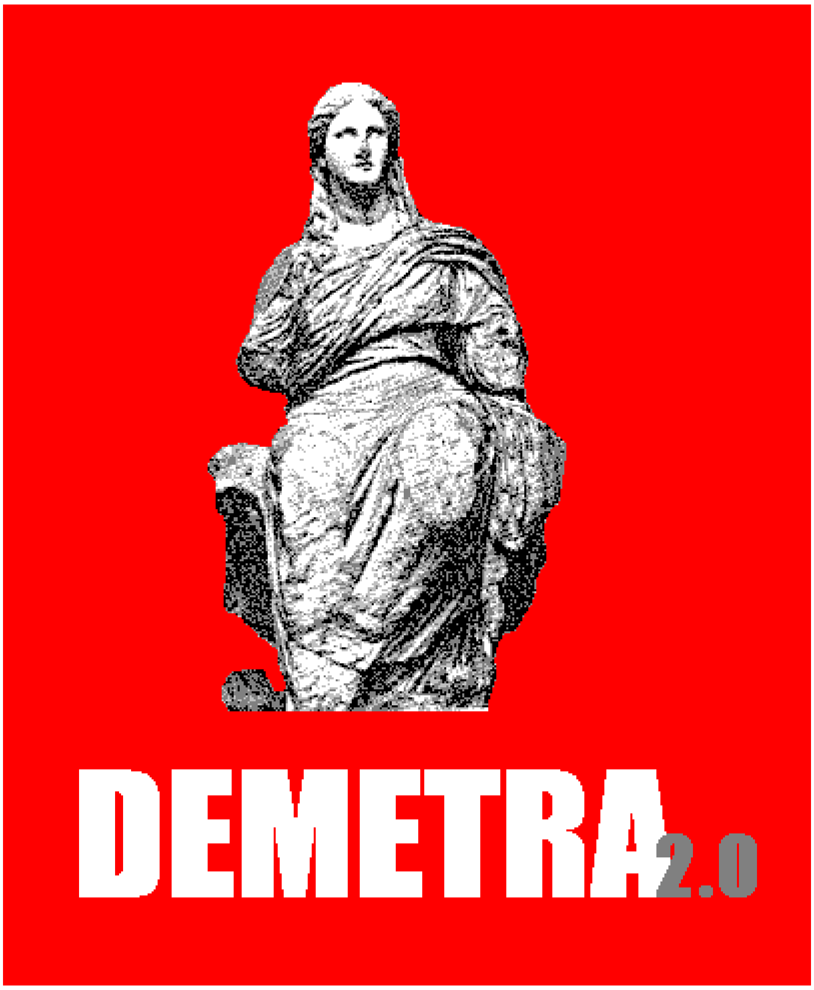
Seasonal Adjustment with Demetra



Pedagogical Manual

May 2002

Eurostat

the Statistical Office of the European Commission

TABLE OF CONTENT

I. SEASONAL ADJUSTMENT **3**

I.1. TIME SERIES 3
I.1.1. Stock and flow *3*

I.2. PRE-ADJUSTMENT 4
I.2.1. Seasonality *5*
I.2.2. Pre-adjustment schemes *6*
I.2.3. Irregular component and validation of the model *7*
I.2.4. Events and outliers *10*
I.2.5. Other calendar effects *14*
I.2.6. Seasonally adjusted data *16*

I.3. SEASONAL ADJUSTMENT METHODS 17
I.3.1. X12-ARIMA *17*
I.3.2. TRAMO-SEATS *18*

II. DEMETRA USER GUIDE **20**

II.1. INTENDED USE 20

II.2. FIRST STEPS 22
II.2.1. Selecting the method *24*
II.2.2. Options for trading days *24*
II.2.3. Exploring results by batch processing *25*

II.3. IMPROVEMENT OF REJECTED SERIES 30
II.3.1. Improvement of rejected adjustments *31*
II.3.2. Improving models by batch processing *36*
II.3.3. Expert system *44*
II.3.4. Processing in the detailed analysis module *45*

II.4. SPECIFIC CHARACTERISTICS OF X12-ARIMA 48
II.4.1. Parameterisation of X12-Arima *48*
II.4.2. Exploring results in the case of X12-Arima *49*

II.5. UPDATING AND CORRECTION STRATEGY 49

II.6. SOME TYPICAL SITUATIONS 51
II.6.1. Changing seasonality over the period under analysis *51*
II.6.2. Non-idempotence: seasonality in the trend *55*
II.6.3. Significant Ljung-Box: structure of residuals *57*
II.6.4. Too many outliers *57*
II.6.5. Incorporating a priori information *58*
II.6.6. Three statistics on residuals significant at 5% *65*
II.6.7. Non-seasonal data *66*

II.7. TO CONCLUDE 67

I. SEASONAL ADJUSTMENT

I.1. Time series

A time series is a string of observations about a variable over time. In almost all cases, and particularly for social and economic data series, the observation of this variable takes place at regular and consecutive time intervals: every year, every quarter or every month. In the following manual, the time series under analysis (original series) will be marked *yt* and we will only examine monthly or quarterly data.

I.1.1. Stock and flow

Variables of different types may be concerned. For the purposes of this manual, we will distinguish between two major types of variable:

4 stock type, or level; its value is associated with a given moment, for example, the balance of your bank account on 30 November 2001, the money mass in circulation or the number of jobseekers at the same date;

4 flow type; this concerns an amount associated with a time period as well as the total of expenses that you made during the month of November 2001, the total banknotes issued or the number of unemployment registrations at the same date.

It is important to know the stock or flow nature of the variable. Indeed, when it is necessary to change the time unit (grouping months into quarters or quarters into years), the processing methods are completely different. Thus, let us suppose that we have to build up quarterly data from monthly data. In the case of a stock, we can obtain either the value corresponding to that observed at the end of the quarter or provide an average of the three values observed at the end of each month in the quarter. For a flow, the quarterly value is the sum of the values for the three months.

Similarly, the unit required for expressing the value differs according to whether it relates to a stock or a flow. Thus, if we study a monetary total, a stock will be simply expressed in euro (€), or any other monetary unit, and a flow must be expressed in euro by time unit (€/month or €/year). In this last case, think about how to express your salary: monthly or annual?

The distinction between stock and flow may be pertinent in the case of seasonal adjustment, because the impact of calendar effects is not the same for the two types of variables.

The study of a time series can be motivated by two different contexts:

4 conjuncture,

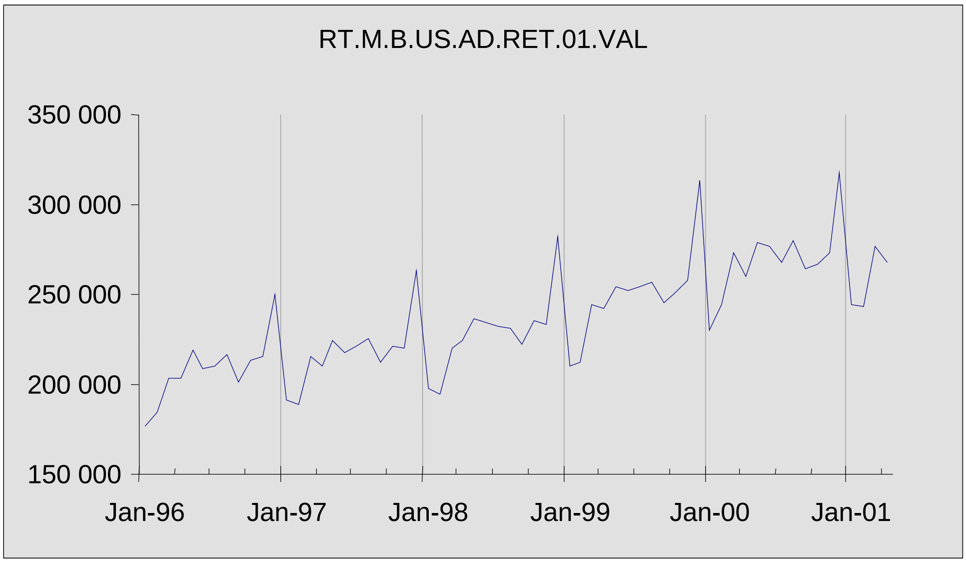
4 forecasting.

Conjunctural analysis aims to interpret the situation corresponding to the last data observed: what happened over the last months? What is happening today? Are we in a period of growth, stagnation, decline or upturn? Is the development observed structural or accidental?

Forecasting aims to … forecast! All automatic methods are based on the implicit supposition that information concerning the future can be found in past and present data. It therefore requires analysis of this information.

