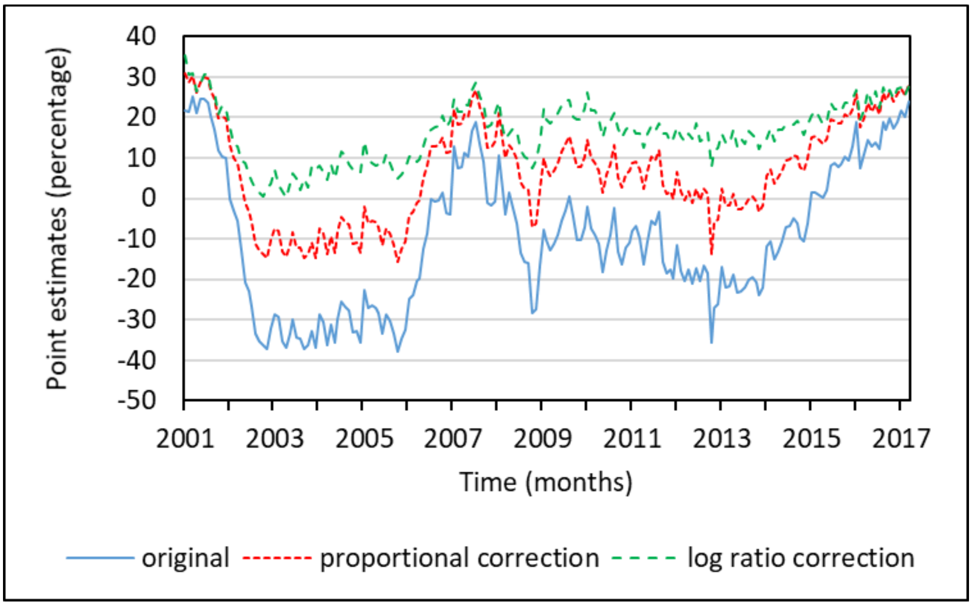

Figure 4.6: comparison of backcasting methods for indicator of Major pur. and fictive discontinuities: $\mathbf{\widehat{\Delta}(p_5^+) = -5}$ ***and*** $\mathbf{\widehat{\Delta}(p_5^-) = -10}$

5. Official publications based on STM

Since 2017 the time series model developed in this paper is implemented in the production process of the Dutch CS. The series for $\hat{p}_{i,t}^{j,O}$ are corrected for the discontinuity for the years until December 2016. Based on these corrected series, baseline series $\hat{y}_1, \dots, \hat{y}_5$ are computed to obtain backcasted series for the period until December 2016 that are corrected to the level of the new design. These series are extended with estimates under the new design until the current month. With these series, the structural time series model (3.4) is estimated. The filtered trend estimates are published and replace the former seasonally corrected figures. These series of filtered estimates, corrected for the discontinuity, starts in 1986.

Filtered estimates for month t are published at the end of the month. Under this time series modelling approach it is possible to update or revise the filtered estimates if more data become available in t+1, t+2, … These estimates are more precise but require that figures have to be revised after the first publication. This is a disadvantage in official statistics, as this can be confusing for the users of the publications. In the Dutch CS it is decided not to revise the filtered estimates published in t+1. For a similar reason the discontinuity estimates used to construct backcasted input series $\hat{y}_1, \dots, \hat{y}_5$ are based on the parallel run only. Improving the discontinuity estimates with the time series model proposed in Section 4 implies that initial backcasted series, required for the production of the regular CS figures in the period directly after the change-over to the new design are revised after one year. Since the updates of the direct estimates of the discontinuities with the time series model are relatively small in this application, it was decided to accept less optimal estimates for the discontinuities and avoid publishing revisions after one year.

In Figure 5.1 the unadjusted and adjusted direct estimates for consumer confidence are compared with the filtered trend estimates obtained with time series model (3.4). The direct estimates observed under the old design (dashed line) are plotted until December 2016. These series are adjusted to the level observed under the new design, using the backcasting procedure proposed in Section 4.2 for the series of the percentage distributions across the positive, neutral, and negative response categories of the five questions underlying the consumer confidence index. The backcasting approach results in a positive adjustment of which the size fluctuates over time. The adjusted series of consumer confidence (solid red line) is continued with the direct estimates observed under the new design after January 2017. As a result an uninterrupted series of direct estimates for the monthly consumer confidence is obtained. The filtered trend (solid blue line) replaces the seasonally adjusted figures in the official publications of the Dutch CS and can be interpreted as a trend-cycle component obtained by removing seasonal fluctuations, population irregular term and sampling error from the series of direct estimates. For the period before January 2017, the trend is also corrected to the level observed under the new design, since time series model (3.4) uses the five backcasted baseline indices as input series. As a result uninterrupted trend series are obtained for the period starting in 1986.

