Statistics Netherlands

Discussion Paper

Data Related Issues with Compsite Indicators: Availability, Frequency and Adjustment Techniques

Paper prepared as chapter 3 of the UN Handbook on Composite Indicators, G.-L. Mazzi and A. Ozyildirim editors., in preparation. The views expressed in this paper are those of the authors and do not necessarily reflect the policies of Statistics Netherlands.
The authors thank the editors and anonymous referees for valuable suggestions and comments.

2015 | 06

Jan van den Brakel
Sabine Krieg
Pim Ouwehand
Frank van de Pol
Floris van Ruth
Marc Smeets
Piet Verbiest

Index

1. Overview

It is the role of official statistics to produce undisputed figures, being a benchmark and safe haven in an ever extending sea of data. Constructing long and yet consistent time series is part of this task, as well as composing indicators from varying sources.

The main determinants of the quality of long series of composite indicators are

- Stability of research designs,
- Timeliness and proper definitions of data from several sources or regions,
- Equal frequency of data sources,
- Adequate derivation of indices and
- Wise decisions on outliers or turning points in seasonal adjustment.

In this chapter we focus on each of them. Firstly, components may not be stable enough due to changing classifications or some sort of a redesign. The problem can be mended with methods that are mentioned in section 2. We give special attention to models for interrupted time series. An introduction to structural time series modelling is in an appendix.

Another cluster of problems concerns incomplete or delayed data. Official statistics cannot always collect data according to the desired definition. Sometimes alternative data are freely available that can be used as an imperfect substitute or proxy. Two examples of alternative data are given in section 3. We also pay attention to modelling data that come in too late or that are missing for some regions or branches of a classification.

Section 4 is about combining indicators of mixed frequency. We point out briefly some possibilities to do so and furthermore refer to the abundant literature.

Indices pose specific problems. An important issue is the detrimental effect of collecting data for only the first part of the month in order to speed up publication. In section 5 we also direct attention to pitfalls with chain estimators and to the debate on how to cope with changing weights of index components, that is with a chain-linked or fixed base method.

Finally, sharp turns in a time series pose problems to seasonal adjustment. When should an outlier be considered as the beginning of a sharp turn? In section 6 we plead for a conservative approach.

In this chapter we will often refer to model based approaches. However, it should be kept in mind that a good presentation of the information is vital. Modern animated graphics offer special possibilities.. An example is the business cycle tracer (Van Ruth c.s., 2005) which shows several indicators simultaneously on their way through time. Such a business cycle clock is now being presented by several statistical agencies, among which Eurostat, the OECD and the statistical offices of Denmark, Germany, Korea and the Netherlands.

2. Quantifying discontinuities in sample surveys

2.1 Introduction

In order to forecast economic cycles, time series of macro-economic indicators should be as long as possible. Macro-economic indicators are typically based on repeatedly conducted sample surveys conducted by national statistical institutes. One important quality aspect of figures published with repeatedly conducted sample surveys is comparability over time. To produce consistent series, national statistical institutes generally keep their survey processes unchanged as long as possible. It remains inevitable, however, to redesign survey processes from time to time to improve the quality or the efficiency of the underlying survey process.

Question wording of course affects the outcomes. Also response categories, the way in which these are presented (with or without interviewer, lay-out of the questionnaire, paper or web), the sequence of questions, previous questions, introductory text, explaining text, the length of the questionnaire; they all have effects on the outcomes, on the amount and direction of bias. Factual questions are less sensitive for these effects. The less latitude the respondent has for interpretation, the better. For enterprise surveys, these problems concentrate on internationally standardised nomenclatures, like the activity codes of enterprises (ISIC, NACE) and systems for international trade in goods (HS, CN, SITC), transport codes (NST) and classifications of services (MSITS: United Nations Statistical Division, 2010). See Eurostat (2009). Changes therein contribute to and constitute changes in measurement errors and bias.

Secondly, selective response is influencing the outcomes. Also if a better, more complete, sampling frame becomes available, for an investments survey for instance, then estimates of investments will increase suddenly and the new figures are no longer comparable with the old time series.

Thirdly, alterations in data processing are influential. The introduction of a new imputation method, or an efficient data editing method like top-down editing, or macro editing, all are potential causes of irregularities in the time series. For a classification of sources of errors in registers, we refer to Zhang (2012).

A redesign of a survey process generally affects the different sources of measurement bias in a survey and therefore has a systematic effect on the outcomes of the survey. These kinds of effects are referred to as discontinuities. If the method changes in either of the aspects mentioned above,

i. data collection instrument,

ii. target population or response percentage,

iii. data processing,

it should be checked whether the time series of target variables have been interrupted.

With respect to the measures we can take, we have several options, some of which are not applicable in specific situations. The best option is to apply parallel data collection following both designs, the old and the new. This may be done with a very important figure, like unemployment, for instance after the introduction of a web questionnaire.

A less expensive option is available if only classification variables are affected. In the case of the introduction of a new version of the NACE for instance, we may aim at calculating outcomes with both versions of the classification variable. This implies that we need a questionnaire that supports both classifications in the transition period. Moreover, for a good comparison, sample size in publication cells should be large enough for both the old and the new classification. Therefore, in some cells a larger sample size will be needed in the transition period.

A less expensive alternative for a parallel data collection or parallel recalculation is to "glue" the series before and after the interruption with a step parameter in a time series model. In order to estimate the step parameter one needs an accurate estimate for the first time point next to the interruption and for that a sufficiently long time series is needed. If available, an auxiliary time series that is strongly correlated with the target series may help to bridge the gap more accurately. This approach will be treated in detail in the following subsections.

Lastly, we have of course the zero-option to do nothing. For less important target variables, a statistical office may decide to communicate that the time series has been interrupted and a new series has been started, because specific methodological changes or improvements were necessary. In such a case, the new time series unfortunately may not be comparable with the old series on the same target variable.

However, in an ideal survey transition process, the systematic effects of the redesign are explained and quantified to avoid confounding real developments of the indicators of interest with discontinuities induced by the redesign of the survey process. This keeps series consistent and preserves comparability of the outcomes over time. To sum up, one or more of the following four options may be available,

a. parallel data collection with the old and the new design,

b. parallel data processing in the old and the new way,

c. apply an interrupted time series model,

d. do nothing and take the blame for a discontinuity in the time series.

In survey methodology, time series models are frequently applied to develop estimates for periodic surveys. Blight and Scott (1973) and Scott and Smith (1974) proposed to regard the unknown population parameters as a realization of a stochastic process that can be described with a time series model. This introduces relationships between the estimated population parameters at different time points in the case of non-overlapping as well as overlapping samples. The explicit modelling of this relationship between these survey estimates with a time series model can be used to combine sample information observed in the past to improve the precision of estimates obtained with periodic surveys. This approach is frequently applied in the context of small area estimation, see e.g. Rao and Yu (1994), Pfeffermann and Burck (1990), Pfeffermann and Bleuer (1993), Pfeffermann et al. (1998), and Pfeffermann and Tiller (2006). Time series models are also appropriate to account for differences in measurement bias in rotating panels (Pfeffermann, 1991; Van den Brakel and Krieg, 2009) and for discontinuities induced by a survey redesign, (Van den Brakel, Smith and Compton, 2008 and Van den Brakel and Roels, 2010). Some other key references to authors that applied the time series approach to repeated survey data to improve the efficiency of survey estimates are Scott et al. (1977), Tam (1987), Binder and Dick

(1989, 1990), Bell and Hillmer (1990), Tiller (1992), Harvey and Chung (2000), Feder (2001), and Lind (2005).

Before structural time series models emerged, a regression approach to time series has been common practice (Ostrom, 1978), and it still is widely used. This approach has developed throughout the years into RegArima models, coping with autocorrelation by taking first order differences among other things. Intervention parameters can be estimated also (Box and Tiao, 1975). Nowadays this RegArima approach is often used as a first step, including estimation of discontinuities, before seasonal adjustment filters are applied (Bureau of the Census, 2013; Caporello and Maravall, 2004; Grudkowska, 2013). However, for modelling discontinuities we have a preference for structural time series models because they offer both good interpretability of parameters, like in seasonal adjustment heuristics, as well as statistical models that can be tested.

The need as well as the various possibilities to quantify discontinuities, is discussed in subsection 2.2. In an appendix to this chapter, a general framework for modelling discontinuities in time series observed through repeatedly conducted surveys is constructed. There we elaborate on the state-space representation of these models. Also the Kalman filter, commonly used to analyse state-space models, is briefly reviewed. It is also indicated how external quantitative information about discontinuities can be used in the model. Appropriate software to estimate state-space models is briefly reviewed.

2.2 Quantifying discontinuities in sample surveys

Changes in a survey process can have large systematic effects on the outcomes of a survey. An experiment with a new questionnaire for the Dutch Labour Force Survey (LFS) in 2000 showed that the unemployed labour force, defined as the ratio of the total unemployed labour force to the total labour force, dropped from 3.1 to 2.4 percentage points (Van den Brakel and Van Berkel, 2002). An experiment with alternative data collection procedures in the LFS, conducted in 2001, showed that the estimated unemployed labour force under computer assisted telephone interviewing is 1.1 percent points smaller compared to computer assisted personal interviewing, Van den Brakel (2008). Finally an experiment conducted to estimate discontinuities due to a major redesign of the LFS in 2010 showed that a change-over from computer assisted personal interviewing to a mix of computer assisted telephone and personal interviewing in combination with the introduction of a new questionnaire increases the estimated total unemployed labour force from 420.000 to 475.000 people, Van den Brakel and Krieg (2012). These three examples illustrate that modifications in a survey process can have substantial and significant effects on the figures of a survey. It is therefore crucial that discontinuities due to modifications of a survey process are quantified and that data users account for these effects in their analyses.

There are various possibilities to quantify the effect of a survey redesign, see Van den Brakel, Smith and Compton (2008) for an overview. The required methods depend on the phase of the survey process that is changed. In cases where the underlying sample data remain the same, the differences can be investigated by recalculation, for example the introduction of new editing, imputation or estimation methods. Also a new economic activity classification system in business surveys will generally result in discontinuities in time series. An example is the change-over from the NACE Rev 1.1 to the NACE Rev. 2 in 2008. The effect of such a new classification system can be quantified by adding the new classification system to all units in the sample frame or at least in the sample. As a result, a double coded sample or sampling frame is obtained for a particular reference period, and appropriate design-based domain estimators can be constructed under both classification systems. This enables the quantification of the effects of the change-over. See Van den Brakel (2010) for details and an application to the change-over from the NACE Rev 1.1 to the NACE Rev. 2.

A bottleneck in this approach may be that data editing has to be carried out twice too, requiring more statisticians than are actually available. Then a simplified procedure may be followed, involving a transformation of publication figures. Results according to the old NACE after data editing can be distributed over relevant cells of the new NACE. In order to do so an enterprise transition matrix from the old classification to the new classification has to be constructed that holds the proportions that stay and that move to specific cells.