I.2. Pre-adjustment

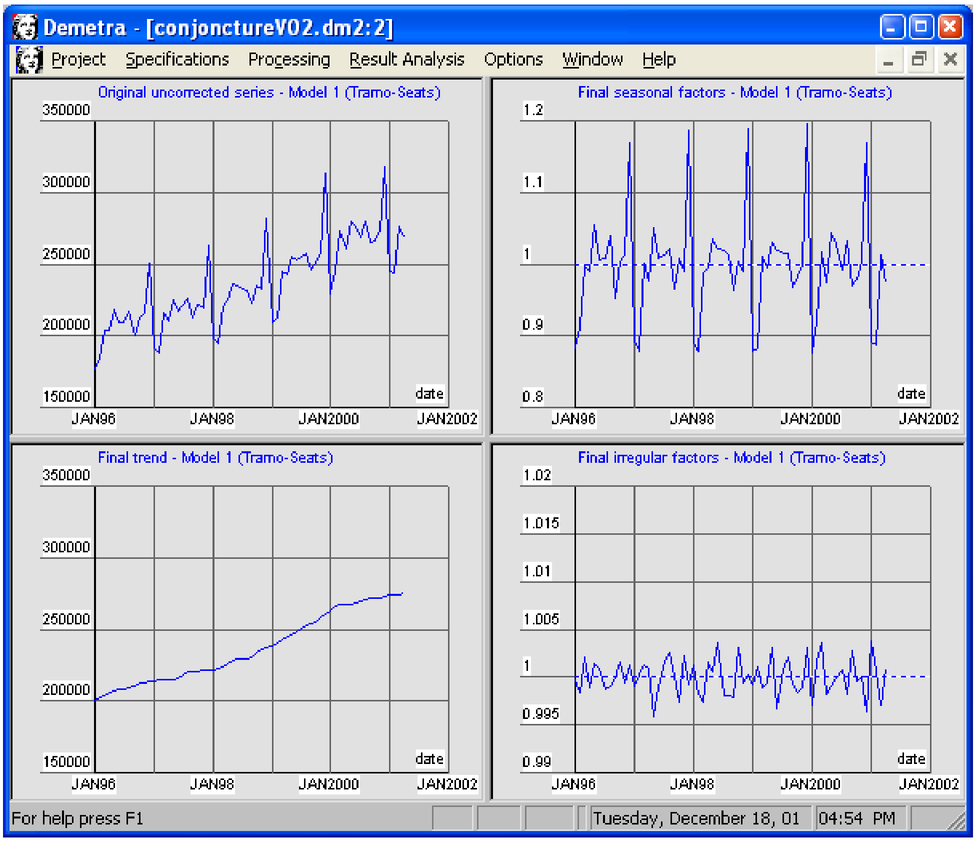
In the two contexts just evoked, a classical approach consists in trying to decompose the series into several elements or components. This approach is very natural; it is already in embryonic form in the cyclical analyst’s last inquiry and is entirely intuitive when one looks at the graphical representation of time series resembling this one (this one concerns retail sales in the United States).



It is visually easy to identify a pre-adjustment from the original series (top-left in the following figure):

- trend (bottom-left),
- and a seasonal component (top-right).

The bottom-right graph below represents the residual, ie. what is missing to reconstruct exactly the original series. You will note that this reconstruction is practically perfect, the residual component reveals a maximum difference of ±0.4%. But note the automatic scalings: the two right-hand graphs represent coefficients varying around 1, but the scale of the one above (seasonal component) is ten times larger than the one below (residuals).



The figure was produced using Demetra's default model, ie. by just validating the options proposed.

This analysis consists of 3 components:

4 trend (t_t),

4 seasonal evolution (s_t),

4 the remainder, marked i_t, for irregular, is at the basis of seasonal adjustment methods.

I.2.1. Seasonality

After this experimental seasonal adjustment method, we can introduce some more theoretical arguments. Human activity is subject to different original rhythms, the effects of which must be able to be observed in the result of this activity. Thus, the major natural rhythms such as the circadian rhythm (the Earth's 24-hour rotation) and the succession of seasons govern social cycles (alternation between night and day, holiday periods etc.). To these can be added purely social rhythms, such as the weekly rhythm and its effect on Sunday rest. The point in common between all these cycles is their perfectly known character and their almost-absolute predictability in the future. If we have ways of measuring the result of this human activity at an infra-annual rhythm, we can certainly find the trace of all the rhythms. It is therefore natural to try to estimate their impact and take account of them in the analysis of time series.

In the case of the economic cycle approach, this would avoid statements such as: in summer, unemployment decreases and rises again in autumn. We then correct the series according to seasonal variations which we call 'seasonally adjusted figures'. Similarly for forecasts, it would be logical to identify a smooth and easily scalable trend and a largely predictable seasonal component to estimate the future evolution of the series. This supposes, of course, that our pre-adjustment model will still apply in the future. In this case, the extent of variations of the residual component could also be used to estimate the uncertainty of forecasting and allow to calculate forecast bands.

To define the analysis, for interpreting more complicated series than that we have just considered, it will be necessary to introduce supplementary components.

I.2.2. Pre-adjustment schemes

Having envisaged splitting the original series into 3 components, we should now focus our attention on the mode of composition. Various solutions are possible, of which the simplest and the most classic (and the two implemented in Demetra) are additive or multiplicative composition.

In the case of additive composition, we will suppose that the three components reconstruct the original series using the following formula:

$$y_t = t_t + s_t + i_t$$

For a multiplicative composition, the formula will be:

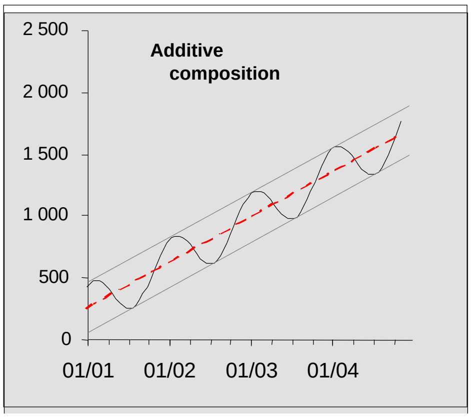
$$y_t = t_t \times s_t \times i_t$$

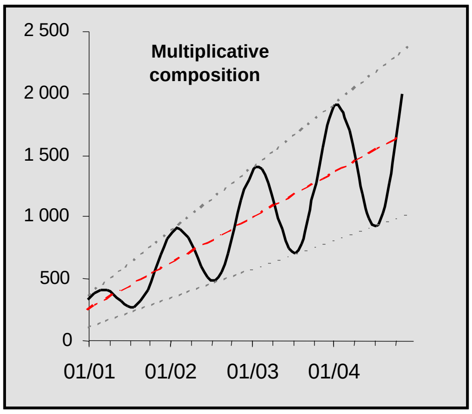
To understand the difference between these two mechanisms, we can consider the units of the different components: in both cases, y_t et t_t are in the same units, those of the original variable, in the additive case, s_t et i_t are also in the same units; however, in the multiplicative case, they become simple dimensionless coefficients. Thus, in the first case, seasonal variations are absolute variations and are expressed in the form + 5 000 € or – 3 000 €. Inversely, in the multiplicative case, the coefficients are expressed in the form 1.20 or 0.95, without any unit, ie. in an equivalent manner + 20 % or – 5 %. To view these coefficients as a percentage, one can also write the multiplicative composition model in the form:

$$y_t = t_t \times (1 + s_t') \times (1 + i_t')$$

In the multiplicative model, the seasonal variations are proportional to the level of the series.

The graphs below illustrate the difference between these two models in the case of an increasing series. For an additive model, seasonal variations are contained within a band or a tunnel parallel to the tendency. In the multiplicative model, the variations increase when the level of the series rises and reveal a cone axed on the tendency. The cone is divergent when the series rises, and convergent when it decreases.





By studying the graphical representation of the series, we can determine the pertinent model of composition. It should be remarked that the two models are indistinguishable from one another if the level of the series remains constant throughout the period. We then say that the pre-adjustment model is unidentifiable.

The additive and multiplicative models are available in Demetra. In Tramo, the multiplicative model is produced by transforming the original series by logarithms (log-transform) beforehand. The default option consists in establishing the necessity of this prerequisite transformation (pre-test); the model selected is thus – in theory – additive, except if it is judged necessary to use a multiplicative model instead. For the rest of this manual, we will suppose, unless otherwise stated, that the composition is produced in an additive manner.

Generally, it should be noted that the identifiability of the pre-adjustment model poses a real problem: starting with **one** data series, we split it into **three** series. It is of course clear that there is an infinite number of ways to produce such a pre-adjustment. They all must satisfy the equation:

$$y_t = t_t + s_t + i_t$$

To limit the number of possibilities and arrive at a reasonable solution, we need to clarify the constraints to which the various components are subject. The trend must be a series which is rather smooth, evolving over the long term and without disturbance.

Seasonality should be periodical or almost, that is, reoccuring from one year to the next in the same manner or approximately. We can envisage observing a slow modification of the seasonal component if the period studied is sufficiently long. The average effect of the seasonality over a year must be zero or, if preferred, the annual total of the original series' values must coincide with the total of the series' values excluding seasonal variations over the same period.

I.2.3. Irregular component and validation of the model

The irregular component appears thus as an adjustment term, a residual, indispensable for preserving the accounting equation. It plays, however, a fundamental role in the statistician's approach: it is a key to validating the model. Indeed, this term of adjustment must not contain information since this would totally invalidate the interpretation of the other components. In technical terms, we can imagine that it is necessary to extract an interpretable signal from the original series and ignore any noise that accompanies the observation. If, after the pre-adjustment, there is still some signal in the noise, this is because the model selected is inadequate. In statistical terms, we will consider that the irregular component resulting from applying the model must present all the characteristics of chance variations. This is sometimes referred to as white noise.