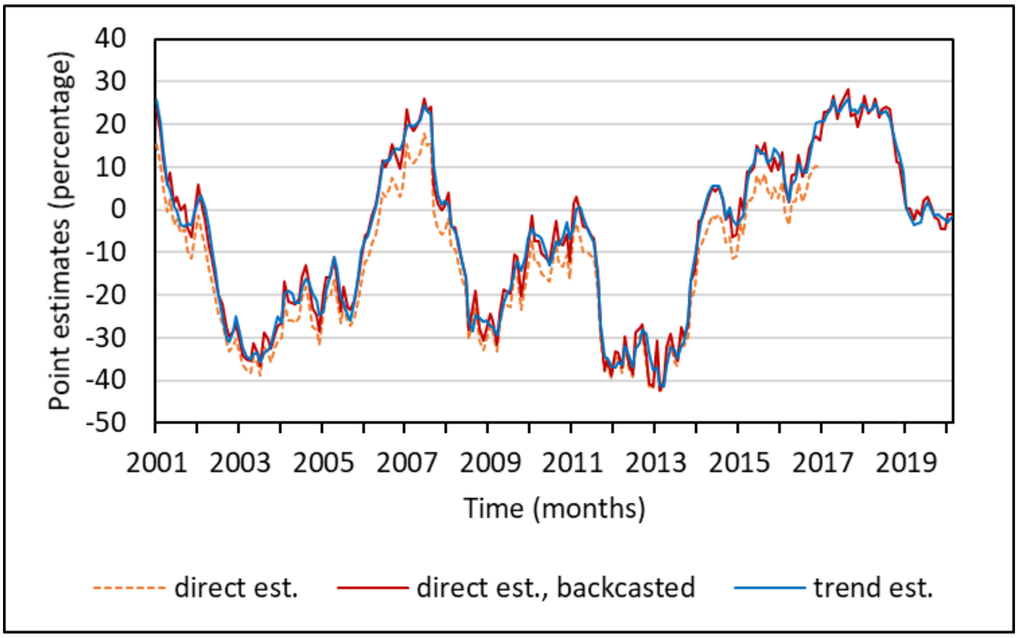

Figure 5.1: comparison direct estimates (original and backcasted) with trend estimates, as published, consumer confidence.

6. Conclusion

In this paper a model-based inference method for the Dutch Consumer Survey (CS) is developed. The method is based on a multivariate structural time series model where the monthly direct estimates of five baseline indices for consumer confidence are used as input. These indices are 1) opinion economic situation last 12 months, 2) opinion economic situation next 12 months, 3) opinion financial situation last 12 months, 4) opinion financial situation next 12 months, and 5) whether it is a right moment to make major purchases. Each index is obtained from the difference between the percentage of respondents with a positive and a negative opinion on that subject. As a result each index can vary between -100 and 100. From the five baseline indices, three composite indices are derived, which are an indicator for economic climate (average of the first two indices), indicator for willingness to buy (average of the last three indices) and the indicator for consumer confidence (average over all five indices). The time series model is developed to produce more accurate monthly consumer confidence indicators and to estimate and correct for discontinuities induced by a change-over to a new survey design in April 2017. The method was implemented for the production of official monthly CS figures in April 2017 and has been used since that month.

The time series model assumes for each input series a separate structural baseline model that consists of a smooth trend model, a trigonometric seasonal model, and a measurement error. The measurement error contains the sampling error and the population irregular term. A full covariance matrix is assumed for the slope disturbance terms of the trends and for the measurement error. For the seasonal disturbance terms a diagonal covariance matrix is assumed. Modelling the correlation between the measurement error is essential, since the five input series are based on variables measured on the same sample units. Model-based estimates including standard errors for the three composite indicators are obtained by calculating the average over the trend and signal (trend plus seasonal) of the relevant baseline indicators. They are derived as linear combinations from the filtered state vector and its covariance matrix.

The time series model produces optimal model-based estimates for the monthly CS figures by using sample information observed in previous reference periods and relations between the five input series. The standard errors of the model estimates for the composite indicators are nevertheless larger than the standard errors of the direct estimates. This finding is not in line with the findings for small area time series methods in the literature and is the result of the following two reasons. In this application the population irregular term is of the same order as the sampling error. The uncertainty of the population irregular term is reflected in the standard error of the time series model estimates but not in the standard error of the direct estimates. Furthermore, the positive correlations between the measurement errors particularly increase the standard errors of the composite indices. It is nevertheless essential to model the correlation between the measurement errors since this results in a better separation of the measurement error from trend and signal.

The survey redesign in 2017 resulted in discontinuities in the series of the Dutch CS. To separate real month-to-month changes from sudden differences in measurement and selection bias due to the redesign, discontinuities are estimated by collecting data under both the old and the new design in parallel during the first quarter of 2017. The parallel run results in direct estimates for the discontinuities. The precision of these estimates can be improved with a time series model where the discontinuity is modelled with a level shift and an exact initialization is applied for the regression coefficient of this level shift in the Kalman filter using the direct estimates and their standard errors obtained in the parallel run. In this way the information observed with the time series before and after the change-over is used to further improve the direct estimates for the discontinuities.