When data collection procedures are affected, the new micro data are not consistent with the past. In these cases the effect of the change-over can be quantified by conducting a field experiment where the regular and new survey designs are run concurrently. Under a well-designed parallel run, two estimates for the variables of interest are obtained for the same reference period; one under the regular and one under the new approach. The contrast between both estimates is a direct estimate for the discontinuity induced by the redesign of the survey. See e.g. Van den Brakel (2008) for applications and details of designing and analysing this kind of large scaled field experiments. The old design may be maintained for several periods with a sample that is small, but just big enough to discern relevant changes in the level of unemployment. If explanations are to be given for the changes that occur, a larger sample will be needed in order to be split up in explanatory subsamples, like age groups.

A parallel run is not always tenable for a national statistical institute due to budget constraints. Moreover, doubling the data collection is not possible with enterprise surveys, as the largest enterprises are included in the sample with probability one. It is not feasible to ask these enterprises to fill out two different questionnaires on the same subject. In such cases a time series modelling approach can be considered as an alternative. In this case the evolution of the series of the variable of interest is modelled with an appropriate time series model. To quantify the effect of the redesign, an intervention variable that describes the moment of the change-over from the old to the new design is added to this model. Under the assumption that the components of the time series model, other than the intervention component, describes the evolution of the series reasonable well, the regression coefficient of the intervention variable will measure the systematic effect induced by the redesign of the survey.

In an appendix to this chapter, we will give an introduction to the use of a state-space framework for modelling time series observed with repeatedly conducted survey samples. Intervention models are developed to construct uninterrupted series for the target variables. It is discussed how additional information about discontinuities, for example from a parallel run, can be included as prior information in these models.

2.3 Adjusted series, back-casting

In the appendix, a state-space framework to construct consistent series is proposed. In some cases, however, a simple regression approach may also suffice (Ostrom, 1978). After having analyzed the intervention model, the observed series can be adjusted or corrected with the estimated discontinuities, in a way comparable to constructing seasonally adjusted data. In the case of a level intervention, for example, the time series after the moment of the survey transition can be adjusted for the estimated discontinuities with $\tilde{y}_{t,k} = \hat{y}_{t,k} - \hat{\beta}_k$, with $\tilde{y}_{t,k}$ the new series and $\hat{\beta}_k$ the amount in which the new series exceeds the old one in category or stratum k. Adjusting the new time series cannot be continued for a very long time, of course. Therefore we prefer the alternative, that is to adjust the series before the survey transition, with $\tilde{y}_{t,k} = \hat{y}_{t,k} + \hat{\beta}_k$. This may be called back-casting. As stated before, an auxiliary series can be added to the model as a proxy, in order to get a more accurate estimate of the intervention step. When for instance a redesign has interrupted the series from a structural business survey, VAT figures may be incorporated in the structural time series model to help bridge the gap.

However, adjusting the old series and favoring the new series implies that we have to argue why the new figures are better, which may be hard when redesigns are motivated by cutting costs. Therefore it is important that official statistics maintain or improve quality when redesigning surveys. Quality can be measured partially as representativety of the responding part of the sample (Schouten et al., 2011).

As an alternative for a correction with $\hat{\beta}_k$, filtered or smoothed estimates for the target parameters can be used. Intervention models explicitly account for discontinuities. Filtered estimates for the target variables are therefore not affected by the systematic effect of the change-over.

The two aforementioned approaches are useful if consistent series are required to produce uninterrupted series of indices or if consistent macro indicators are required as input for other models used to analyze economic cycles.

If the state-space framework itself is used to analyze economic cycles, then the intervention components developed in this chapter can be included in the model to avoid model-misspecification and distortion of the estimated economic cycles by the discontinuities in the input series.

For back-casting in the context of RegArima models we refer to Sartore and Caporin (2006). These authors focus among other things on aggregated time series for the EU. When one of the EU-countries revises its series, this invokes recalculation of the back-casting exercise for the aggregated EU series.

3. Missing, delayed or incomplete information: what alternatives do we have?

Composite indicators cannot be built without some sort of statistical information. However, we don't always have the right information, that can straightforwardly be added up or raised to population figures. The Organisation for Economic Co-operation and Development (OECD, 2008) gives a good exposition about various imputation methods as the solution for missing data problems. In addition, we want to highlight situations where imputation is less relevant. If there is no direct information on your target variable, maybe proxy information is available which is known to be highly correlated to the figure that we want to estimate. Secondly, if the desired information becomes available after a long time, that is if it comes too late, auxiliary variables that become available sooner may be used for a provisional estimate of the current situation. Finally, sample size of a survey may be too small for a detailed breakdown. Information on a specific region or stratum may even be completely missing. In that case a highly correlated variable from a register, or data from similar strata or areas, may be helpful to obtain more detail. If there are no data on a specific area and no better option is available, even a completely synthetic estimate could be made. We consider three cases to illustrate different solutions, 1) newly realised buildings, 2) consumer confidence and 3) the order book of establishments. The section ends by reflecting on administrative data versus surveys.

3.1 Newly realised buildings

The best option to monitor the building production is to maintain a register of addresses, of course. Such a register should be kept up to date by municipalities or any other type of institution that earns money from lending services to these addresses, like maintaining roads and other infrastructure, providing energy, collecting garbage, etc. Nations can simplify administrative tasks for such enterprises by offering one unifying generic register of addresses.

The ideal to have a timely and complete address register is difficult to reach, though. If an address register is not available, a survey may be set up, asking local municipalities or counties to assess the current stock of houses and the proportion that was built after the last occasion the survey was held, for instance ten years ago.

These figures can be updated with a yearly survey on building permits that local municipalities have granted. With these updates, the fact should be taken into account that not every permit will result in a completed building. In some highly regulated countries, people have the right to object against building permits for a specific period of time. To counteract this risk, sometimes more than one version of a building permit is delivered. Moreover, adverse economic conditions may inhibit and delay building plans. Hence an estimate is needed of the fraction of the building permits that will be realised within the first, second and later years. Also the limited scope of a survey on building permits should be taken into account. Usually a lower threshold of for instance €50.000 is maintained and temporary buildings are ignored.

In addition to the building permits survey one may monitor the amount of concrete that is being used in order to estimate an increase or decrease in the percentage of building permits that results in a completed building. Also the turnover of building enterprises, and short term changes therein, known from VAT data, can be helpful to assess this parameter. We will discuss administrative data in general at the end of this section.

3.2 Consumer confidence and social media data

Because consumer confidence is an early indicator for business tendency, several countries carry out a survey on this subject. Daas and Puts (2014) recently found that the time series of consumer confidence is strongly correlated with the sentiment in part of the social media, especially Facebook and to a lesser extent Twitter. Consumer confidence seems to be leading and social media sentiment lags a bit behind. These findings seem promising, but more research has to point out whether consumer confidence can be derived from social media on a regular basis, for instance weekly, in the future. If so, only a yearly survey would be needed to corroborate and tweak the estimate.

This is just an example of the potential of "big data". For instance Google Trends and Google Correlate offer a toolbox for trend watching and exploring the internet for statistical purposes. The underlying methods are also referred to as "data science".

3.3 New orders for establishments

Another useful indicator for economic prospects is the value of the order book of enterprises. "Industrial new orders are among the leading series used for the widely monitored OECD Composite Leading Indicator", state De Bondt et al. (2013). Because this information is no longer an obligatory part of the Eurostat short term statistics (STS), the European Central Bank (ECB) continues the series, notwithstanding the fact that some countries have stopped collecting these "hard" data, due to budget cuts and pressure to reduce the administrative burden.

An interesting question is whether data from countries where the hard information is still available, can be used to obtain valid estimates for the countries where these data are lacking. This problem can be considered as a special case of small area estimation (SAE). In small area estimation (EuroStat, 2013) information on an intermediate level, like region or branch of industry, is used to reinforce survey estimates. Especially in the areas where few survey observations are available, the model based estimate will get a large weight compared to the survey estimate. In the extreme case where no information is available one may rely completely on the model based prediction, which will then be called a synthetic estimate. Most literature on SAE is on data where every observation is equally important, as with persons and households. A reference specific for business surveys, where the data are not normally distributed, but skewed, is Krieg, Blaess and Smeets (2012).

Instead of applying the rather involved SAE methods, or in a preparatory phase before doing so, one may make common sense estimates for countries or regions where hard information on the order book is missing. This may be done for instance by selecting comparable areas, taking a stratification by branch of industry into account. Also other early indicators for the tendency of the market could be taken into account, such as vacancies and investments.

The ECB has developed illuminating time series models that cope with partially missing information about investments. They do not only rely on hard information on new orders, which lacks for some countries, but also use information that is available for all countries, although maybe a bit less leading. Next to the leading survey indicator "managers' assessment of the current level of orders books", also the hard "Industrial turnover index in manufacturing" is used as a second indicator. Several transformations, such as lags and first order differences, have been tested and incorporated into the model where useful. De Bondt et al. (2013) also report various robustness checks.

3.4 Administrative data or surveys?

In order to reduce the administrative burden, data collecting agencies are put under pressure to replace surveys with the use of administrative data as much as possible. This means that instead of survey variables that can be defined in any appropriate way, administrative data are obtained, which have to be accepted as they are. Restrictive is also that registers generally offer a small number of variables which are relevant for statistics, whereas the number of variables in a survey can be larger.

An advantage of administrative data is that registers offer many, sometimes almost all, members of a population, whereas surveys, due to cost restrictions, concern only a limited sample. Hence with surveys weighing and minimising sampling inaccuracy are important issues, whereas with administrative data selectivity of the missing part of the population is an important issue.

Various mixed designs emerged, bringing along estimation issues. For instance one may rely on VAT data for smaller enterprises, whereas a survey with a longer questionnaire is used for the larger enterprises. Another hybrid design emerges when a yearly survey sets the levels and an administrative source, like VAT, is used for short term extrapolation. However, some administrative sources are too slow to be used for a short term indicator. Therefore some countries use a survey to catch the short term general trend.

Not only survey data, but also administrative data can hold erroneous information (Zhang, 2012). Register holders will have a keen eye for their core variables, like value added in a tax setting or the number of employees in a social security setting. However, register holders will be less strict when it comes to for instance the timing of transactions or the categorisation of employees as temporary or permanent. Needless to remind the reader that survey respondents make errors too.

In sum, switching to administrative data, in part or completely, poses many different additional problems compared to surveys.

4. Mixed frequency data and delayed results

Policy makers rely on several indicators for the business cycle, but these indicators will not become available simultaneously. Indicators that appear monthly can be available earlier than quarterly indicators. Moreover, some monthly indicators require a longer fieldwork period or more processing time than others. This ragged edge of the available time series, some of them already available in the most recent month, and others not, is a problem that has been tackled by time series modelling in various ways recently.

Quarterly data can be transformed to monthly data in various ways. The time series analysis program Eviews has a tutorial for frequency conversion that shows several options, one of them being linear interpolation. Other methods of temporal disaggregation are Chow-Lin interpolation (Chow and Lin, 1971) and the Denton method (Denton, 1971). Next to temporal disaggregation, one can also overcome the multi-frequency problem by aggregation to the slowest, that is yearly, frequency, but this does not provide very recent now-casts, as yearly figures become available late.