To be concise, the irregular or residual component, must absolutely test the following hypotheses:

1. Average zero; in practice, this hypothesis is rarely restrictive and is automatically tested by the methods used.

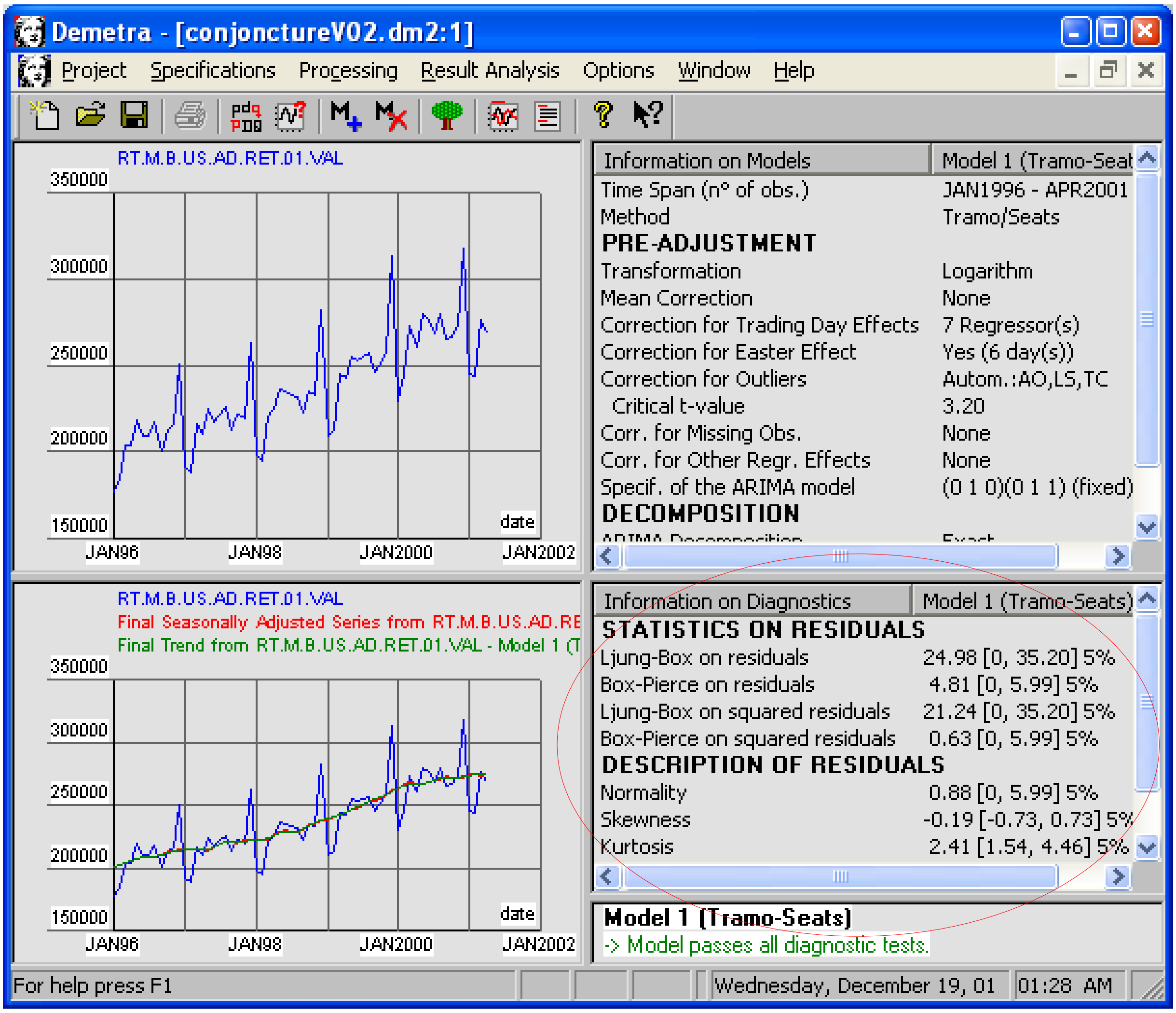
2. Different values of the irregular component must be independent. This means, among others, that one must not be able to predict a value of the residual according to the previous values.

3. Be distributed according to a normal law. The normal law or Gauss law is the model of a purely random variable. This is mathematical modelling of a noise without any information.

Testing the last two hypotheses is imperative for validating the pre-adjustment model. This is why Demetra always shows the elements that allow us to appreciate these hypotheses in the results provided in order to diagnose the model selected (see the following image: bottom-left). The information is presented slightly differently in the automatic processing module and in the detailed analysis module (reproduced version below).

The calculated values must be within the interval which follows the value; if this is not the case, the corresponding line is shown in red.

The statistics on residuals concern the first set of hypotheses. A value outside the specified interval is a sign that there is outstanding information in the noise. We then say that the value is significant. The descriptive part of residuals compares the distribution of values to the distribution of a normal law. In the same way, if one of the numbers is outside the interval, we can conclude that the distribution of residuals is significantly different from a normal law.



Finally, it is useful to compare the magnitude of seasonal variations to that of the variations resulting from the irregular component. Indeed, if the seasonal effect is lost in the noise, it is probable that the use of a seasonal adjustment model for the analysis of the series is not indispensable. We then say that the seasonal component is not significant.

We have just presented a seasonal adjustment model with three components (trend, seasonality, irregular component) as well as the principle of validating the model, from which we can see the necessity that the irregular must imperatively present ‘good properties’. However, in this irregular component, some useful information can remain unexploited by our first, simple model. It is therefore necessary to give details of the contents of our first residual by introducing new components.

This will include:

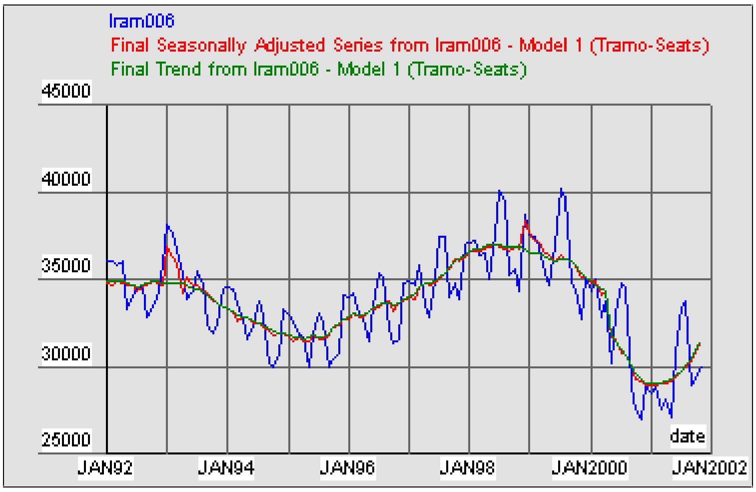
4 Events having marked the history of the series. Many events can affect the time series: technical problems, strikes, changes in nomenclatures or calculation methods, follow-up measures, ‘announcement’ effects. These events can be known and, in theory, introduced into the model by the statistician; they can also be automatically detected by Demetra. It can therefore be worthwhile to try to identify the events that Demetra discovers.

4 Other calendar effects not linked to a seasonal variation: correction for trading days, impact of Easter or the correction for leap years. The characteristics of the calendar are known and predictable, their effects can impact upon the level of the series; Demetra can detect and estimate them.

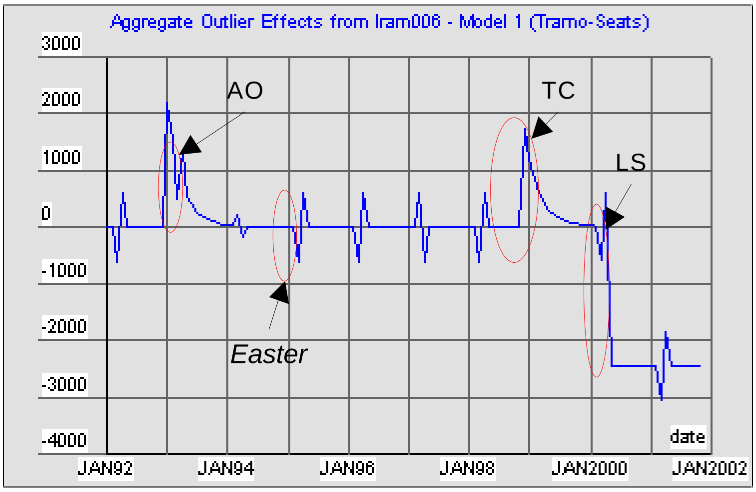
These different components are going to provide as many prior corrections. They are summarised in the top-right section of the project window under the rubric *Pre-Adjustment.* For the series presented in the copy of previous the screen, most of these options have been activated.

