The role of outlier detection and replacement in practical seasonal adjustment*

Kirchner, Robert
Deutsche Bundesbank, Statistics Department
Wilhelm-Epstein-Strasse 14
60431 Frankfurt am Main, Germany
E-mail: robert.kirchner@bundesbank.de

Mehrhoff, Jens
Deutsche Bundesbank, Statistics Department
Wilhelm-Epstein-Strasse 14
60431 Frankfurt am Main, Germany
E-mail: jens.mehrhoff@bundesbank.de

ABSTRACT

The particular role of outlier detection and replacement in seasonal adjustment of a time series is to allow an unbiased estimation of typical seasonal and calendar effects. Only if abrupt, strong and atypical changes are explicitly modelled, is it possible to estimate typical calendar influences and seasonal effects, which recur with similar intensity in the same season each year. The seasonal and calendar components are then eliminated from the unadjusted data in order to obtain the seasonally adjusted results.

It is very likely that problems will arise if outliers are not modelled. It may be the case, for example, that initial estimates of seasonally adjusted figures will identify turning points too late and be misleading over a period of months – compared with the final results, which are established years later. Furthermore, one year after mistakenly deciding not to identify outliers, problems may arise at the current end of the time series because parts of the one-off special effects have been erroneously ascribed to the seasonal component as recurring each year. This would result in inaccurate movements of the seasonally adjusted series which have to be corrected at a later date. In order to prevent such problems and the corresponding revisions, it is advisable to treat abrupt, strong and atypical events as outliers in the process of seasonal adjustment.

The reasoning behind this is illustrated using empirical examples such as the circulation of Argentine currency during and after the Argentine crisis in 2001-2002 as well as building permits granted for residential construction in western Germany during and after discussions about phasing out the grant for homebuyers. It is shown that applying the Guidelines on Seasonal Adjustment of the European Statistical System helps to avoid flawed modelling of the seasonal and calendar components and, thus, of the seasonally adjusted results.

1 The role of outlier modelling in seasonal adjustment

The goal of seasonally adjusted data is to reveal the “news” contained in a time series (cf. Eurostat, 2009, item 0). Seasonally adjusted results should show changes in long-term growth trends or in the business cycle as well as the impact of special effects such as strikes, large orders, atypical weather conditions or natural disasters.

In terms of methodology, “news” is filtered out of unadjusted data using time series analysis tools. As a rule, it is assumed that the time series of the unadjusted values Y_t can be decomposed into three unobservable components

* This paper represents the authors’ personal opinions and does not necessarily reflect the views of the Deutsche Bundesbank or its staff.

(1) $Y_t = T_t\, S_t\, I_t$,

where T_t stands for the trend-cycle component, S_t for seasonal fluctuations (and calendar influences) and I_t for irregular effects at time *t*.

To then extract the "news" from time series Y_t, the seasonal effects (and calendar influences) S_t, which recur with similar intensity in the same season each year, are eliminated:

$$(2)\quad \frac{Y_t}{S_t} = T_t I_t \,.$$

For the seasonally adjusted time series $T_t\, I_t$ to also show the full extent of all special effects, care must be taken to ensure that these effects are not erroneously ascribed to the seasonal component S_t and accidentally also eliminated. This is the task of outlier modelling. In other words, the particular role of outlier detection and replacement in seasonal adjustment of a time series is to allow an unbiased estimation of typical seasonal and calendar effects. Only if abrupt, strong and atypical changes are explicitly modelled, is it possible to estimate typical calendar influences as well as seasonal effects and, therefore, seasonally adjusted results that measure the "news".

2 Detecting outliers

As seasonal estimation in the X-12-ARIMA and TRAMO/SEATS procedures comprises a number of steps, outliers have to be taken into account separately at each stage. The treatment of outliers is particularly important in RegARIMA models, which form the basis for the TRAMO/SEATS seasonal adjustment program and which are also used in the X-12-ARIMA procedure (cf. Gómez and Maravall, 1996, and Findley et al., 1998). The aim of these models is to obtain an unbiased estimation of the seasonal component where the detection of outliers

- is intended to achieve a linearisation of the time series that produces Gaussian residuals,
- allows an unbiased estimation of the regular and seasonal differencing orders as well as of the model parameters that result in the selection of seasonal filters in the TRAMO/SEATS procedure and thus in the calculation of the seasonal components and the seasonally adjusted series,
- can be used to calculate optimal one-step-ahead forecasts that have an impact, in particular, on the estimation of the seasonal component towards the end of the series.