In order to take advantage of the latest information in the "ragged edges", bridge models are often used for now-casting (Foroni and Marcellino, 2013). In this approach the high frequency, usually monthly, data are used to create forecasts or a now-cast for the current month. The equations for forecasting may then be used to extend quarterly time series with estimates for the latest months. Durbin and Quenneville (1997) proposed a state-space framework to benchmark time series observed on a monthly frequency to more reliable series observed on a quarterly or annual frequency.

A more recent development is the use of distributed lag polynomials for mixed data sampling (MIDAS). Alternatively, the state space models that were described at the beginning of this chapter may be used, extended to mixed-frequency vector autoregression (VAR) and factor models. An advantage of the latter type of models is that estimates for various frequencies can be forced to be consistent. For details the reader is referred to Foroni and Marcellino (2013) and the literature and software mentioned therein.

Statistical agencies also rely on imputation of the missing, most recent, data, as a practical alternative for structural or reduced form statistical models. Statistical imputation can be performed with simple time series models or other ways to impute or extrapolate the missing data with instrumental variables. Whereas statistical models give more opportunity to compare and test several ways for predicting missing information, imputation procedures are easier to handle in the production of statistical data. This is an important advantage when figures should be produced as soon as possible. The U.S. Leading Economic Index (LEI) for instance, which has been published since 1968 , suffered from "ragged edges" with some of the constituting time series. McGuckin, Ozyildirim and Zarnowitz (2007) proposed an imputation procedure that makes better use of the real time financial data. As a result, the official "Conference Board" U.S. Leading Index can be published two weeks earlier than before. The Conference Board implemented this procedure in its 2001 benchmark revision of the LEI. See also the Conference Board handbook on Business Cycle Indicators (2001).

5. Incomplete data and indices

5.1 Introduction

In the ideal case we get a 100% response throughout the whole observation period, not during just a part of the observation period. Also overdue response from previous periods is to be ignored in reports about later periods. Practice, however, is different. Response rates may be low at the earliest publication occasions. Moreover, some institutions collect data from the middle of one month to the middle of the next month and label the results as representing the whole latter month. Other institutions erroneously count overdue response as representing the month when it was received, whereas it should be counted with an earlier month, the month that the respondent reported about. We pay for these inconveniences and errors with several potential biases. Developments of not responding units are not represented in the published figures and figures get published with a more recent tag than is justified by the contents. In the present section we will elaborate on these problems.

In the third subsection we will look into the problem of following changes in the population of enterprises when an index like turnover is to be computed.

Fourth, we note that the economy is changing rapidly. Frequently new products like smart phones appear on the market while others like tube televisions disappear. In order to make 'the best' estimates for price and volume indices, weighting schemes should be kept in line with the dynamics in the economy. To reach this in practice, often the previous year is used as base year in index formulae. A consequence of applying such a 'moving base year' is that a consistent time series of indices cannot be made. A solution for this is the use of chain linked indices.

5.2 An observation period that does not match with the reporting period

For some time now, an important issue in European statistics has been the improvement of the timeliness of statistics. This research is driven by the demands of the ECB, Eurostat, financial institutions and economic analysts for the faster publication of data. One method for achieving this goal is to take measurements only during a part of the period on which one will report. For example, only during the first two weeks of a month. The data for the observed part are assumed to represent the whole period. Thus earlier publication is achieved. The problem which then arises is how the accuracy of a statistic measured by this method is affected. All developments in the ignored part are missed. Surprisingly, very little research into this problem has been performed. This is all the more relevant as a large number of statistics is routinely produced in this manner. Lack of data probably has been a major cause of this deficit in research. We will summarise here results of a study into the influence of timing and length of the measurement period on monthly price indices (Van Ruth, 2002). Specifically, it has been studied whether there are weeks in the month which yield more accurate monthly indices than other weeks. Further, the influence on the accuracy of lengthening the price collection period is examined as well.

New data collection techniques have resulted in data becoming available at higher frequencies, for example prices. These are used to construct inflation measures such as consumer price indices (CPI). Virtually every country in Europe publishes a monthly consumer price index. The measurement period varies widely, from one day in the middle of the month to almost the whole of the month . In the Netherlands some price data on a daily or weekly basis have become available. These have been used to investigate the influence of using different measurement periods on the price index series.

Three types of data are available: petrol prices, traditionally collected prices for items that exhibit frequent price variations, such as vegetables, and so-called scanner data from super-market chains. From each of these series an index based on the whole month and on different parts of the month has been constructed. These then were compared. It will be attempted to generalise the results to other statistics.

It is difficult to make a general judgement on the use of shorter measurement periods for the production of price indices. Indices based on measurements during just one week of a month result in large monthly deviations from the index as based on data for the whole of the month. Average errors are generally as large as or larger than average monthly changes in the reference series. On top of this, maximum single errors can be very large, up to 35%. Errors in the monthly year-on-year price change rate, usually the statistic of interest, are large as well. The average error for single week indices is about half the average monthly rate of the whole month series. Therefore, in that case, inflation rates will possess a large amount of uncertainty.

On the other hand, for conventionally measured price indices, the price development in time of the different indices is quite comparable. So if the long-term evolution of inflation is the main interest, measurement in one week can be acceptable. And using the first two weeks of a month will then result in improved accuracy. Also, developments missed in one month will usually show up a little later. This is not so for price indices based on scanner data. The inclusion of special offers, usually present in single weeks, means that important price changes can be either missed or overstated in single week indices. This is reflected as well in the relatively very large error in the single week inflation rates. The average error is about as large as the average monthly inflation rate of the reference series. It is clear that, for the products considered here, short price collection periods cause a large amount of relevant information to be missed. Therefore, if the short-term development of prices is of primary interest, collecting prices in only a (minor) part of the month will probably be unacceptable. This is especially so if stringent accuracy requirements are present.

All these problems are mitigated when longer period averages are used. Two-week averages more or less halve the errors considered, and three-week averages bring these down even more. This last method also diminishes the influence of price variability on the magnitude of the errors. It has been shown that, as was to be expected, both average monthly errors and maximum errors increase with increasing price variability. The effect is strongest on the level of the maximum errors as measured in the individual item's time series. For the classes of items with the lowest price variability, there is actually not much difference in performance between the different measurement periods. Unfortunately, these constitute only a minority of the sample.

The acceleration of price index measurement by considering only part of the reference period clearly carries a cost in accuracy. How serious this cost is, and if the improvements in timeliness are worth it, depends on the users of the statistics.

Economists from the ECB, an important user, have stated that frequent revisions of the HICP of more than 0,1 percentage point would be as harmful as late publication . If for example petrol prices are considered, this already proves to be a difficult constraint. In the Netherlands, petrol has a weight of 2.7% in the CPI and even in the best case an average error of 1,4 in the index is present. The average error in the CPI resulting from this would be about 0,04 percentage point. Thus just one of the items, with a relatively small weight in the overall index, already consumes a large part of the margin of error.

Finally, it is interesting to look at the concept of shorter measurement periods in a broader perspective. The question is, whether given these results for price indices, the method is applicable on other statistics as well. It depends on the type of statistic. For statistics which do not fluctuate very much over shorter periods and for which the general development in time is the most important, including only part of the reference period seems to be warranted. These conclusions are in line with the ILO CPI manual on price indices (2003), chapter 6.9.

5.3 On chain estimators and changes in population or sample

Establishment surveys should represent the population as closely as possible. Therefore the sampling scheme should not only reflect the fact that some enterprises disappear, but should also allow the inflow of new enterprises and enterprises that have moved to the activities that are the subject of the present survey. However, this may result in serious distortions of an index that is derived from that survey. To illustrate this we discuss an economic indicator like turnover, that is represented as a percentage of a base year, which is set to 100. In such a case the index for period *t*, I^t, should be

$$I^t = 100\ \frac{Y^t}{Y^0},$$

with Y^t the aggregate at period *t*, for instance turn over, and Y^0 the same figure in the base period. This index may be thought of as a product of growth rates, $G^{t-1,t} = Y^t / Y^{t-1}$, for all intermediate periods,

$$\begin{aligned} I^t &= 100\ G^{t-1,t} G^{t-2,t-1} \cdots G^{0,1} \\ &= 100\ \frac{Y^t}{Y^{t-1}} \frac{Y^{t-1}}{Y^{t-2}} \cdots \frac{Y^1}{Y^0} \\ &= 100\ \frac{Y^t}{Y^0} \end{aligned} .$$

Rewritten in this way, it is readily seen that this product of growth rates is equivalent to the ratio of the current aggregate to the aggregate at its base period.

Things may go wrong when the population changes. Suppose that a growth rate can only be computed with that subsample of the population for which comparable data are available for both subsequent periods. Then the current estimate for the previous occasion, $Y^{t-1,t}$, is not exactly equal to the estimate $\tilde{Y}^{t-1}$ at the previous occasion. Therefore the desired index is polluted with a bias factor that is not automatically corrected,

$$
\begin{aligned}
I^{t} &= 100 \;\; G^{t-1,t} G^{t-2,t-1} \cdots G^{0,1} \\
&\approx 100 \;\; \frac{Y^{t}}{Y^{t-1,t}} \frac{\tilde{Y}^{t-1}}{Y^{t-2,t-1}} \cdots \frac{\tilde{Y}^{1}}{Y^{0,1}} \\
&= 100 \;\; \frac{Y^{t}}{Y^{0}} \left[\frac{\tilde{Y}^{t-1}}{Y^{t-1,t}} \frac{\tilde{Y}^{t-2}}{Y^{t-2,t-1}} \cdots \frac{\tilde{Y}^{0}}{Y^{0,1}} \right] \\
&= index\ (t) * bias \quad factor\ \ (t).
\end{aligned}
$$

The same problem arises in any situation where an estimate of Y^{t} takes another value when compared with Y^{t-1} than when compared with Y^{t+1}. This can be the case if an error is edited in between the publication of $G^{t-1,t}$ and $G^{t,t+1}$. Also robust estimation methods can introduce a small bias, usually an underestimate of the growth rate, that will build up to become a large bias as the chain of growth rates gets longer. Van Delden (2006) studied and quantified the various components that contributed to bias in the index of a specific NACE code, the supermarkets.

These bias inducing problems with the present chain index become visible as soon as a new basis for the index is chosen and the current Y^{t} appears to be inconsistent with $Y^{0} I^{t}$. They can also become apparent before that time, when subseries are inconsistent with the generic index for both subseries. Suppose that next to the economic growth rate two subseries are published, namely the domestic and foreign growth rate. Suppose we have the following gross turnover for periods 0, 1 and 2.