I.2.4. Events and outliers

In the following example (number of women receiving unemployment benefit in Ireland) Demetra's default options, the Tramo-Seats model, give the following the results: seasonal adjustment (accepted) and trend:



A priori adjustment term:



This last graph is obtained using the *Aggregate Outlier Effects option*. It groups the various elements detected automatically, AO, LS and TC as well as the calendar effect (*Easter*). It can be remarked that some of them are sufficiently significant to be spotted on the representation of the original series, of the series corrected for seasonal variations and the trend.

We arrive thus at a pre-adjustment of the irregular component into two sub-components:

- ev_t: grouping the 'events', ie. all the *a priori* corrections,

- i^*_t: the 'true' residual, consisting only of a random component.

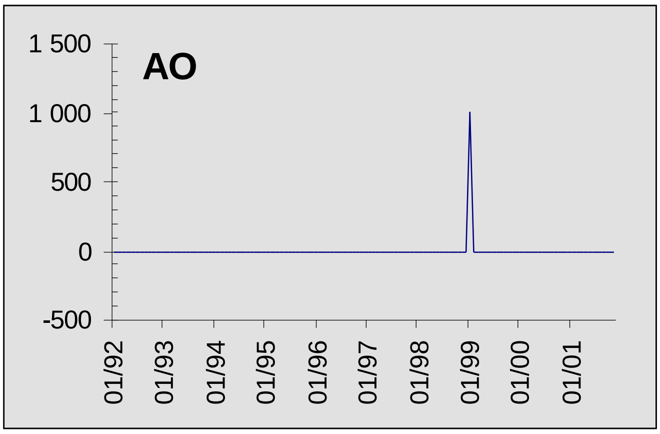
The events described by a correction term are, in principle, identifiable or 'susceptible of being identified'. Ideally, the cyclical analyst should identify them; either to find well known events of economic life (significant strike, violent weather etc.) or to find out what could have impacted on the series at a certain date. In this last case, we often initially think of 'true' events, in the sense of climatic, social or political events. But 'false' events should not be overlooked: the artefacts connected with the production process of the data under analysis. Almost anything can be imagined here: from changes in nomenclature to input, transmission error or, to remain in the environment of the example presented, from the destruction of postal sacks in an employment agency following a flood, to the implementation of a more intense policy of striking people off a list.

As can be seen from the several examples cited, the effective identification of accidents, ie. their interpretation requires a thorough knowledge of the socio-economic environment but also of the rules and conventions determining the definition of the data under analysis, their evolution and the administrative environment in which the data are produced.

In the framework of this manual, and unfortunately, as regards the current practice of the cyclical analyst, we will limit ourselves to describing the different types of event that could arise. Each one is presented below, in additive version, at the same date and in an identical characteristic value.

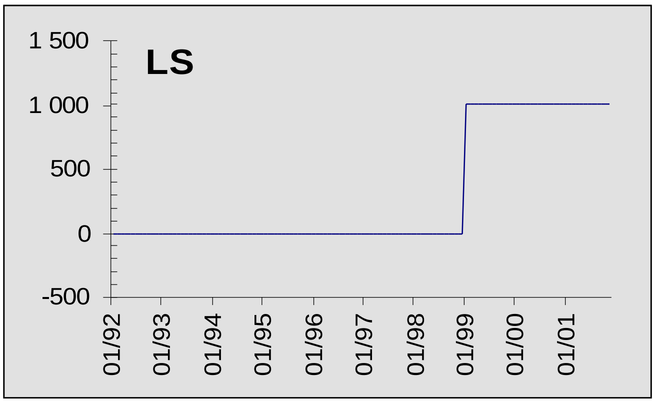
Ø *Additive outliers (AO)*

'Additive values' correspond to an isolated point and may concern the effect of a strike, a weather event or a flood, etc. In the pre-adjustment model, this type of event is integrated into the irregular component.



Ø *Level shift (LS)*

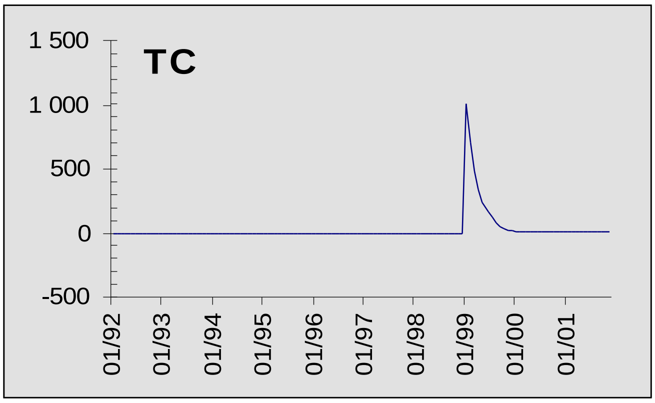
'Level shift' corresponds to a break in the average value of the series. The change remains a given. This type of event arises, for example, when the definition of the series is changed, the nomenclature is modified …



Level shift is processed by Demetra as a component modifying the trend.

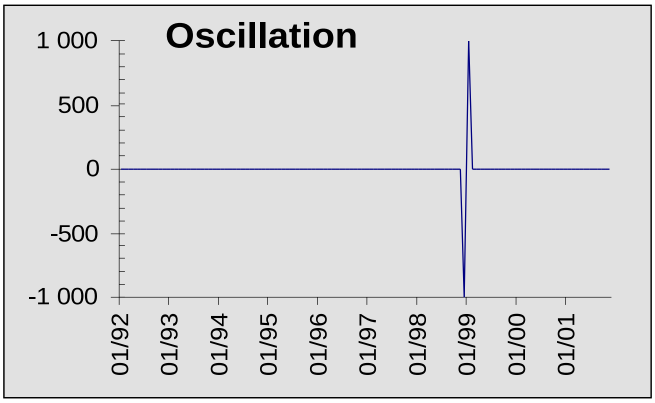
Ø *Transitory change (TC)*

'Transitory change' describes a disruption with a return to the initial situation. To describe it, we need, like the previous ones, a size, but also the time it takes to return to the initial situation.



A transitory change is an event of an 'irregular' type.

We should not forget one type of shock that we frequently find in certain types of data: oscillation.



It manifests itself by phenomenons where something is postponed from one period to another. It can result from announcement effect: setting up of a bonus to take effect from a given month or, inversely, the date when the bonus ends.

Although it is not considered as a type of event, we can find (or describe) an oscillation in the form of consecutive outliers in the opposite direction. We should also note that the Easter effect can easily be located because of such an oscillation; it is just simply a postponement from one period to the following when Easter falls at the end of March (from March to April for monthly data; from the first to the second quarters for quarterly data).

Thus, to conclude this part, we can consider that our pre-adjustment can be written as follows:

$$y_t = t_t + s_t + ev_t + i^*{}_t$$

in the additive version, with an equivalent formula for a multiplicative model.

I.2.5. Other calendar effects

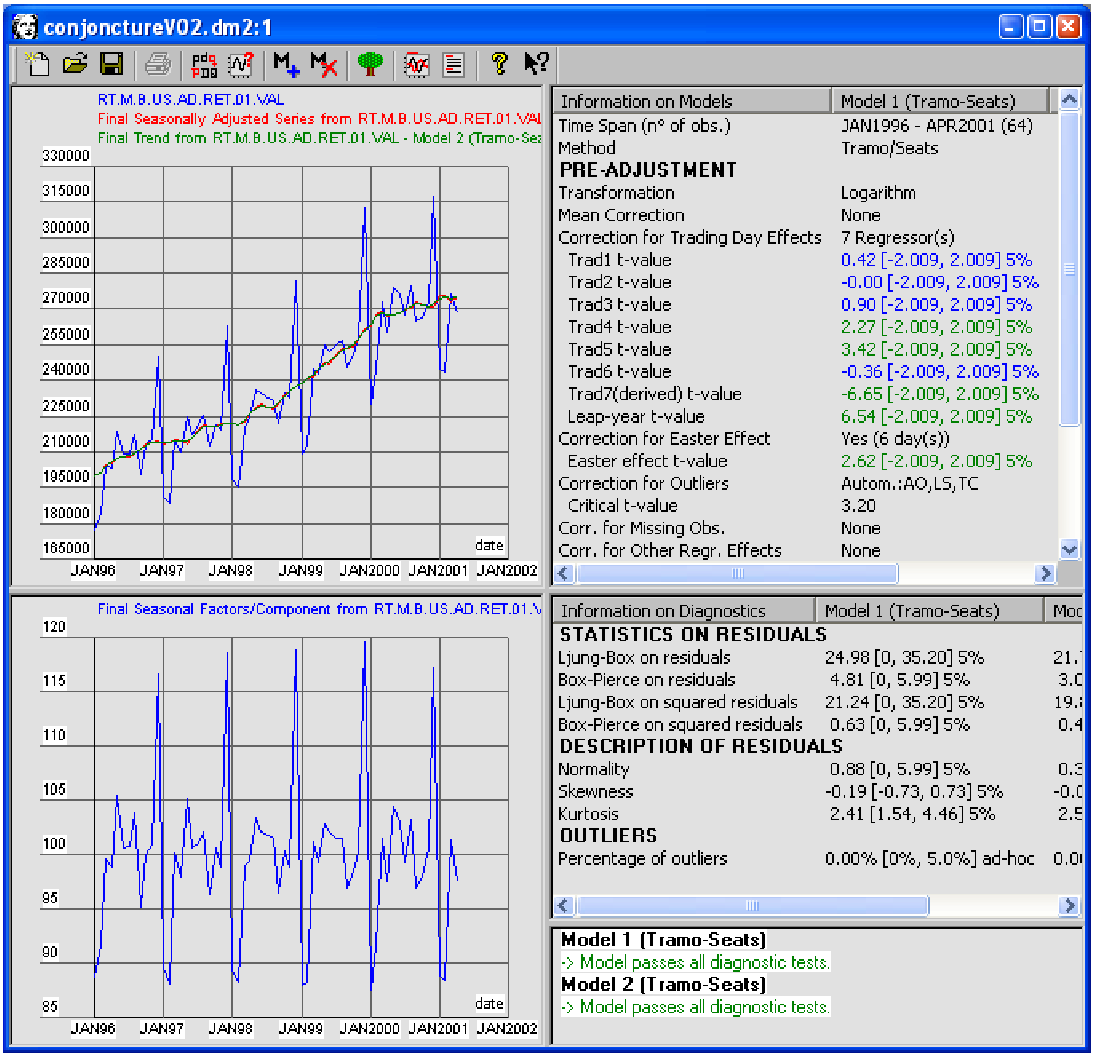
Several techniques can be used to take seasonality into account and, more generally, for integrating calendar effects.