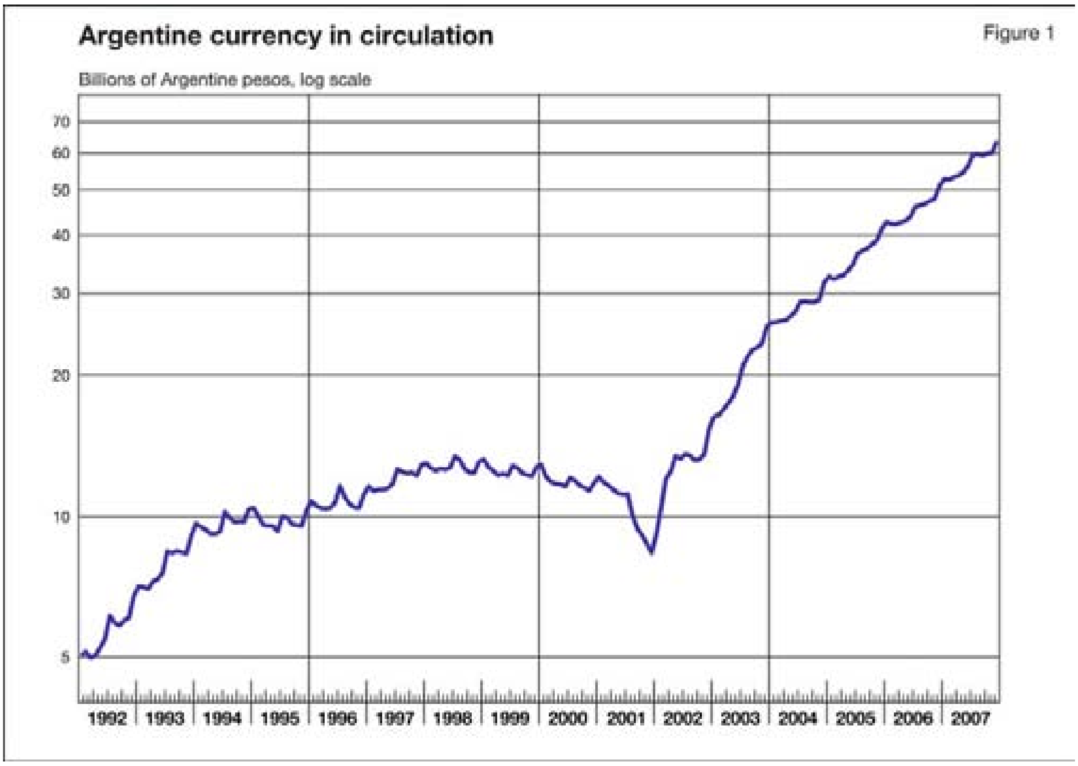
Hence, there are many ways in which outlier modelling may influence the estimation of the seasonal component and thus the seasonally adjusted series. Given the major impact that the treatment of outliers has on the seasonally adjusted results, the presence of outliers is checked statistically using *t*-tests in the X-12-ARIMA and TRAMO/SEATS program packages. Different types of outliers are distinguished (additive outliers, level shifts, temporary changes etc). In the case of purely statistical decisions, there are of course also errors of type I and II. A value is treated as abnormal although it is not (type I error) or is seen as "normal" although it is not (type II error). To avoid such purely statistical errors wherever possible, the ESS Guidelines on Seasonal Adjustment specify that outliers should also be interpreted in economic terms (item 1.4). In an ideal situation, the statistical significance of the estimated parameters on the one hand coincides with the economic interpretability of the outlier variable on the other hand.