Table 5.1. Turnover and index values

Period	domestic	abroad	total		domestic	abroad	total
0	80	20	100		100	100	100
1	80	5	85		100	25	85
2	60	5	65		75	25	65

Then the index values are as the right hand side of the Table shows. However, activity codes for several enterprises have changed. Therefore, the following corrections were carried through. Please note that the corrections for domestic and foreign turnover add up to total turnover, i.e. they are consistent.

Table 5.2. No longer valid turnover (to be ignored in comparison with the next occasion)

Period	domestic	abroad	Total
0 to 1	0	5	5
1 to 2	2	3	5

Table 5.3. Newly added turnover (to be ignored in comparison with the previous occasion)

Period	domestic	abroad	total
0 to 1	0	0	0
1 to 2	10	1	11

Now growth rates are found that differ from those that appear from Table 5.1. The domestic growth rate from period 1 to 2 is no longer 0.75, but (60-10) divided by ((80-2), that is 0,64. (Enterprises contributing to Table 5.2 have left the present activity code and should not be considered in the comparison with the first period where their observations are no longer available in the present NACE section. Enterprises in Table 5.3 just entered the present NACE code and should not contribute to the comparison with the previous period, when they were classified elsewhere.)

Table 5.4. Inconsistent index values after data mutations

Period	domestic	abroad	total
0	100	100	100
1	100	33	89
2	64	67	60

After computing growth rates and deriving index values from that, it appears that the total index value at period 2 is no longer consistent with the index values of the constituting categories: 60 is not in between 64 and 67, though it should be if the figures were to be consistent.

Such inconsistencies cannot always be avoided, but some choices make their occurrence less influential. Firstly both estimates in a growth rate, Y^{t} and Y^{t-1}, should be consistent with respect to the treatment of errors. Secondly, it is advisable to benchmark the series at regular intervals. In the case of the turn-over index one can use quarterly or yearly VAT data for this purpose. (In some countries VAT data become available too late for short term indicators, at least for small enterprises.) If benchmarking is not feasible, one should avoid those (robust) estimation methods that tend to reduce the influence of positive outliers only. Moreover, moving enterprises from one activity code to another should be avoided or postponed to the next base year (at the cost of less validity). If a benchmark is available, activity codes can be adapted more often, though.

The issue of actuality versus consistency does not only pop up at the level of sample selection criteria, but also plays a role in determining weights in price indices, as will be discussed in the next subsection.

5.4 Chain linked versus fixed base price indicators

Changes in prices and volumes are important indicators for judging the performance of an economy. The consumer price index and the real growth of gross domestic product are well known examples. For this purpose national accounts of an increasing number of countries make use of a chain linked method in which the weights for prices and volumes of the components of the economy are updated annually. Generally Laspeyres type indices are used for price and volume estimates[1]. In a Laspeyres index formula individual price and volume changes are weighted with the values of the concerning transaction in a "base year". Generally speaking there is a choice between a fixed base year and a moving base year.
With the method of fixed weights for a series of years, the weights are derived from a single year in the past. An advantage of this method is that in longer series of level estimates of deflated components of aggregates at prices of the base year, these level estimates exactly add up to the deflated aggregate. However, a very serious disadvantage is that volume changes of aggregates are calculated with outdated

[1] In national accounts a Laspeyres volume index is combined with a Paasche price index in order to get a full breakdown of value changes in price and volume changes and warrant additivity in the system.

weights. This disadvantage is especially severe when relative prices change rapidly. As a result, economic growth can be significantly overestimated. Also appearance of new products (mobile telephony, tablets) or the disappearance of others (tube television) can lead to distortion of estimates of economic growth.

Applying a moving base year means that weights are updated every year and are usually derived from the previous year. Since those weights are more up-to-date, a better approximation of the "real world" price and volume changes is obtained than with the method of fixed weights. However due to the changing weights, compilation of time series of indices is not straightforward. Time series results can be obtained by chaining of indices, i.e. multiplying separately estimated year-to-year price and volume indices. (The time series starts in a specific year for which the index is set to 100.) An important advantage of the chain linking method with moving weights is that the above mentioned overestimation of growth rates is avoided. There is also a disadvantage with the time series of indices when a moving base year for prices is applied. In the time series in terms of prices of a specific reference year, the deflated components of an aggregate no longer exactly add up to the deflated aggregate. As a result "mathematical discrepancies" will appear that cannot be removed without distorting the underlying "actual" volume and price movements.

More details, tips and tricks can be found in Brueton (1999) and in leading manuals from the International Labor Organization (2003), Eurostat (2001, 2009, 2015) and the International Monetary Fund (2010).

6. Seasonal adjustment in times of strong economic changes

6.1 The problem and the options

Starting in 2008 the world has been subject to a severe economic crisis. A collapsing confidence in the solidity of sub prime mortgages hit the financial sector, and later spread to most parts of the real economy. In most countries, the strongest fluctuations took place in the period 2008-2010.

Due to these fluctuations, the economic crisis had its implications for the seasonal adjustment process. On the one hand, seasonally adjusted figures help in interpreting short term economic developments. On the other hand, the seasonal adjustment process itself is subject to this increased volatility in the data. Producing accurate seasonally adjusted figures is more difficult under these circumstances.

Specifically, there are two problems making seasonal adjustment more difficult:

- Due to the strong fluctuations and rapid changes estimating the seasonal pattern is more difficult, especially at the end of a series. On the same time, these are often the most important figures of a series. The main question is whether these movements at the end of a series are temporary or whether they imply more structural economic changes.
- The seasonal pattern itself may change due to the crisis as well. Again, this can be a temporary change or a more permanent one. At the time seasonal adjustment is carried out, it would be good to know whether the change in the seasonal pattern is a temporary or more permanent change, or whether we are dealing with just heavy fluctuations in the data. The problem is that we can not know at that point.

Now that the crisis has been going on for quite a while, it is important to evaluate the approaches applied to deal with the above problems. In the following subsection we describe seasonal adjustment practice during the crisis, and summarize lessons learned and give recommendations for future steps.

In the following we first give an overview of the series used for this research. Then we describe how the crisis first became visible in the series and could be recognized. Next the influence of the crisis on the seasonal adjustment process is described. Finally we compare approaches of various countries, draw tentative conclusions and give recommendations.

Several software packages are available for seasonal adjustment. In the context of state space models and time series we mention Stamp, SsfPack and Ox (Koopman et al., 2009; Koopman et al. 2008). For seasonal adjustment often a RegArima approach is applied first, to be followed by a filtering method for seasonal adjustment (Findley et al., 1998). This can be accomplished with X13Arima-Seats (Bureau of the Census, 2013) or Tramo-Seats (Caporello and Maravall, 2004). Both approaches are implemented in Demetra+ (Grudkowska, 2013). For a general treatment of seasonal adjustment using X-12-Arima, we refer to Van Velzen et al. (2011). Eurostat (2009; 2015 in preparation) gives general guidelines for seasonal adjustment.

6.2 The case of the economic crisis of 2008-2010 as reported in the Netherlands

Overview of series

We investigated a wide range of series that are seasonally adjusted at Statistics Netherlands. The following gives an overview of the series we looked into:

- National accounts (quarterly, e.g., GDP, imports, exports)
- Macro economy (monthly, e.g., industrial production index)
- Confidence indicators (monthly, e.g., Consumer/producer confidence index)
- Labour market (monthly, e.g., vacancies, unemployment benefits)

Presence and identification of the crisis

The first indication of what later turned out to be a crisis came with the consumer confidence index of September 2007. A sharp decline was witnessed at this moment, which can be seen in Figure 6.1. After that, the first signs of a crisis in the real economy were not seen until the third quarter of 2008. (See GDP series in Figure 6.2.)

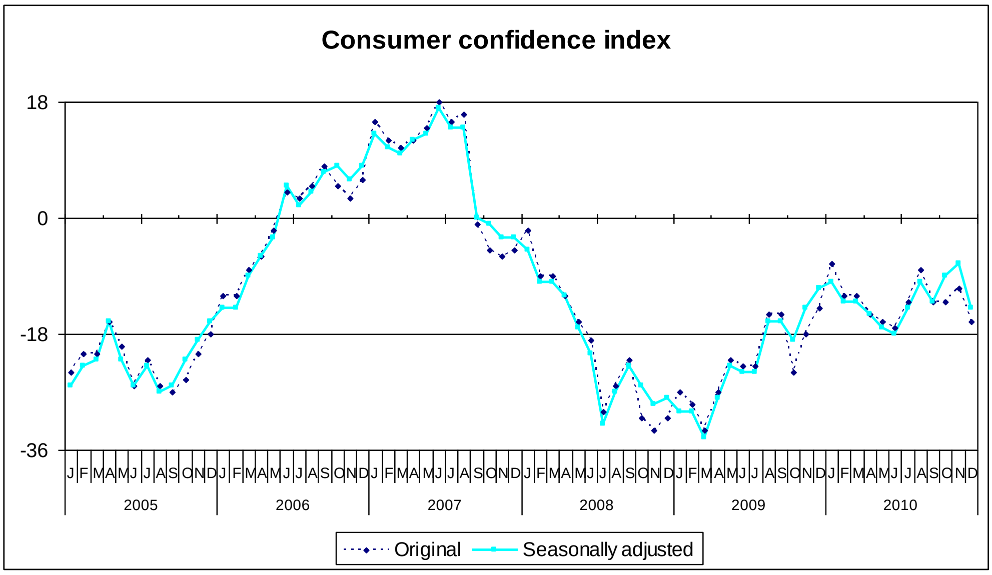

Figure 6.1: Consumer confidence index

For quarterly national accounts series, the start of the crisis was relatively clear. In the second quarter of 2008, several important indicators (GDP, imports, exports) showed a decline in seasonally adjusted figures. In the third quarter, in all other quarterly series a decline could be seen. Since the quarterly series are published only 45 days after the quarter has finished, this decline did not become apparent until October 2008.

This strong decline is illustrated in Figure 6.2 and Figure 6.3. The decline in the second quarter of 2008 can be clearly seen in Figure 6.2. While the previous quarters exhibited high quarter-on-quarter growth rates, the second quarter suddenly drops. At the same time the year-on-year growth rates (Figure 6.3) for the same quarter are still quite high (3,3%). However, the year-on-year growth is approximately equal to the sum of the four most recent quarter-on-quarter growth rates. Since the previous three quarters were strongly positive, the year-on-year growth rate only decreases only partly when adding the second quarter of 2008. The decrease is thus a turning point after a period of growth.

In 2005, the year-on-year growth rate shows a strong decline as well. Unlike in 2008, here the decline is only brief. In order to judge whether the economy was in a recession at that moment, we should look at the quarter-on-quarter growth rate. According to the definition, a recession requires two consecutive quarters with a negative quarter-on-quarter GDP growth rate. In 2005, however, there was no negative growth.