The different calendar months, like the four quarters of the year, are of varying lengths (28, 29, 30 or 31 days for the months; 90, 91 or 92 days for the quarters). This can have consequences on flow type data, whose value depends on the length of the period. Usually, this influence is incorporated into the 'pure' seasonal effect. But it is also possible, in X12, to distinguish the pure seasonal effect, linked to the month, from the length effect of the month. Finally, months are, in general, the same length from one year to the next, *except* for February where it is therefore necessary to introduce a correction for leap years.

In addition, months do not contain a whole number of weeks. Excluding non-leap year February months, they contain therefore one, two or three supplementary days. Therefore, the different days of the week do not all appear the same number of times in each month: some appear 4 times, others five. However, a large amount of social activity is conditioned by the weekly rhythm: traditionally, you marry on a Saturday, Sunday rest is the rule in Europe, and in France many shops are also closed on a Monday, and so on. We can see therefore that the composition of the month in weekdays can have a direct effect on the level of the series.

Here, too, this remark is particularly well-founded for flow type data. As an example, the sales level of the retail and distribution sector in December is directly linked to the number of Saturdays in the month; December numbering 31 days, there are, *a priori*, 3 out of 7 years containing 5 Saturdays and 4 containing just 4. The comparison between these December months must imperatively take account of this parameter. These adjustments are grouped under the umbrella term of 'adjustment for trading days' or components (*trading day effects*). They allow to calculate a 'coefficient' for each weekday describing the intensity of the activity for this weekday relative to the average value.

In the example below, which presents the monthly retail sales evolution in the United States, we can note the presence of an adjustment for trading days: thus, Sunday (day 7) presents a very negative factor, the Thursdays and Fridays are days of greater sales. The sum of the effects of the seven weekdays is, by construction, zero.



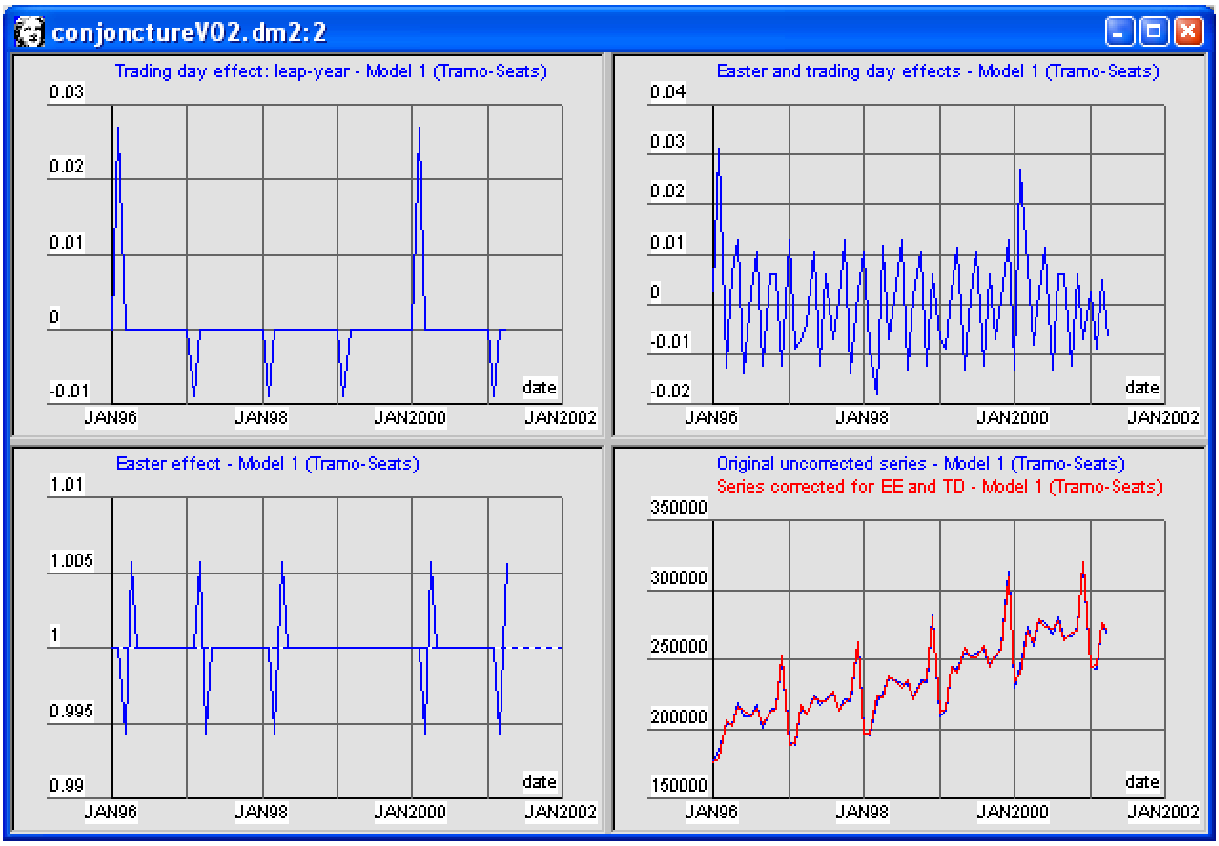
We can also note the presence of an Easter effect and a leap year effect. The copy of the screen below, extracted from the detailed analysis module, presents the various components linked to calendar effects:

4 in the top-left, the leap year effect, each February, zero on average over 4 years, positive in the leap years, negative in the other three;

4 in the bottom-left, the Easter effect;

4 in the top-right, all the adjustments linked to the calendar, other than seasonal variations. They cover 2 previous components plus all the weight effects of each weekday; and finally,

4 the last graph allows us to compare the original series and the series adjusted for calendar effects. When adjustments are pertinent, the adjusted series should look more regular than the original: the periodicity should be more marked.



Lastly, we should note that Demetra proposes, by default, to test the necessity of these different adjustments and to apply them only when the estimated effects are judged significant. In the example above, the effects are judged significant if their value lies outside the interval mentioned (for all the parameters, the interval is [-2.009 ; +2.009]). The further the value is, the more the effect is significant. Concerning the weights of weekdays, if one of the days has a significant effect, the adjustment for trading days is used.

I.2.6. Seasonally adjusted data

Traditionally, a seasonally adjusted series is one from which the impact of the seasonal component has been withdrawn. This data, of course non-observed, makes it possible to estimate what the value of the series would be if the seasonal effect were not taken into account.

By applying this definition with the notations of our pre-adjustment and by using the accounting equation, we obtain the following different versions:

$$\begin{aligned} cvs_t &= y_t - s_t \\ &= t_t + i_t \end{aligned}$$

We will remark therefore that seasonally adjusted data can be viewed as the sum of the trend AND the irregular component. In other words, it contains the trend, but also the fluctuations, which are as random as the events. The practice, very widespread today, of publishing seasonally adjusted data means that it is commonplace to consider that this statistic is combined with an indication of trend. When the cyclical analyst publishes seasonally adjusted data, they have to understand that these data still need to be interpreted: is it possible to distinguish, for this last observed value, between what is structural and what is short term? The answer is obviously no! It would be an illusion to hope to find the answer by consulting the (various) parameters and indicators generated by the model: if a new event is currently underway, or has been for a few months, we cannot see how the model, however sophisticated, could originate in the past.

A recent example clarifies this position. If the cyclical analyst observes that their data for September 2001 are ‘atypical’, it is up to them to pose questions about the cycle and look around, in relation with the series under analysis, to see whether there has not been some event which could have an influence on what is described by this series. Nothing, in any model, will tell them what they should look for. It is their knowledge of the economic, political, social, meteorological situation as well as their knowledge of how a sector functions that will suggest the hypotheses that they can postulate.