3 Examples of the empirical relevance of outlier treatment

The following section presents two examples. The adjustment of Argentine currency in circulation (cf. Burdisso et al., 2010, and Mehrhoff, 2010) illustrates that mistakenly deciding not to model outliers at the end of the series can lead to the wrong conclusions being drawn with regard to current economic developments. Furthermore, data about building permits granted for residential construction in western Germany (cf. Deutsche Bundesbank, 2006) demonstrate that failure to conduct outlier modelling can still produce misleading patterns in seasonally adjusted results one year later.

3.1 Argentine currency in circulation

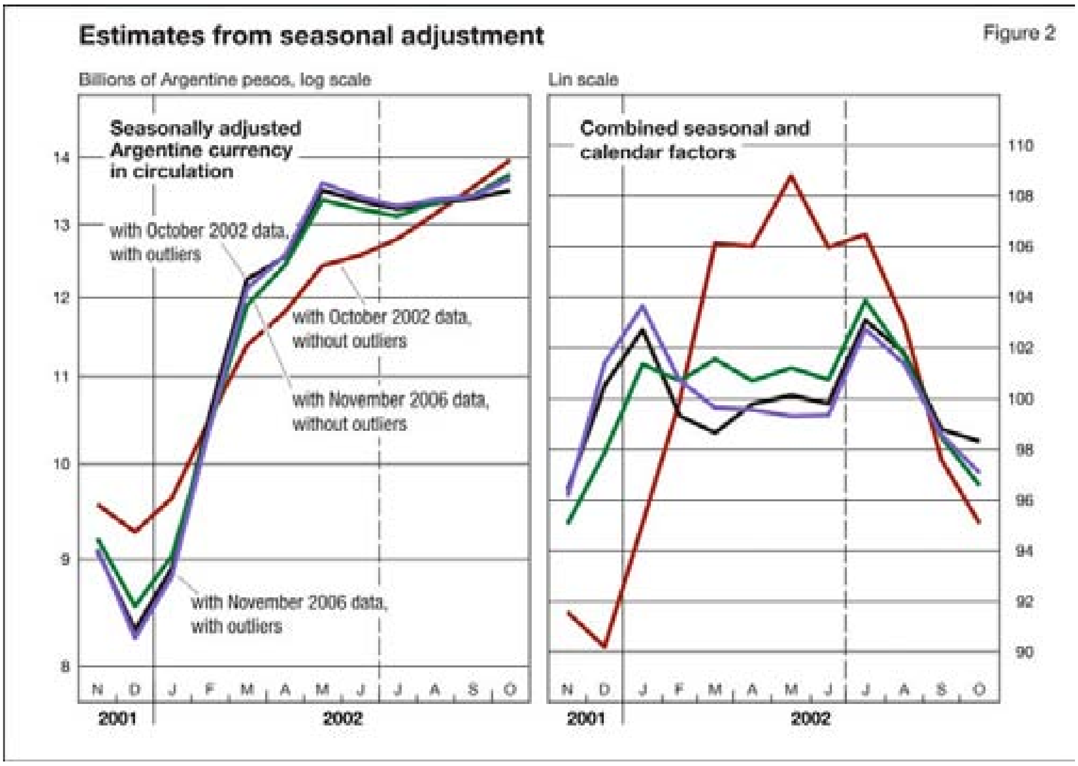
Figure 1 depicts the circulation of Argentine currency. The seasonal pattern is marked by legally guaranteed half-wage payments in June-July and December-January. It also shows the effects of the Argentine crisis in 2001-2002.



Four different estimations were carried out using TRAMO/SEATS to examine the effects of outlier modelling on seasonally adjusted results, in particular, during the crisis. The first stage was to determine the adjusted figures using the unadjusted values up to October 2002. If outliers are not modelled, this produces the red curve in the left panel of Figure 2 that rises continually from December 2001 to October 2002. However, if the statistically significant and economically explainable outlier variables for modelling the abrupt and strong effects of the Argentine crisis are taken into consideration, the pattern is different (black curve). Here the seasonally adjusted values of currency in circulation increases at a faster pace between December 2001 and May 2002 but then stagnates at this level. Seasonally adjusted data and economic conclusions thus differ considerably depending on whether or not outliers are modelled.