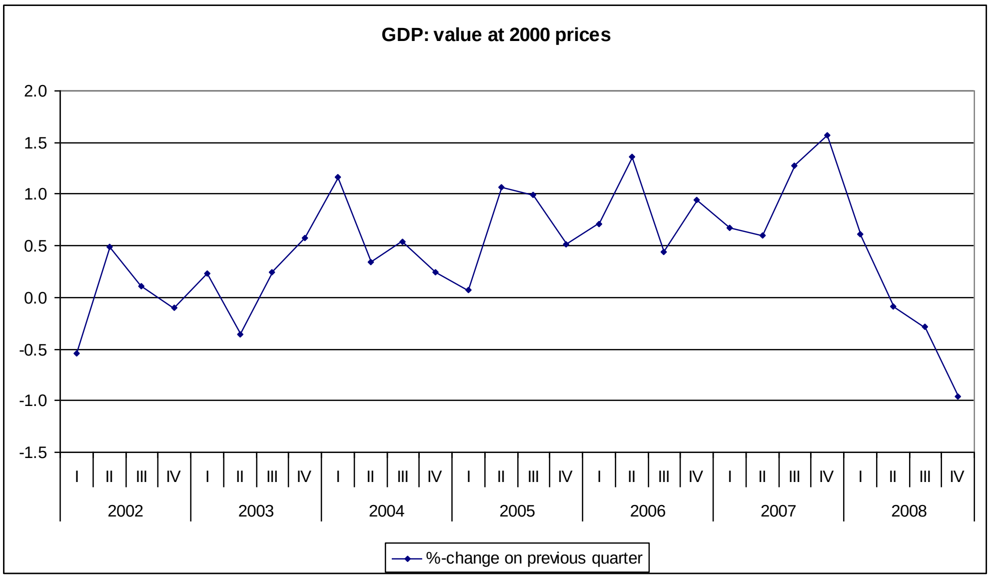

Figure 6.2: Quarter-on-quarter percent change in GDP

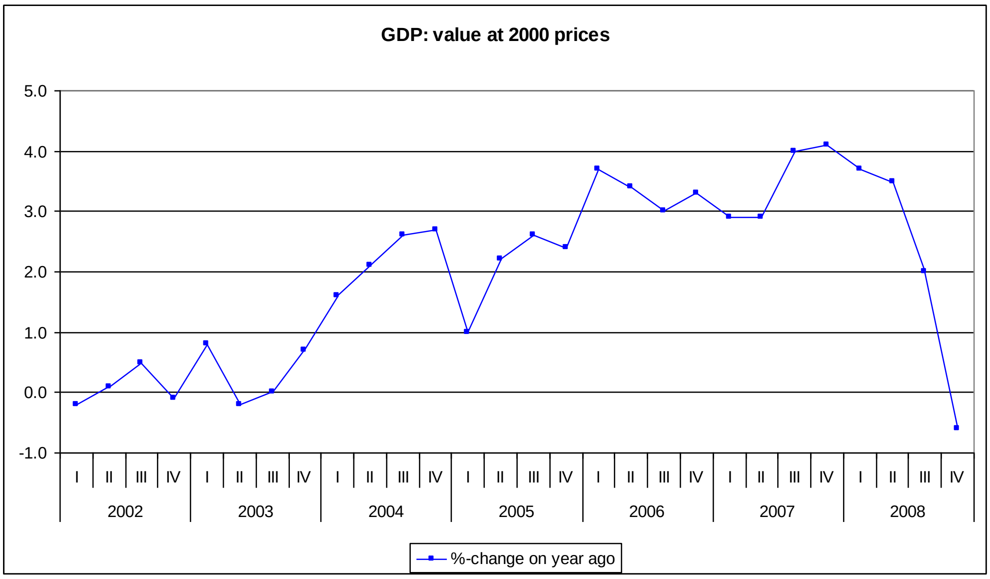

Figure 6.3: Year-on-year percent change in GDP

The judgement that the strong declines should be interpreted as a crisis was based on consensus of experts within and outside Statistics Netherlands (economists, academics). The above definition of a recession was applied. According to this definition, there was a recession in the third quarter of 2008 (See Figure 6.2).

For the monthly series, there was a clear influence of the crisis on the further development of these series. The regular seasonal pattern was clearly less visible than before the start of the crisis. The severity of the crisis caused the series to decline rapidly. Because of this, no seasonal peaks could be seen, but instead, deeper troughs. It is not clear whether actually the seasonal pattern was not present or whether the seasonal movements were relatively small compared to the other movements in the series. This problem was especially present in macro-economic series and for confidence indicators (consumer and producer confidence indices), and to a lesser extent in labour market series. See Figure 6.4 and Figure 6.5 for examples of this 'masked' seasonal pattern in series of unemployment benefits and vacancies.

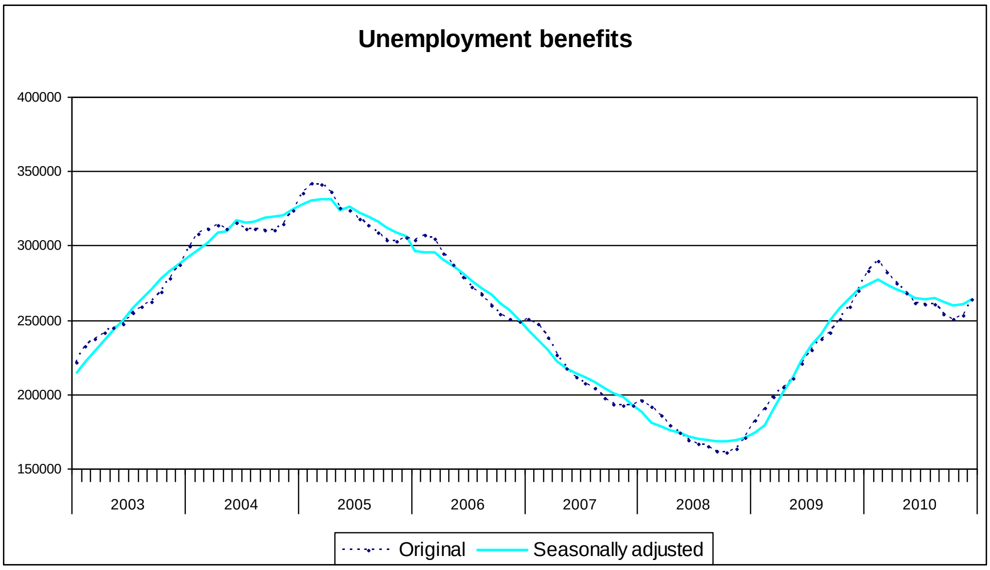

Figure 6.4: Employment benefits, example of a series where the seasonal pattern is 'masked'

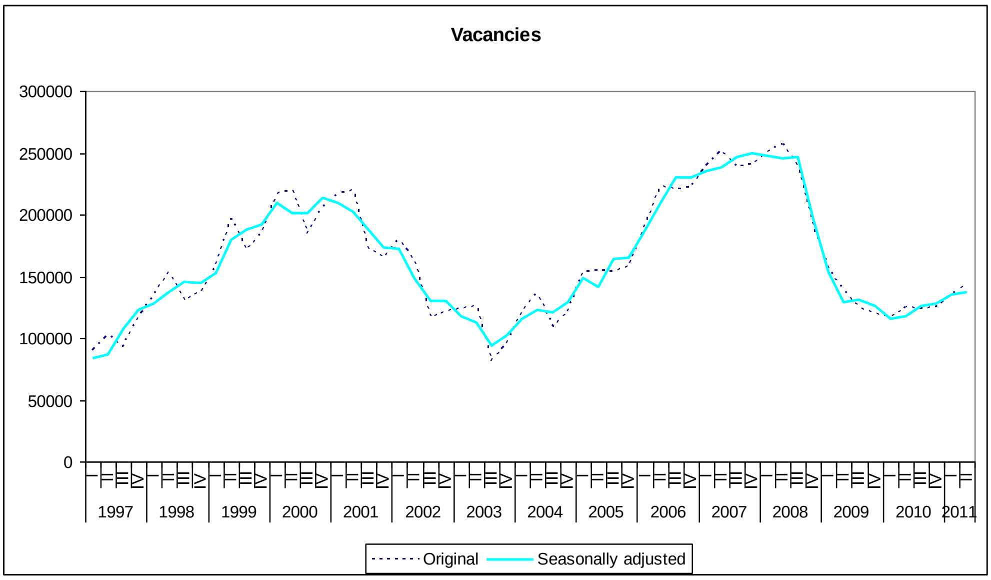

Figure 6.5: Vacancies, example of a series where the seasonal pattern is 'masked'

For both series, an almost linear increase in the original series can be seen in 2009, while the seasonally adjusted series hardly differs (as opposed to the previous years, where strong seasonal patterns can be seen). Later, the seasonal pattern returns, but still not as strong as before. For quarterly series, a similar effect of the crisis on the seasonal pattern could not be seen. The seasonal pattern was clearly present at all times.

For both monthly and quarterly series a clear recession could be seen, with strong movements. These movements mostly did not seem to be structural, when series returned to levels (close to those) before the crisis. However, it remains to be seen if the effects of the crisis will completely disappear.

Influence on seasonal adjustment

For monthly series, the crisis actually made seasonal adjustment more difficult. This became clear from the substantial differences between the consecutive (monthly) estimates of the seasonal patterns. The main difficulties were:

1) Determining whether there actually was a crisis present in the series
2) Deciding whether to intervene in the seasonal adjustment process
3) The number of series that had to be published, and that could have been affected by the crisis
4) Possible consequences of an intervention on other steps of the production process
5) Other adaptations of the production process. For example, the production index was subject to a revision of the NACE classification at the same time of the crisis, so that the effects of these two events could not be separated. In this specific case, it was possible to recalculate the historical series based on this new classification, so that long time series for the new classification were available. Consequently, this made it possible to generate good estimates despite the crisis. If recalculating historical series would not be possible, a new classification could lead to difficulties.

For quarterly series, there were not many disturbances, as mentioned above, the seasonal was clearly visible (at least at a higher level of aggregation).

For the monthly series, most series are revised according to a concurrent adjustment policy, i.e. the data are kept fixed during the current year, and only updated once per year ('concurrent with annual review', see Eurostat guidelines, 2009). For these series, the approach used before the crisis was continued during the crisis, i.e. a fine-tuned setup for each individual series, where outliers were detected automatically (i.e. if the data exceed a critical value). Only for a small number of important series, a revision of the complete series was done every period.

Only part of the outliers was detected in an automatic fashion. Many monthly economic series fluctuated strongly and sometimes showed large peaks, leading to a case-by-case decision of what was an acceptable critical value for outlier detection. In these cases the standard critical value was lowered in steps of, for example, 0.1 (e.g., from 3.3 to 3.2), such that not too many nor too few outliers were detected for a reliable extraction of the seasonal pattern, or for acceptable adjustments at revisions, leading to plausible figures. This process of setting the critical value iteratively was done by seasonal adjustment experts, based on their experience.

The general strategy for the monthly series was to first wait and see how the series developed. After all, at first it is difficult to judge whether movements are temporary or more structural. A quick reaction could lead to a large adjustment later on. The development of the series and effects of seasonal adjustment procedures were monitored constantly, with a monthly decision on whether to intervene or not. These decisions were based on a combination of common sense (plausibility of results) and several quality indicators from X-12-ARIMA. Also, a possible change in parameters such as regression coefficients was closely monitored.