I.3. Seasonal adjustment methods

Demetra makes it possible to choose, for each series under analysis, between two methods belonging to distinct families:

- Methods based on econometric modelling of time series, or parametric methods. The choice is SEATS of TRAMO/SEATS; the method was proposed in 1996.
- Methods of an empirical origin, not requiring explicit recourse to a model, also called non-parametric methods. Their origins go back in time, based essentially on moving average methods. Demetra proposes the X12-ARIMA method, derived, also in 1996 from a line passing via X11, in 1965, then X11-ARIMA in 1975.

I.3.1. X12-ARIMA

In principle, this method is based on moving averages. We are going to present it rapidly in supposing monthly data.

With the seasonal component – supposedly periodic or quasi-periodic, of an annual zero average effect – the idea comes very naturally to calculate the averages of the series over a period of 12 consecutive months. The result of this calculation, attached to the median period of time (the month located in the middle of the 12-month period) is therefore ‘filtered’ of its seasonal component, and we obtain an estimate of the trend.

By withdrawing this trend from the raw series, the deviations from the trend can be calculated. The equation of pre-adjustment indicates to us then that these deviations from the trend contain a seasonal component and a residual term. The following stage therefore consists in grouping together the deviations corresponding to the same season (the same month) and calculating the moving averages (for example, the average of deviations of March over 5 consecutive years). The values thus obtained can then be used as an estimate for seasonal coefficients.

Using this rather simple procedure, the methods of types X11 or X12 will multiply the repetition of this process a certain number of times. They are the ‘stages’: after stage A (the preliminary stage), the B, C and D stages will each repeat these filtering techniques twice.

Part of the complexity of this multiplication stems from the fact that in the B and C stages, the method used aims to detect additive outliers (AO) by trying to locate those values that are too significant in the irregular component. The series is then adjusted for these outliers and used for the next stage.

Another tricky point regarding the X11 type methods arises when it comes to focussing on the ends of the series under analysis. If the starting point is not necessarily a major issue, the last observed value, however, has particular importance; also in the cyclical analyst’s approach. It is especially because of these last points that interpreting the seasonal-trend-irregular pre-adjustment becomes very difficult. However, as we have indicated above, the moving average over 12 months necessarily leads to a 6-month time lag in estimating the trend.

Two solutions are proposed for dealing with ends of series:

- Use a different moving average from the simple equal weighted arithmetic average (the average of 12 values). This implies introducing a weight for each of the 12 values whose average must be calculated, and this weight increases the influence of the very last points in the series. However, these points are, in theory, less well known: they can be provisional or subject to revision. This method is the one initially retained in X11.
- The other solution consists in stretching the data series at both ends. The idea is to adjust a model to the data and to use it for extrapolating (or retrapolating, to the past) these data in order to have (fictitious) points going beyond the ends observed, and on which it is possible to base a calculation of a moving average which is not (or less) unsymmetrical. The prediction uses the Box and Jenkins methods, identification and estimation methods for ARIMA models. Hence the name X11-ARIMA or X12-ARIMA.

Although we meet again the modelling technique at the basis of TRAMO/SEATS, it is important to understand that its role, in the framework of non-parametric methods, comes into play at a much less fundamental level: it is more like a method for improving the estimation of the ends of a series. What is more, it is very possible that X11-ARIMA is unable to adjust an ARIMA model to complete its points; this does not affect the method which falls back on the initial algorithm.

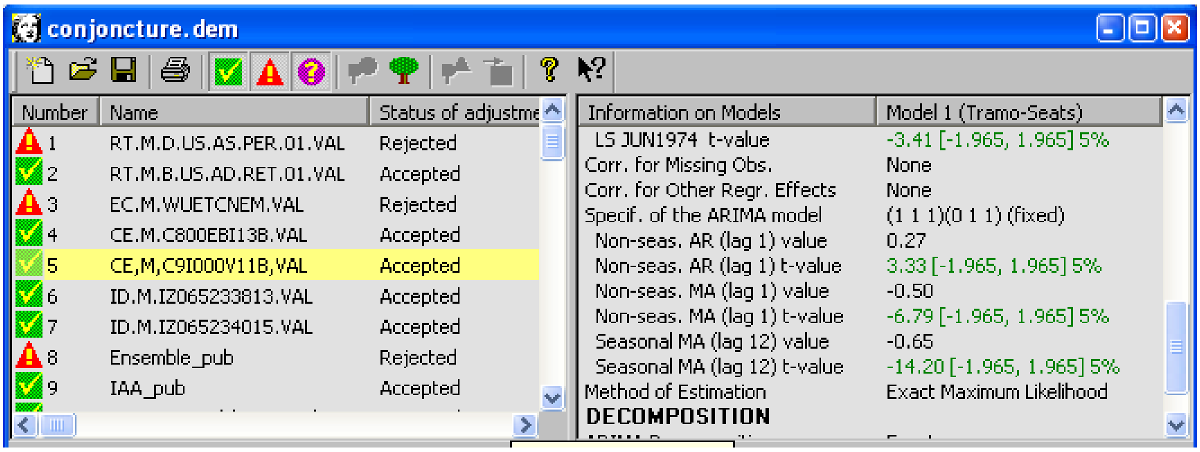
I.3.2. TRAMO-SEATS

Contrary to the 'empirical' methods presented above, SEATS fundamentally focuses on the modelling of the data series. The family of usable models is provided by the theory of stochastic processes and is known under the name of ARIMA models. In such models, the value of the series at a given moment is determined by several components:

- A part called 'auto-regressive' – the AR of ARIMA – which means that the value is determined by a relation with the previous values of the series. The number of previous values (or lags) used in this relation is said to be the order of the AR part. It is often denoted by the letter p, in the models used, it can vary between 0 to 3.
- A part called 'moving average' – or MA. This is a random component. It brings in the value of a noise variable, but this variable can be a structure-less noise, or white noise, which means that the different values of the noise are independent from each other, or, to the contrary, possess a certain structure: the successive random terms are correlated with one another. The temporal extent of this structure is described by the order of the MA part, denoted by q. It indicates the number of lags to use for determining the random part from a white noise. Usually, its value is between 0 and 3.
- Finally, the I, for *integrated*, indicates whether the series itself should be used or its variations (or derivatives). The differential order, denoted by d, indicates whether the model simply uses the variable itself ($d = 0$), if it must contain its derivative (its variations, $d = 1$) or its second derivative ($d = 2$, maximum value used by SEATS).

Integrating seasonal variations in ARIMA models is done by building up another seasonal model using the basic model presented above as a template. The basic model deals with expressed lags as a number of observed values, ie. in months for monthly data and in quarters for quarterly data. The seasonal model describes the link between the value of the series at a given moment and its value in the previous year. The lags are now expressed as a number of years. To describe this part of the model, we introduce therefore three new parameters corresponding to the respective orders of the auto-regressive, integrated and moving average parts of the seasonal model. They are respectively denoted *sp*, *sd*, *sq* and are limited to the values 0 and 1 in SEATS.

To sum up, the type of model used to describe the series under analysis is described (we say specified) by the two triplets data (*p*, *d*, *q*)(*sp*, *sd*, *sq*). This information on the model used is indicated in Demetra outputs. The selection phase of a model is called model identification. By default, Demetra identifies the model to use for describing the series, but allows specialists in theory the possibility to choose the model themselves. Afterwards, the ARIMA methods must estimate the coefficients of this model. The number of coefficients to calculate corresponds exactly to the orders retained, excluding differencing.



In the figure above, the series selected is described by an (1, 1, 1)(0, 1, 1) order model. We therefore find three coefficients: AR(1) and MA(1), equal to 0.27 and –0.50, associated to the orders *p* and *q* (**1**, 1, **1**)(0, 1, 1) and MA(12) corresponding to the order *sq* (1, 1, 1)(0, 1, **1**), equal to –0.65. Their values are followed by indicators of significance (*t-value*); in the case of a model retained automatically by Demetra, they are significant (*t-value* outside the interval) for the coefficients corresponding to the highest order of each part.

One of the difficulties in ARIMA type methods stems from the sensitivity of the method to disturbances. In fact, if a time series is subject to events or breaks (see above), they are going to have a strong influence on the calculations necessary for identifying the model and estimating the parameters. It is therefore very important to set up a fine detection mechanism for such disturbances in order to guarantee the quality of the model selected. This is where TRAMO from TRAMO/SEATS comes into play. It is also the reason why the *a priori* adjustments of the series are more detailed in the TRAMO/SEATS option than in the version of X12-ARIMA used in Demetra.

II. DEMETRA USER GUIDE

II.1. Intended use

Demetra has been designed by statisticians – in an environment of cyclical analysts – for cyclical analysts. The programme therefore takes their work organisation into account. They receive the latest updates of raw data and must provide an interpretation of them; this repetitive task is likely to concern a sizeable, if not very significant, number of series. For this they use, among others, seasonal adjustment methods and therefore need to:

- choose a method and a model,
- apply this model for calculating seasonally adjusted data.

These two parts are subject to distinct cycles: the use of the model, deseasonalisation, is carried out on a monthly basis for publishing data, the choice of the model or the reestimation of its parameters is not done, generally, for each publication but less frequently, typically about once a year.