In a second stage, these results are compared with those obtained – for the same observation period – from estimates based on unadjusted data up to and including November 2006. This ensures that the results incorporate not only the figures before and during the crisis, but also those from the ensuing period. All in all, the results show that the curve calculated without modelling outliers (green) is quite close to the series estimated with outlier modelling (blue). In addition, both series confirm the findings made with the black line, which was estimated with the unadjusted values up to October 2002 and using outlier modelling. The red curve is far away from the other functions.

This is because the red curve estimation incorrectly interprets the special effects of the crisis as a change in the seasonal component and thus eliminates them (see the right panel of Figure 2). However, the seasonal conditions are guaranteed by law and did not change during or after the crises, as is also evident from the results on the basis of the long unadjusted series. The estimation using the unadjusted values up to October 2002 can take this insight into consideration only if outlier variables are used to model the crisis.

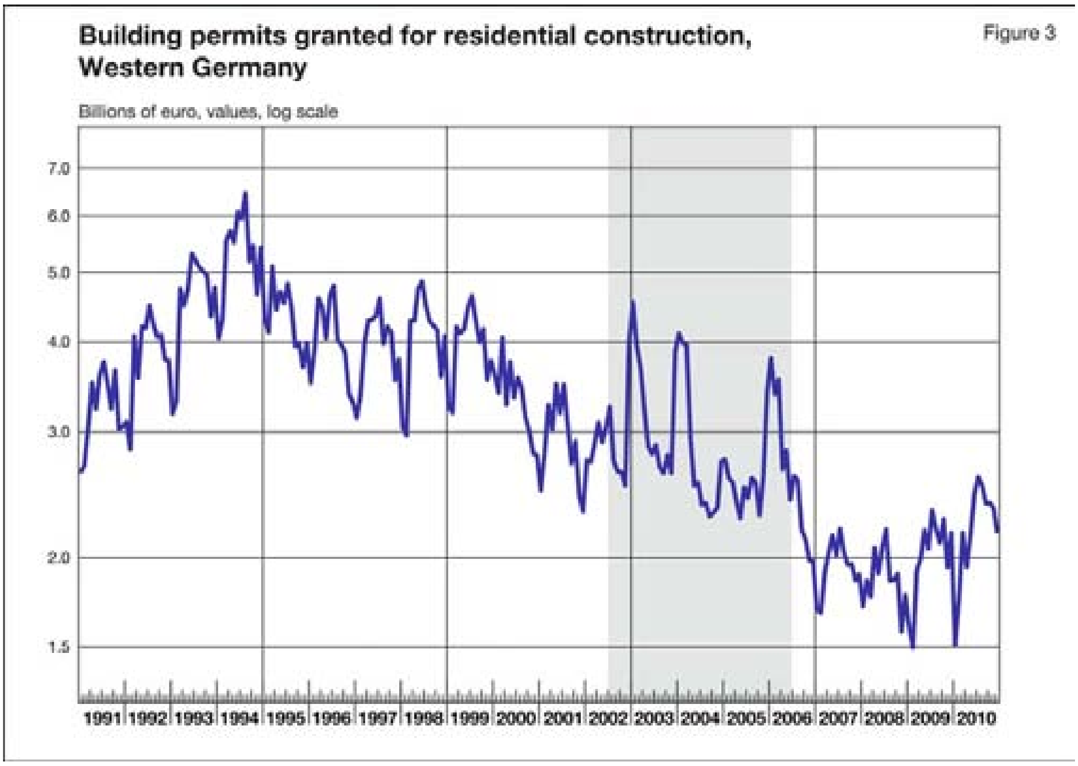


This example leads us to the following conclusion. If there is no indication that seasonal conditions have changed but there are significant and abrupt trend-cyclical or irregular influences, outlier variables (additive outliers, level shifts and the like) should be used to avoid a biased estimation of the seasonal component and consequently of the current end of the seasonally adjusted time series. Mistakenly deciding not to model outliers distorts the pattern of seasonally adjusted figures, which are gradually revised using new estimations as new unadjusted values become available.