To prevent the need for strong adjustments, in these situations a relatively slow seasonal filter was preferred (slower than the standard X-12 filters). But even these sometimes proved not slow enough: the estimated seasonal patterns fluctuated more despite the slow filter. In Figure 6.6, monthly adjustments for the Industrial Production Index are shown. These adjustments were calculated by comparing the most recent series (June 2011 in this case) to first adjusted figures for each period. It can be seen that the adjustments fluctuate strongly in 2008 and 2009, and are very high several times. Table 6.1 shows the absolute average adjustment per year. Here as well, we can see that 2008 and 2009 contained substantially higher adjustments than surrounding years. In order to prevent these strong adjustments, it is recommended to build in more stability into the seasonal adjustment process in times of strong fluctuations.

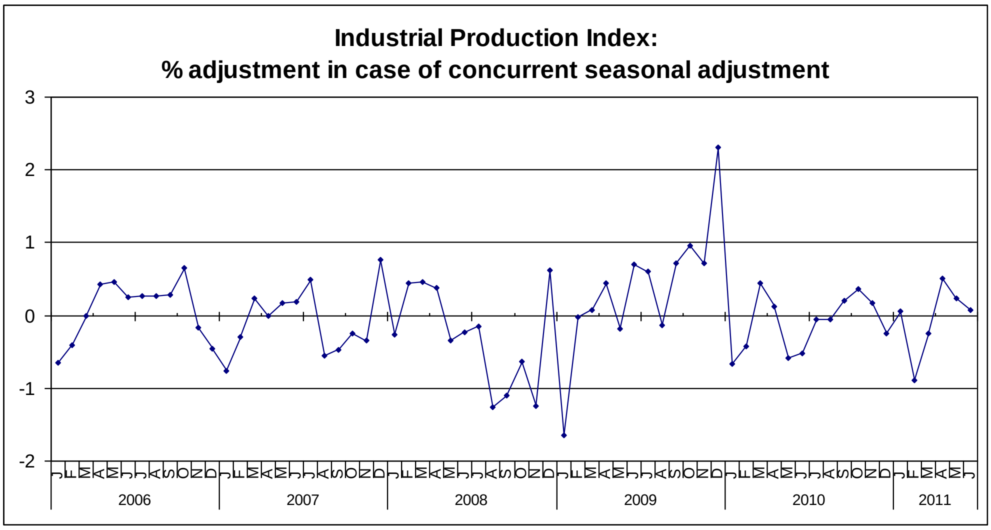

Figure 6.6: Adjustments in case of concurrent seasonal adjustment

Table 6.1: Average absolute adjustment

Year	*Average absolute adjustment*
2006	0.36
2007	0.38
2008	0.60
2009	0.71
2010	0.32
2011	0.33

For several of these important series an intervention was considered necessary. In these cases, several outliers were set manually, after first waiting for several months. In Figure 6.7, an example of a series is shown where two manually set outliers have a strong influence on the Industrial Production Index. It was also considered whether the X-12-ARIMA parameters should be adjusted, which was done in several cases (for the regular annual review of setups).

For many quarterly series no specific adaptations to the seasonal adjustment process were done. Here, the usual approach was continued, including an automatic outlier detection (using default X-12-ARIMA settings). Since quarterly series are generally more stable than monthly series, automatic outlier detection did not give significant problems. For the highest aggregates, the crisis did not lead to more outliers than in other years, while for lower aggregates there were more outliers than usual. Although no significant adaptations were necessary, the series were more closely monitored during the crisis.

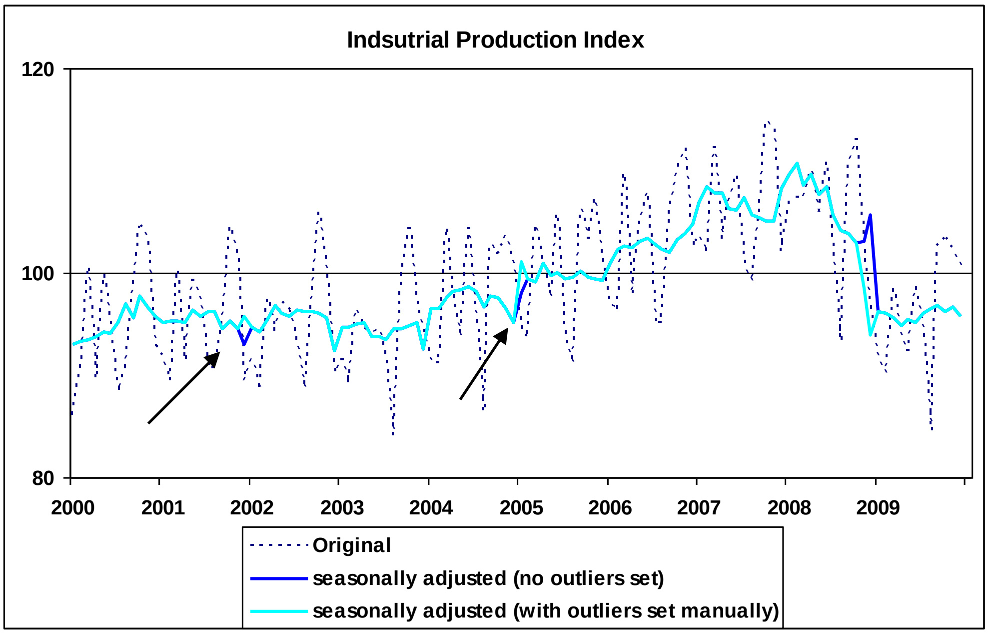

Figure 6.7: Example of a series with manually set outliers

An important question that could not be answered yet is whether the seasonal patterns have in fact changed due to the crisis. They may have changed at least temporarily, but whether the changes are permanent is not clear. This may off course vary from series to series. For example, the seasonal pattern of the monthly series of turnover in hotels and restaurants has likely not changed significantly. Also for higher aggregates of quarterly series, no significant changes are expected, while for lower aggregates this is more likely. Knowing whether there was a change in pattern, is very useful. However, establishing this can only be done when the crisis is over.

6.2.1 Comparison between countries, conclusions and recommendations

Based on information from other European countries, a comparison could be made with their experiences, including the adaptations made to the seasonal adjustment process. The present approach (wait first, then set outliers carefully, using slow filters, or forecasting seasonal factors and then keeping these fixed for a while) is in accordance with international practice (although some countries adopted a more adaptive approach than others). Some countries chose to set more outliers than others, and in general, choices in setting outliers seem to be important. In order to quickly react to changing patterns, sometimes shorter series were used. In case of structural changes this may be a good approach.

The most important conclusion is that seasonal adjustment in times of an economic crisis requires specific and constant attention. Since developments in time series can differ substantially from usual developments, the standard seasonal adjustment procedures may not be applicable. It may be difficult or even impossible to quickly judge whether the developments seen over a short period of time are structural or only temporary changes. However, this aspect is crucial in deciding whether and how to adapt the seasonal adjustment process. Therefore, in these times of strong changes, we recommend a careful approach. Only if a sufficient number of observations was available to make a good judgement, a decision was taken to adjust the process or not. Most often a decision was made to manually set outliers.

Furthermore, it can be concluded that the crisis that started in 2008 has had strong effects on the seasonal adjustment of several series. For some important series (e.g., GDP) there is only a normal decline instead of a deep trough. This could be an explanation why the standard seasonal adjustment process led to acceptable results even in times of a crisis.

The seasonal adjustment process however, required more attention during this crisis than in other years, where it was not always clear whether and at what moment to intervene in the process. Besides, it was only possible to monitor the most important series, where common sense and the use of several indicators were used to decide to intervene or not. A single indicator that shows that there is an extraordinary situation, and that the seasonal adjustment process requires special attention, can have an added value. With such an indicator, more series can be monitored with less effort. Therefore, we recommend developing such an indicator. This indicator may use information from several series at the same time, since simultaneous extraordinary developments are an indication that it is likely that a structural change is taking place. Next to this indicator, recommendations should be developed on how to act when the indicator recommends an intervention.

General recommendations on all aspects of seasonal adjustment can be found in Eurostat (2009; 2015 in preparation). There is a section on the treatment of outliers at the end of the series and at the beginning of a major economic change.

In publications, it can be desirable to keep outcomes of the seasonal adjustment process stable. More research is needed on how this can be done without ignoring real developments and reducing the size of adjustments. Eurostat (2009; 2015 in preparation) gives recommendations on the revision strategy for seasonally adjusted data and the strategy adopted for the seasonal adjustment of long time series, since both can affect the stability and the path of the seasonally adjusted data.

At present it is not yet known whether and to what extent seasonal patterns have changed due to the crisis. For some series this may be the case, while for others it seems the pattern has not changed. For the cases where there was a change in pattern, the change is sometimes only temporary, and appeared during the most severe periods of the crisis. After that the old seasonal pattern then seems to return, as is the case for series where the seasonal pattern was 'masked.' We recommend to do more research on this when the crisis is over, in particular for all important series. Based on that, a decision can be made in what way the seasonal adjustment process should be adapted to deal with a future economic crisis.

References

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

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*, *Series B*, 35, 61-66

Bondt, G.J. de, H.C. Dieden, S. Muzikarova and I. Vincze (2013). Introducing the ECB Indicator on Euro Area Industrial New Orders. European Central Bank, Occasional Paper Series #149, http://www.ecb.europa.eu

Box, G.E.P., and G.C. Tiao (1975). Intervention analysis with applications to economic and environmental problems. *Journal of the American Statistical Association*, 70, p. 70-79

Brueton, A. (1999) The development of Chain-Linked and Harmonised Estimates of GDP at Constant Prices. Economic Trends 552, pp. 39-45, http://www.ons.gov.uk/ons/rel/elmr/economic-trends--discontinued-/no--552--november-1999/the-development-of-chain-linked-and-harmonised-estimates-of-gdp-at-constant-prices.pdf

Bureau of the Census (2013). X-13ARIMA-SEATS Reference Manual. Time Series Research Staff, Statistical Research Division, U.S. Census Bureau, Washington http://www.census.gov/srd/www/x13as/

Caporello, G. and A. Maravall (2004). Program TSW, Tramo Seats Windows, revised reference manual. Banco de Espana, Madrid

Chow, G.C. and A.-L. Lin (1971). Best linear unbiased interpolation, distribution and extrapolation of time series by related series. *Review of Economics and Statistics* 53, pp. 372-375

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

Conference Board (2001). Business Cycle Indicators Handbook. New York, NY: The Conference Board, http://www.conference-board.org/publications/publicationdetail.cfm?publicationid=852

Daas, P.J.H., and M.J.H. Puts (2014). Big Data as a Source of Statistical Information. Paper written for the ECB workshop on using Big Data for forecasting and statistics, Frankfurt am Main, 7-8 April 2014.