Discontinuities appear in the percentage distribution across the positive, neutral, and negative response categories obtained with the questions about the economic situation, the financial situation and major purchases. The discontinuities in the five baseline indices that are derived from these percentages depend on the percentage distribution across these three categories. These distributions, and therefore the discontinuities in the baseline series, varied strongly over time. It was therefore decided to estimate and model the discontinuities in the time series of the percentage distribution across the positive, neutral, and negative response categories. To maintain uninterrupted series, the time series of these percentages observed under the old design are backcasted to the level of the new design. An adjustment method that accounts for the fact that the valid range of adjusted values of the proportions is in the range [0,100] is proposed. Under this method the size of the adjustment in each period is made proportional to the variance of the estimated proportion. Alternatively a log-ratio transformation can be considered. The drawback of the log-ratio transformation is that the adjustments can be extremely large in periods where the value of the original ratio is smaller than 1. For this reason the percentages series are backcasted using the method where the adjustment is proportional to the variance of the observed percentages. The results obtained with this adjustment method seem more plausible. The validity of this adjustment method, however, cannot be verified and is based on the strong assumption that the required adjustment is proportional to the variance of the observed percentages.

Backcasted series of the five baseline indices are derived from the backcasted series of the percentage distribution across the positive, neutral, and negative response categories. In this way the method correctly accounts for differences in adjustments over time, due to the fact that the composition of respondents over the positive, neutral, and negative response categories changes over time. These adjusted series are extended with the direct estimates observed after April 2017 and are used as input series for the multivariate time series model to estimate monthly trends for the CS. This method has been implemented in April 2017 for the publication of uninterrupted direct monthly CS figures and trends that start in 1986. The trend estimates replace seasonal adjusted figures and can be interpreted as a trend-cycle derived from the direct estimates from which seasonal component, population irregular component and the sampling error is removed.

The lockdown in March 2020 due to the Covid-19 pandemic resulted in strong shocks in the input series of the time series model and resulted in a temporary model-misspecification. The methods considered to accommodate in the model for the sudden changes in the dynamics of the underlying series and how this was handled in the production process of the CS, are described in a separate paper.

References

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

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

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

Blight, B.J.N. and A.J. Scott (1973). A Stochastic Model for Repeated Surveys. Journal of the Royal Statistical Society, B series, vol. 35, pp. 61-66.

Boonstra, H.J. and J.A. van den Brakel (2019). Estimation of level and change for unemployment using structural time series models. Survey Methodology. Vol. 45, pp. 395-425.

Brakel, J.A. van den, en S. Krieg (2009). Estimation of the Monthly Unemployment Rate through Structural Time Series Modelling in Rotating Panel Design. Survey Methodology, 16, 2, 177-190.

Brakel, J.A. van den, en S. Krieg (2015). Dealing with small sample sizes, rotation group bias and discontinuities in a Rotating Panel Design. Survey Methodology, 41, 2, 267-296.

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

Brakel, J.A. van den, E. Söhler, P. Daas en B. Buelens (2017). Social media as data source for official statistics; the Dutch Consumer Confidence Index. Survey Methodology, 43, 2, 183-210.

Brakel, J.A. van den (2013). Design-based analysis of factorial designs embedded in probability samples. Survey Methodology, 39, 232-349.

Brakel, J.A. van den, X. Zhang, and S.-M. Tam (2020). Measuring discontinuities in time series obtained with repeated sample surveys. International Statistical Review, 88, pp.155-175.

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

Doornik, J.A. (2007). An Object-oriented Matrix Programming Language Ox 5. London: Timberlake Consultants Press.

Durbin, J., en Koopman, S.J. (2012). Time series analysis by state space methods. Oxford: Oxford University Press.

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

European Commission, Directorate-General for Economic and Financial A_airs (2014). A user manual to the joint harmonised eu programme of business and consumer surveys. Eu-publication.

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

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

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

Koopman, S.J., N. Shephard and J.A. Doornik, (1999). Statistical Algorithms for Models in State Space using SsfPack 2.2. Econometrics Journal, 2, 113-166.

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

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

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

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

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

Rao, J.N.K. and I. Molina (2015). Small Area Estimation, 2nd edition. Wiley.

Rao, J.N.K., and M. Yu (1994). Small area estimation by combining time series and cross-sectional data, Canadian Journal of Statistics, 22, pp. 511-528.

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

Tam, S.M. (1987). Analysis of repeated surveys using a dynamic linear model. International Statistical Review, 55, pp. 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, pp. 149-166.

Explanation of symbols

Empty cell	Figure not applicable
.	Figure is unknown, insufficiently reliable or confidential
*	Provisional figure
**	Revised provisional figure
2017–2018	2017 to 2018 inclusive
2017/2018	Average for 2017 to 2018 inclusive
2017/’18	Crop year, financial year, school year, etc., beginning in 2017 and ending in 2018
2013/’14–2017/’18	Crop year, financial year, etc., 2015/’16 to 2017/’18 inclusive

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

Colophon

Publisher
Centraal Bureau voor de Statistiek
Henri Faasdreef 312, 2492 JP Den Haag
www.cbs.nl

Prepress
Statistics Netherlands, CCN Creation and visualisation

Design
Edenspiekermann

Information
Telephone +31 88 570 70 70, fax +31 70 337 59 94
Via contactform: www.cbsl.nl/information