3.2 Building permits for residential construction in western Germany

The number of building permits granted for residential construction in western Germany measured by the estimated cost of the building in Figure 3 shows a clear seasonal pattern up to mid-2002 and from mid-2006 onwards. During the winter (roughly from December to February), in which construction output is hurt by icy and snowy conditions, the number of building permits granted is much lower than during the rest of the year. However, every year from 2002 to 2005, there was debate as to whether government subsidies for owner-occupied housing (grant to homebuyers) would be significantly reduced or even phased out at the end of the year. In 2002, there were concerns that the grant to homebuyers would soon be cut. However, this did not happen initially. According to a first draft law, which followed in the summer of 2003, the grant to homebuyers was to be abolished from January 2004. The law was rejected and agreement could be reached only on a number of changes. The debate was reopened in 2004. It was then assumed that the grant would be abolished at the beginning of 2005. When this did not happen either, the start of 2006 was proposed as the date on which the grant to homebuyers would actually be abolished. As a result, just before the end of every year, homes were purchased or applications for building permission were submitted, which were then processed during the winter. Hence the effect continued beyond the end of 2005/beginning of 2006.

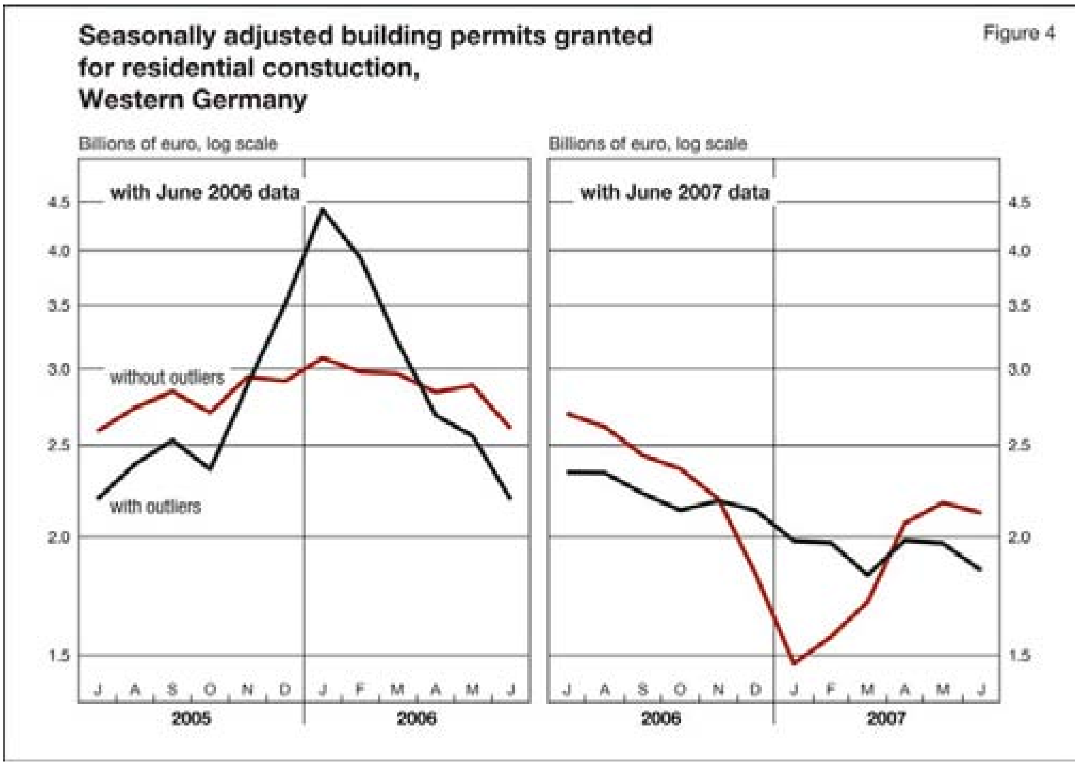
If it were assumed that the effects of the discussions about abolishing the grant to homebuyers inevitably recurred with similar intensity in the same season each year, then the effects observed in the period from 2002 to 2006 would be regarded as changing the seasonal component. Seasonal adjustment using X-12-ARIMA without outlier and extreme value modelling on the basis of the figures up to June 2006 therefore produces a fairly smooth curve for seasonally adjusted figures at the end of 2005/beginning of 2006 (red curve in the left panel of Figure 4.