Denton, F.T. (1971) Adjustment of monthly or quarterly series to annual totals: An approach based on quadratic minimization. *Journal of the American Statistical Association 66*, pp. 99-102.

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

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

Durbin, J. and Quenneville, B. (1997). Benchmarking by state-space models. *International Statistical Review*, *65*, pp. 23-48

Eurostat (2001). Handbook on price and volume measures in National Accounts, http://ec.europa.eu/eurostat/en/web/product-manuals-and-guidelines/-/KS-41-01-543

Eurostat (2009). Overview of international system of classifications, http://ec.europa.eu/eurostat/statistics-explained/index.php/File:International_system_of_classifications.PNG, http://ec.europa.eu/eurostat/en/web/products-manuals-and guidelines/-/CA-26-99-069

Eurostat (2009; 2015 in preparation). ESS guidelines on seasonal adjustment, http://ec.europa.eu/eurostat/documents/3859598/5910549/KS-RA-09-006-EN.PDF ,

Eurostat (2013). Weighting and Estimation- 11-M-Preliminary Estimation Model-Based v3.2p. In: Memobust, www.cros-portal.eu/content/small-area-estimation

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

Findley, D.F., B.C. Monsell, W.R. Bell, M.C. Otto and B.-C. Chen (1998) New Capabilities and Methods of the X-12-ARIMA Seasonal-Adjustment Program, Journal of Business and Economic Statistics 16, pp. 127-152

Foroni, C. and M. Marcellino (2013) A survey of econometric methods for mixed-frequency data. Norges Bank, Working paper 2013/06

Grudkowska, S. (2013). Demetra+ user manual. National Bank of Poland

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

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

International Labor Organization, (2003). Consumer Price Indices: An ILO Manual. International Labor Organization: Geneva

International Monetary Fund (2010). Producer price index manual, http://www.imf.org/external/pubs/ft/ppi/2010/manual/ppi.pdf

Koopman, S.J., A.C. Harvey, J.A. Doornik, and N. Shephard (2009). *Structural Time Series Analyser, Modeller and Predictor: STAMP 8.2:* London: Timberlake Consultants Press

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

Krieg, S., V. Blaess and M. Smeets (2012). Small area estimation of turnover of the Structural Business Survey. Statistics Netherlands discussion paper 2012 03, http://www.cbs.nl/en-GB/menu/methoden/onderzoek-methoden/discussionpapers/archief/2012

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

McGuckin, R.H., A. Ozyildirim, V. Zarnowitz (2007) A more Timely and Useful Index of Leading Indicators. Journal of Business & Economic Statistics 25, pp. 110-120.

Organisation for Economic Co-operation and Development, OECD, (2008) Handbook on constructing composite estimators: methodology and user guide, http://www.oecd.org/std/42495745.pdf

Ostrom, C.W. (1978). Time series analysis. *Sage University paper series* 9

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., M. Feder and D. Signorelli (1998). Estimation of Autocorrelations of Survey Errors with Application to Trend Estimation in Small Areas. *Journal of Business & Economic Statistics*, 16, pp. 339-348

Pfeffermann, D., and 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 M. Yu, (1994). Small area estimation by combining time series and cross-sectional data, *Canadian Journal of Statistics*, 22, pp. 511-528

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

Sartore, M., and D. Caporin (2006). Methodological Aspects of Time Series Back-Calculation. Working paper 56 Ca' Foscari University of Venice, Dipartimento Scienze Economiche

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

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

Schouten, B., Shlomo, N. & Skinner, C. (2011). Indicators for Monitoring and Improving Representativeness of Response. *Journal of Official Statistics* 27, pp. 1-24

Statistics Estonia (2009). Frequently asked questions about the chain-linking method. www.stat.ee/dokumendid/29861

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

United Nations Statistical Division, UNSD (2002, 2010). Manual on Statistics of International Trade in Services
MSITS 2002: http://unstats.un.org/unsd/pubs/gesgrid.asp?id=457 or http://ec.europa.eu/eurostat/documents/3859598/5855477/MANUAL-EN.PDF ,
concept MSITS 2010: http://unstats.un.org/unsd/statcom/doc10/BG-MSITS2010.pdf .

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

Van den Brakel, J.A. (2010). Sampling and estimation techniques for the implementation of new classification systems: the change-over from NACE Rev. 1.1 to NACE Rev. 2 in business surveys. *Survey Research* Methods*.* 4, pp. 103-119

Van den Brakel, J.A. and C.A.M. van Berkel (2002). A design-based analysis procedure for two-treatment experiments embedded in sample surveys. An application in the Dutch Labour Force Survey. *Journal of Official Statistics*, vol. 18, pp. 217-231

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

Van den Brakel, J.A. and S. Krieg (2012). Dealing with small sample sizes, rotation group bias and discontinuities in a rotating panel design. Working paper, Statistics Netherlands, Heerlen

Van den Brakel, J.A., and R.H. Renssen (2005). Analysis of experiments embedded in complex sampling designs, *Survey Methodology*, vol. 31, nr. 1, pp. 23-40

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

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

Van Delden, A. (2006). Diverging growth rates of supermarkets: causes and remedies (in Dutch) Statistics Netherlands research paper TMO-R&D-2006-02-09-ADLN

Van Ruth, F.J., 2002, The influence of the measurement period on consumer price indices, HCPI 02/398, paper presented at the Eurostat Working Party "Harmonisation of Consumer Price Indices", Luxemburg, 8-10 June 2002

Van Ruth, F., B. Schouten and R. Wekker (2005). The Statistics Netherlands' Business Cycle Tracer. Methodological aspects; concept, cycle computation and indicator selection , http://www.cbs.nl/NR/rdonlyres/253FD272-B93E-46FF-A474-1E5A396C81F1/0/2005methodebusinesscycletracerart.pdf

Van Velzen, M., R. Wekker en P. Ouwehand (2011). *Seasonal adjustment*, Statistics Netherlands Methods Series, The Hague, http://www.cbs.nl/NR/rdonlyres/75B108E9-FE3E-4506-904D-166AD9F785A6/0/2011x37062art.pdf

Zhang, Li-Chung (2012). Topics of statistical theory for register-based statistics, *Statistica Neerlandica* 66, pp. 41-63

Appendix. Modelling interrupted time series

Structural time series models

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

Consider a repeatedly conducted survey to produce series of official statistics for target variables. The variables of interest are defined as categorical variables measured on a nominal or an ordinal scale, means or totals. The population values of a categorical variable specify the distributions in the population over say *K* categories. For each variable a *K*-dimensional vector $\mathbf{y}_t = (y_{t,1}, \dots y_{t,K})$ is defined where the elements of $\mathbf{y}_t$ specify the proportions over the *K* categories. These *K* variables are subjected to the restriction $\sum_{k=1}^{K} y_{t,k} = 1$ for all *t*.

Population values for means and totals can be decomposed in say *K* subcategories and are defined as $\mathbf{y}_t = (y_{t,+} y_{t,1}, \dots y_{t,K})$. The first component, $y_{t,+}$, stands for the total and is broken down over *K* categories specified by the remaining estimates $\hat{y}_{t,k}$, $k = 1 \dots K$. These *K*+1 variables are subjected to the restriction $y_{t,+} = \sum_{k=1}^{K} y_{t,k}$ for all *t*.

Based on the data observed in a sample, direct estimates for the unknown population values are obtained with design-based or model-assisted estimators, known from sampling theory. Examples are the Horvitz-Thompson estimator or the generalized regression estimator, Särndal et al. (1992). As a result, for each categorical variable *K* series are observed that specify the estimated proportions over *K* categories and are collected in the *K*-dimensional vector $\hat{\mathbf{y}}_t = (\hat{y}_{t,1}, \dots \hat{y}_{t,K})$, *t* = 1, …, *T*. For variables defined as means or totals, *K*+1 series are observed that are collected in the *K*+1-dimensional vector $\hat{\mathbf{y}}_t = (\hat{y}_{t,+}, \hat{y}_{t,1}, \dots \hat{y}_{t,K})$, *t* = 1, …, *T*. The sample estimates obey similar restrictions as their population values, that is $\sum_{k=1}^{K} \hat{y}_{t,k} = 1$ for proportions and $\hat{y}_{t,+} = \sum_{k=1}^{K} \hat{y}_{t,k}$ for means and totals.

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

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

To keep the notation as parsimonious as possible, it is assumed that the autonomous development of the series of the indicator is modelled with a stochastic trend, a regression component and an irregular component. The regression component consists of an intervention variable with a time independent regression coefficient that describes the effect of the survey transition. This approach is initially proposed by Harvey and Durbin (1986). Seasonal, cyclic, ARMA, and other auxiliary regression components can be included in the model depending on the application at hand.

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

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

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

Surveys are often based on a rotating panel design. Such designs result in partially overlapping samples with correlated sampling errors. Particularly in these cases a separate component for the sampling error in the time series model might be required to capture this serial correlation. Through this component the estimated variances for the $\hat{y}_{t,k}$, which are generally available from the survey, can be included in the time series model as prior information. Binder and Dick (1990) proposed the following general form for the sampling error model to allow for non-homogeneous variance in the sampling errors:

$$e_{t,k} = \omega_{t,k} \tilde{e}_{t,k} , \qquad (2)$$

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

If the observed series are based on non-overlapping cross-sectional samples, then there is no serial correlation between sampling errors. As a result the sampling error and the irregular component of the true population parameter cannot be separated. Therefore both terms are combined in one irregular term,

$$\nu_{t,k} + e_{t,k} = \varepsilon_{t,k}, \qquad (3)$$

which is assumed to be normally and independently distributed with zero mean and a variance that is proportional to the variance of $\hat{y}_{t,k}$, i.e. $\varepsilon_{t,k} \cong N(0, \sigma^2_{\varepsilon,k} V\hat{a}r(\hat{y}_{t,k}))$, where $V\hat{a}r(\hat{y}_{t,k})$ is the estimated variance of $\hat{y}_{t,k}$. This variance structure still allows for non-homogeneous sampling variation, caused by differences in the yearly sample size, differences in the sampling design or differences in the variance of measurement error due to redesigns of the data collection procedures. The parameter $\sigma^2_{\varepsilon,k}$ is added, since the variation of $\varepsilon_{t,k}$ will not exactly follow the variance of the direct estimator. It is also assumed that the irregular components of (3) at different time points are uncorrelated: $Cov(\varepsilon_{t,k}\varepsilon_{t',k}) = 0$ for $t \neq t'$. As a result model (1) simplifies to

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

By combining $e_{t,k}$ and $\nu_{t,k}$ into one irregular term, it is implicitly assumed that the sampling error $e_{t,k}$ dominates the irregular term $\varepsilon_{t,k}$. The estimates for $\sigma^2_{\varepsilon,k}$ provide an indication to what extent the sampling error dominates the irregular term. They are expected to take values around one if the sampling error dominates the irregular of the population parameter, since the variance of $\varepsilon_{t,k}$ is taken proportional to the variance of $\hat{y}_{t,k}$. Estimates for $\sigma^2_{\varepsilon,k}$ larger than one indicate that $\nu_{t,k}$ contributes substantially to the variation of $\varepsilon_{t,k}$.