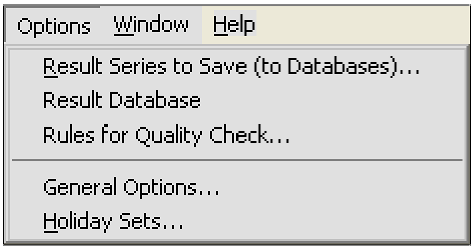
The large volume of data to process justifies maximum automation, as much for analysis as for seasonal analysis. This is where the 'automated module' comes into play. It aims to cover current production needs and supposes that the series are organised by 'project'.

The 'detailed analysis module' allows the statistician to carry out a particular study on a series to which they can devote their time to carry out research on an adapted model; if the series is especially important, or if it presents strong particularities requiring more advanced work. It makes it possible to test different methods or choices of a parameter on the same series, to store the results of these tests and to make a comparison especially in graphic form.

Let us now follow the production steps and accompany the cyclical analyst in his work. We shall start with the definition of a new project. This requires:

- defining the series to process: choice of file containing the data and determining the period for analysis,
- choosing the results to store,
- as well as choosing where (which file) to store them,
- the statistics used and the rules for selecting models.

For this, the *Demetra Project Wizard* takes the user through five screens of options, the last point above in two screens. All these choices hold for the series of the same project and are stored in the definition of the project. They can be modified at a later date by going into the *Options* menu. The first three elements of this menu correspond exactly to the four screens of the assistant for new projects.

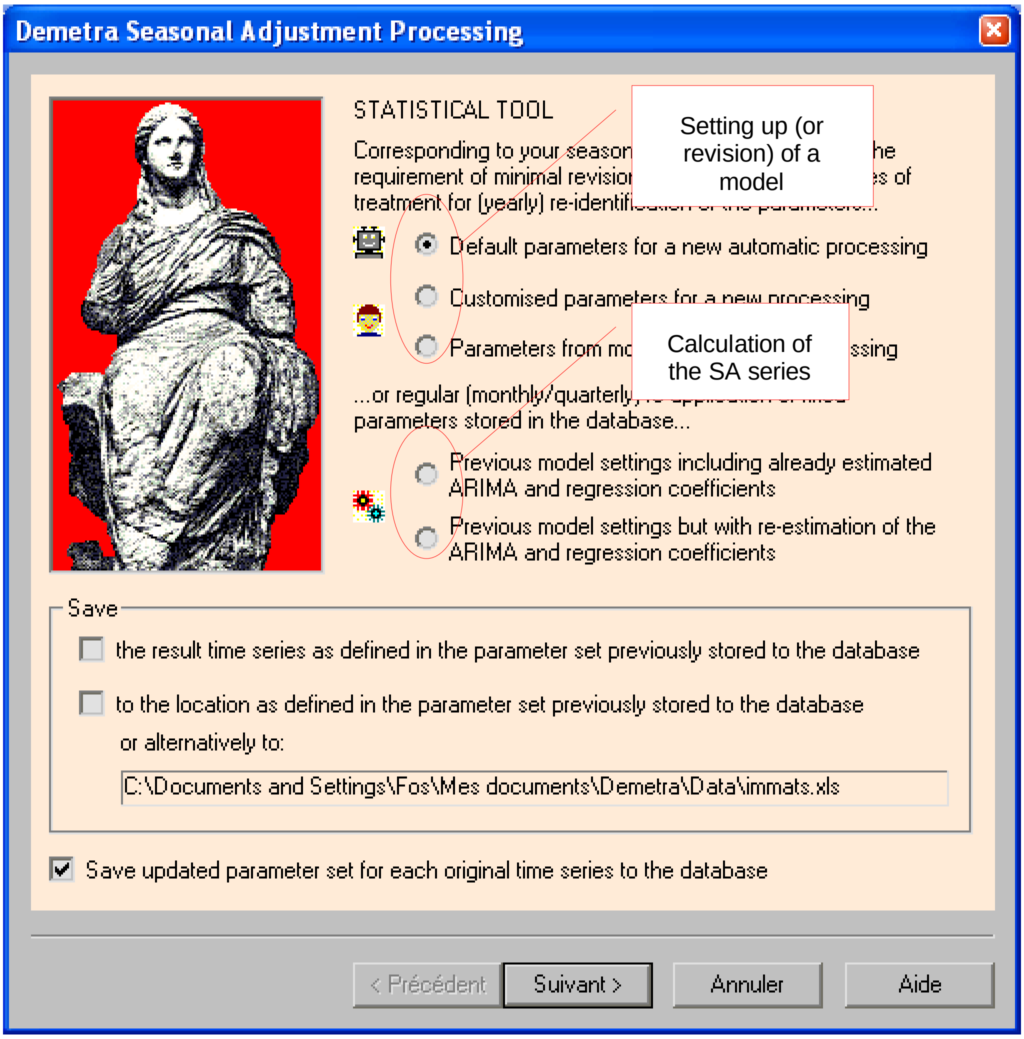


The different general options can be modified at a later stage using the menu above. The analysis period can be modified each time a new adjustment is launched.

After selecting the project's parameters, the next step is analysis using *Demetra Seasonal Adjustment Processing*. First, we need to choose the type of process to apply:

4 seasonal analysis (identification and estimation of the model), or

4 deseasonalisation (calculation of the seasonally adjusted data).



In effect, these two activity levels are clearly distinguished in the initial dialogue.

The first three radio buttons correspond to the determination of seasonal pre-adjustment models (seasonal analysis). The term, 'new processing', indicates identification, or reidentification of the seasonal adjustment model. The three options proposed are:

- The default option. In short, Demetra tests the necessity of different possible corrections and includes them if it judges that they improve the model.
- The second choice allows users themselves to fix the different corrections to implement or disactivate.
- The third supposes that these choices have been made and stored beforehand in a file describing the model.

The two following options correspond to the 'deseasonalisation' phase:

- application of the model to the new data (fourth button),
- calculation of the seasonally adjusted data after recalculation (reestimation) of the model's coeffecients.

The tickable boxes correspond to the options available for storing the results of the calculations (the first two) and the parameters of the models (the third). You will note that their default position depends on the choice of process made by the two radio buttons. For an estimation of the models, it is the third box which is ticked; for a calculation of seasonal adjustment, it is the first two. There is little reason to change this preset choice.