However, in accordance with the available information, all unadjusted figures from July 2002 to June 2006 should be seen as extraordinary. This is possible in seasonal adjustment by estimating the seasonal component on the basis of unadjusted values up to mid-2002 and using this as a basis for a forecast up to mid-2006 using a RegARIMA model. The adjusted figures estimated using this seasonal component clearly show the extraordinarily high number of building permits that were granted before the grant to homebuyers was phased out (black line). If seasonal adjustment is not conducted in this way, the extraordinary nature of the situation at the end of 2005/beginning of 2006 would not be visible in the adjusted figures.

The way in which atypical values in the period from 2002 to 2006 are treated does not just influence the pattern of seasonally adjusted results during this period but also the estimation of the adjusted values in the future, for example at the end of 2006/beginning of 2007. If outliers are not modelled then this produces an artificial turning point (red curve in the right panel of Figure 4) although nothing out of the ordinary occurred during this time. The reason for this error is the incorrect estimation of the seasonal component. As no outliers were modelled, the high figures around the turn of the year 2002/2003, 2003/2004, 2004/2005 and 2005/2006 were wrongly classified as typical for the season. As a result, the seasonal component rose in the months in question. The normal unadjusted values at the end of 2006/beginning of 2007 were then adjusted with the values of this component that were too high. This then resulted in seasonally adjusted results being artificially reported as too low and an artificial lower turning point that is not in line with real economic developments.

This problem does not occur if the values for the period from July 2002 to June 2006 are treated as outliers. However, as an unusually high number of variables would be needed, an alternative method can be considered. This involves estimating the seasonal component for the period from July 2006 onwards using a modified unadjusted series. This series can be obtained by eliminating the abnormal figures from July 2002 to June 2006. For instance, the unadjusted value for June 2002 would be followed directly by the value for July 2006. The jump in levels which may arise from linking these values would have to be modelled as a level shift. This method would likewise ensure that the atypical values from 2002 to 2006 have no impact on the estimation of the seasonal component and the seasonally adjusted series from then onwards. Figures calculated in this way (black line) do not display an artificial turning point at the end of 2006/start of 2007.



Knowledge of the extraordinarily strong effects of the discussion about abolishing the grant to homebuyers results in a split procedure being applied in this case.

- The adjusted figures for the period up to June 2002 are calculated in accordance with the usual method.
- Forecasted seasonal factors are used for the period from July 2002 to June 2006.
- The seasonal component and the seasonally adjusted series are then calculated without taking account of the unadjusted values for the extraordinary years.

4 Conclusion

The role of outlier modelling in seasonal adjustment is to allow an estimation of typical seasonal fluctuations that is not distorted by special effects and thus to fully reveal "news" in seasonally adjusted results. Outlier modelling tools should be used wherever there are statistically significant effects that can be explained by economic developments. Otherwise there is a price to pay. The analysis of seasonally adjusted developments at the end of the time series can easily be misleading. Furthermore, one year after an incorrect decision not to model outliers, artificial seasonally adjusted movements can arise, potentially resulting in erroneous economic analyses. Both situations can be avoided by the proper use of outlier methods.

REFERENCES

Burdisso, T., Blanco, E., and Sardi, M. (2010), "El Ajuste Estacional y los Efectos del Calendario Doméstico en un Agregado Monetario para Argentina," *Ensayos Económicos*, 57-58, 145-186.

Deutsche Bundesbank (2006), "Economic Conditions in Germany," *Monthly Report*, May, 40-51.

Eurostat (2009), *ESS Guidelines on Seasonal Adjustment*, Luxembourg: Office for Official Publications of the European Communities.

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

Gómez, V., and Maravall, A. (1996), "Programs TRAMO and SEATS: Instructions for the User," *Banco de España*, Beta Version: November 1997.

Mehrhoff, J. (2010), "Ajuste Estacional en Tiempos de Cambios Económicos Fuertes," *Ensayos Económicos*, 59, 57-73.