In many cases, there are no direct estimates available for the variance of $\hat{y}_{t,k}$. In such cases the variance can be approximated, for example by taking the variance inversely proportional to the sample size: $\varepsilon_{t,k} \cong N(0, \sigma^2_{\varepsilon,k} / n_t)$. This variance approximation implies that the only difference between the sample designs between the subsequent editions of the survey, is the sample size. In the case of categorical variables, the $\hat{y}_{t,k}$ are proportions, whose variances can be approximated with $V\hat{a}r(\hat{y}_{t,k}) = \hat{y}_{t,k}(1 - \hat{y}_{t,k}) / n_t$.

For the stochastic trend, the so-called smooth trend model is assumed as an example since this trend model is widely applied in econometric time series analysis, see e.g. Durbin and Koopman, (2001). In each particular application, other trend models should be considered. The smooth trend model is defined as:

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

with $L_{t,k}$ the level component and $R_{t,k}$ the stochastic slope component of the trend, and $\eta_{t,R,k}$ an irregular component. The smooth trend model (5) is a special case of the local linear trend model, which also has an irregular term for $L_{t,k}$, see e.g. Durbin and Koopman, (2001) equation (3.2), which can be considered as an alternative. It is assumed that the irregular components of (5) are normally and independently distributed, i.e. $\eta_{t,R,k} \cong N(0, \sigma^2_{R,k})$, and that they are uncorrelated at different time points, i.e. $Cov(\eta_{t,R,k}\eta_{t',R,k}) = 0$ for $t \neq t'$. Furthermore, it is assumed that the irregular components of (4) and (5) are uncorrelated: $Cov(\varepsilon_{t,k}\eta_{t',R,k}) = 0$ for all t and t'.

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

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

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

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

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

The interventions described so far assume that the redesign only affects the point estimates of the survey. A survey redesign could, however, also affect the variance of the measurement errors. An increase or decrease of the variance of the measurement errors will be reflected in the estimated variance of $\hat{y}_{t,k}$. A straightforward way to account for such effects is to incorporate the estimated variances of the survey estimates as prior information using sampling error model (2). Another possibility is to define separate model variances for the irregular term $\hat{\varepsilon}_{t,k}$ in the measurement equation for the period before and after the implementation of the survey redesign, i.e. $Var(\varepsilon_{t,k}) = \sigma^2_{s,k.1}$ if $t < T_R$ and $Var(\varepsilon_{t,k}) = \sigma^2_{s,k.2}$ if $t > T_R$. The ratio between $\sigma^2_{s,k.1}$and $\sigma^2_{s,k.2}$ can be used to test hypotheses about the equivalence of both variance components. This approach, however, requires a sufficient number of observations under both surveys to test the equivalence of these variance components with sufficient power.

The discontinuity in the series is modelled with an intervention variable that describes the moment that the survey process is redesigned. This approach assumes that the other components of the time series model approximate the real development of the population variable reasonably well and that there is no structural change in e.g. the trend or the seasonal component at the moment that the new survey is implemented. If a change in the real development of the population variable exactly coincides with the implementation of the new survey, then the model will wrongly assign this effect to the intervention variable which is intended to describe the redesign effect. Information available from series of correlated variables can be used to evaluate the assumption that there is no structural change in the real evolution of the parameter. Such auxiliary series can also be added as a regression component to the model, with the purpose to reduce the risk that a structural change in the evolution of the series of the target parameter is wrongly assigned to the intervention variable. An auxiliary series can also be included as a dependent variable in a multivariate model, which accounts for the correlation between the parameters of the trend and seasonal components, Harvey and Chung (2000) and Van den Brakel and Krieg (2012).

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

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

State-space representation

The structural time series models developed in subsection 2.3 for the separate parameters $\hat{y}_{t,k}$ of the vector $\hat{\mathbf{y}}_t$ comprise a *H*-dimensional structural time series model, where *H*=*K* in the case of a *K*-dimensional categorical variable and *H*=*K*+1 in the case of means and totals including a subdivision in *K* categories. The general way to proceed is to put this model in state-space representation and analyse the model with the Kalman filter. The state-space representation for this *K*-dimensional structural time series model reads as:

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

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

The measurement equation (8) describes how the observed series depends on a vector of unobserved state variables $\boldsymbol{\alpha}_t$ and a vector with disturbances $\boldsymbol{\varepsilon}_t$. The state vector contains the level and slope components of the trend models and the regression coefficients of the intervention variables. The transition equation (9) describes how these state variables evolve over time. The vector $\boldsymbol{\eta}_t$ contains the disturbances of the assumed first-order Markov processes of the state variables. The matrices in (8) and (9) are given by

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

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

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

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

with $\mathbf{I}_{[H]}$ the $H \times H$ identity matrix. The disturbance vectors are defined as

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

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

It is assumed that

$$E(\boldsymbol{\varepsilon}_t) = \mathbf{0}_{[H]},\ Cov\ (\boldsymbol{\varepsilon}_t) = Diag\ (V\hat{a}r(\hat{y}_{t,1})\sigma^2_{\varepsilon,1},..., V\hat{a}r(\hat{y}_{t,H})\sigma^2_{\varepsilon,H}),$$

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

with $\mathbf{0}_{[p]}$ a column vector of order *p* with each element equal to zero. In the case that each measurement equation and each transition equation has its own separate hyperparameter, then (10) is a set of *H* univariate structural time series models. If the measurement equations or the transition equations share common hyperparameters, then (10) is a *H* dimensional seemingly unrelated multivariate structural time series model. This is for example the case if $\sigma^2_{\varepsilon,1} = ... = \sigma^2_{\varepsilon,H} = \sigma^2_{\varepsilon}$. The current state-space representation assumes a diagonal covariance structure for the measurement equation. A further extension is to allow for non-diagonal covariance structures, estimated from the survey data. For example, in the case of categorical data, the design covariances between the categories can be approximated with $Cov\ (\hat{y}_{t,k}\hat{y}_{t,k'}) = -\hat{y}_{t,k}\hat{y}_{t,k'} / n_t$. See Särndal et al. (1992), chapter 5, for more details about design covariances between survey variables.

The time independent regression coefficients of the intervention variables are also included in the state vector, as described by Durbin and Koopman (2001), subsection 6.2.2. The Kalman filter can be applied straightforwardly to obtain estimates for the regression coefficients. An alternative approach of estimating the regression coefficients is by augmentation of the Kalman filter, see Durbin and Koopman (2001), subsection 6.2.3 for details.

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

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

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

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

with $\mathbf{1}_{[p]}$ a column vector of order *p* with each element equal to one. Due to $\mathbf{T}_{iv}$, defined in (10f), the regression coefficients as well as their Kalman-filter estimates obey the restriction $\sum_{k=1}^{K} \beta_k = 0$.

In the case of a population mean or total, the first element of $\hat{\mathbf{y}}_t$ specifies the estimate for a variable and the remaining *K* elements a subdivision over *K* categories. These *K*+1 variables are subjected to the restriction $\hat{y}_{t,+} = \sum_{k=1}^{K} \hat{y}_{t,k}$ for all *t*. As a result, the *K*+1 regression coefficients of the intervention variables must obey the restriction $\beta_{+} = \sum_{k=1}^{K} \beta_k$. This restriction is obeyed if design matrix (10e) is used in the transition equation (9) where

$$\mathbf{T}_{iv} = \begin{pmatrix} 0 & \mathbf{1}^{T}_{K} \\ \mathbf{0}_{K} & \mathbf{I}_{K} \end{pmatrix}. \qquad (10g)$$

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

Kalman filter

After having expressed the multivariate structural time series model in state-space representation and under the assumption of normally distributed error terms, the Kalman filter can be applied to obtain optimal estimates for the state variables as well as the measurement equation see e.g. Durbin and Koopman, (2001). Estimates for state variables for period *t* based on the information available up to and including period *t* are referred to as the filtered estimates. The filtered estimates of past state vectors can be updated if new data become available. This procedure is referred to as smoothing and results in smoothed estimates that are based on the completely observed time series. So the smoothed estimate for the state vector for period *t* also accounts for the information made available after time period *t*. A widely applied procedure to smooth filtered point estimates and standard errors for the state variables is the fixed interval smoother. See Harvey (1989) or Durbin and Koopman (2001) for technical details.

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

The use of a diffuse prior for the regression coefficients of the intervention variables is appropriate if no additional information about the size of discontinuities is available. As explained in subsection 2.2, external information about the size of the discontinuities can be available through the conduction of a parallel run or through recalculation of existing data. This information can be used in the time series models by using an informative prior for the initialization of the regression coefficients. This can be done by using the direct estimate for the discontinuity obtained from the parallel run or through recalculation in the initial state vector for the regression coefficients and the estimated variance of this direct estimate as an uncertainty measure in the covariance matrix of the initial state vector. Another possibility to use the direct estimate of the discontinuities as prior information in the time series model, is to assume that the regression coefficient for the intervention are equal to the observed discontinuity in the parallel run. In this case the direct estimate for the discontinuity is treated as if it is a fixed value, known in advance. The uncertainty of using a survey estimate for the discontinuity is ignored.

Software

Several general statistical software packages contain procedures for the analysis of state-space models, e.g. SAS, R, and Eviews. There are, however, also two software packages that are specialized on the analysis of state-space models. A menu driven and user friendly package is STAMP where most common structural time series models are implemented (Koopman et al., 2009). This package also allows the inclusion of interventions in structural time series models. The additional restrictions on the regression coefficients of the intervention variables in multivariate models, developed in subsection 2.4, are not supported. This kind of models can be implemented and estimated with Ssfpack 3.0 in combination with OxMetrics, see Doornik (2009) and Koopman et al. (2008). Ssfpack is a library of subroutines that can be used in a very flexible way to implement virtually all possible state-space models. It is less user friendly than STAMP and intended for the more experienced user of state-space models. Strong advantages of Ssfpack are that most of the advanced methods documented in Durbin and Koopman (2001) are implemented and that it handles the large sparse matrices, which are characteristic for multivariate state-space models, in an efficient way. This avoids many numerical problems, generally encountered by the implementation in programming environments of the aforementioned more general software packages.

For those who prefer a RegArima approach there is X13Arima-Seats (Bureau of the Census, 2013), Tramo-Seats (Caporello and Maravall, 2004) and Demetra+ (Grudkowska, 2013).

Explanation of symbols

Empty cell	Figure not applicable
.	Figure is unknown, insufficiently reliable or confidential
*	Provisional figure
**	Revised provisional figure
2014–2015	2014 to 2015 inclusive
2014/2015	Average for 2014 to 2015 inclusive
2014/’15	Crop year, financial year, school year, etc., beginning in 2014 and ending in 2015
2012/’13–2014/’15	Crop year, financial year, etc., 2012/’13 to 2014/’15 inclusive

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

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

Prepress
Studio BCO, Den Haag

Design
Edenspiekermann

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