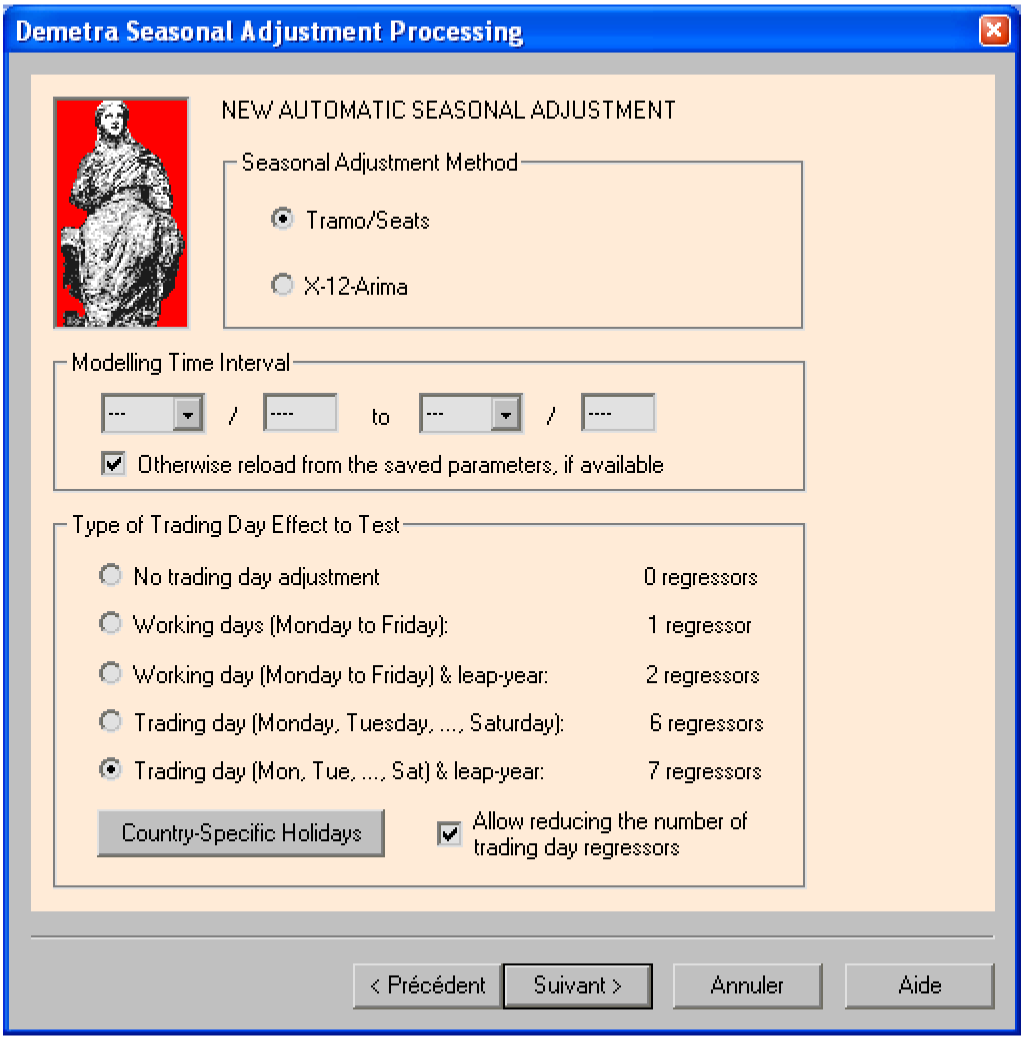
II.2. First steps

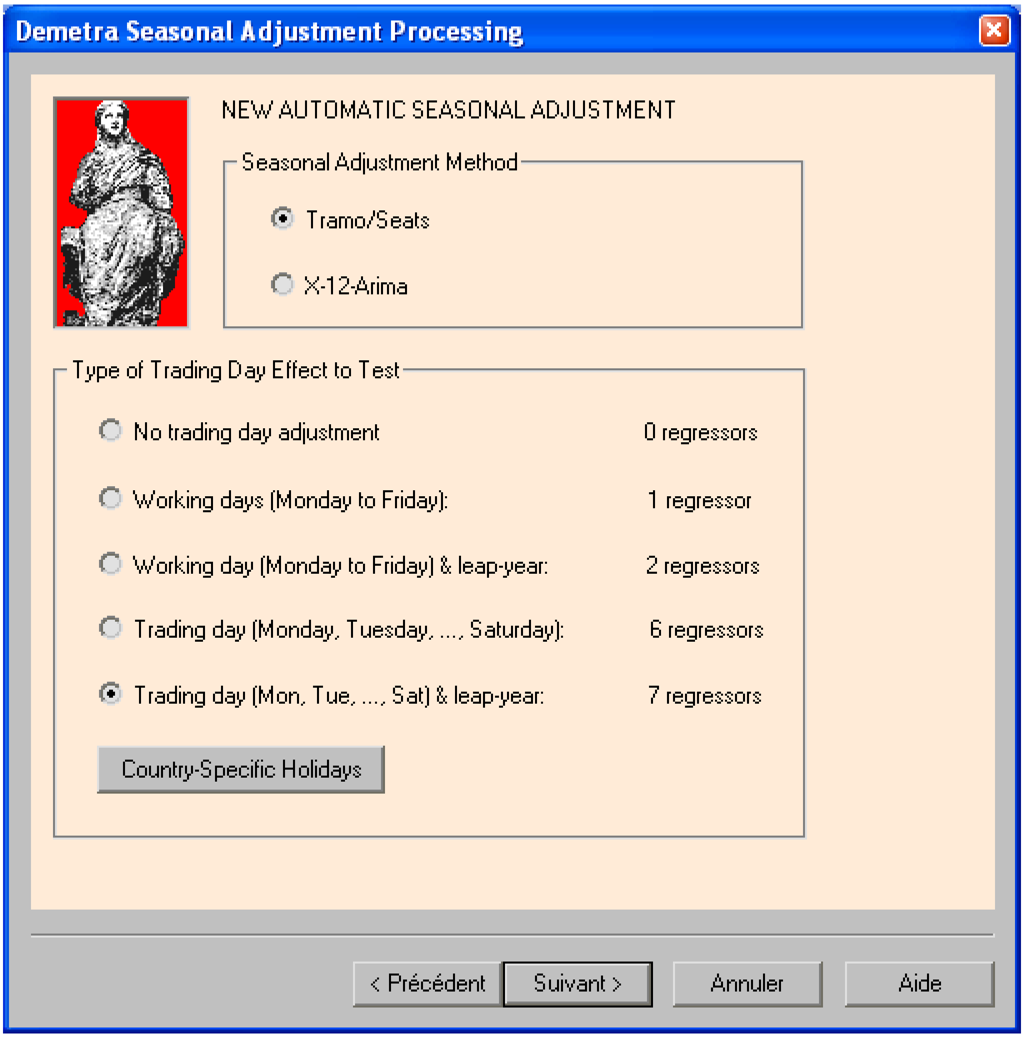
Let us accompany the user in the steps to take. We will suppose that it is the first time that they process the project's group of series. They have chosen therefore a process that implies searching for a pre-adjustment model. They now have to determine the method to use: Tramo-Seats or X12-Arima, then according to these choices, the different user parameters of these methods. The table below summarises the possible options according to the type of process selected.

	Process selected			
	Automatic		**Manual**	
User options	**Tramo/Seats**	**X12-Arima**	**Tramo/Seats**	**X12-Arima**
Calendar effects (CTD)	x	X	x	x
Holidays	x		x	
Series transformation			x	x
X11 parameter setting				x
Outliers			x	x
Correction for bias (SA)			x	x
ARIMA model choice			x	x

All the options listed correspond to a dialogue (or tab of a dialogue) excluding the fixing of the holidays which is one of the dialogue's options, 'calendar effect'. You will have noted that in the current version of Demetra it is not possible to fix the list of holidays for X12-Arima.

Should we desire the most automated process (at one click), we have two choices: the method and the correction of holidays.





II.2.1. Selecting the method

The question about which method to use is still open: Eurostat has made a number of comparisons between the two methods and has concluded that it was not possible to say whether one method is better than the other. Therefore, the choice essentially depends on the choices or habits of the user and of the organisation for which they work. From a purely pragmatic point of view, we can remark that from some tests covering numerous series, where only the default process is used (no user intervention), Tramo-Seats accepts a little more series than X12-Arima; on the other hand, series rejected by one method may be accepted by the other. Finally, in practical terms, the possibility of introducing holidays into the model is an additional facility of Tramo-Seats that, unfortunately, Demetra does not propose for X12-ARIMA.

To wrap up, note that the choice of the method implies an ‘ideological’ aspect: the debates between the modellers (the ARIMA models of Tramo-Seats) or followers of the empirical approach (X11 or X12) could be a lively one …

II.2.2. Options for trading days

The options proposed for correcting trading days are the following:

- 4 No correction.
- 4 Holidays (1 regressor). This option boils down to considering just two types of days: working days, from Monday to Friday and the days not worked, Saturday and Sunday. It counts, on the one hand, the days worked over the month and, on the other, the days of the weekend. This option would correspond to a correction option for the number of days worked in the strict sense of the term.
- 4 Estimation of weight for each day of the week (6 regressors). This is the most flexible option that allows the user to weight each day. It therefore seems to have been selected *a priori* but should not be retained if they think that there is no significant difference between the different days worked over the week, because in correcting, it would introduce random fluctuations.

With the correction tool activated, it is possible and, no doubt desirable, to also incorporate the correction for the leap year (choice of 2 or 7 regressors). This variant is advisable.

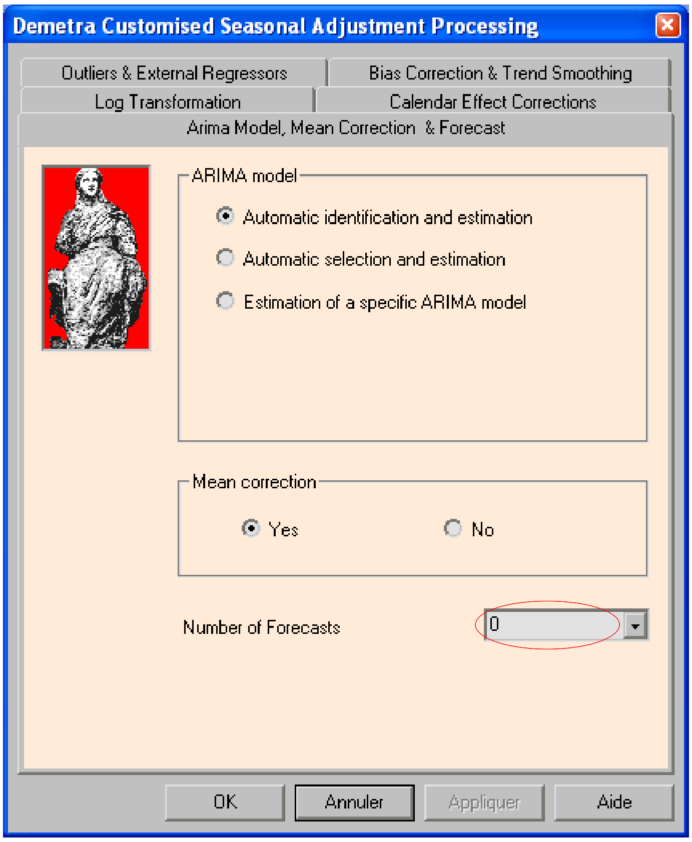
II.2.3. Exploring results by batch processing

Having determined the options, the last thing to do is to launch the processor. The result is then shown for the whole series in the top-left part of the window; the status of the model appears by way of an icon at the start of the line, and it is possible, using the tool bar, to filter what is shown on the screen to show only the series accepted, rejected or remaining to be processed. If time for studying the series is limited, we only need to retain the rejected series, trusting Demetra's default options.

An initial remark: in using Tramo-Seats, certain series are rejected, with an error message:

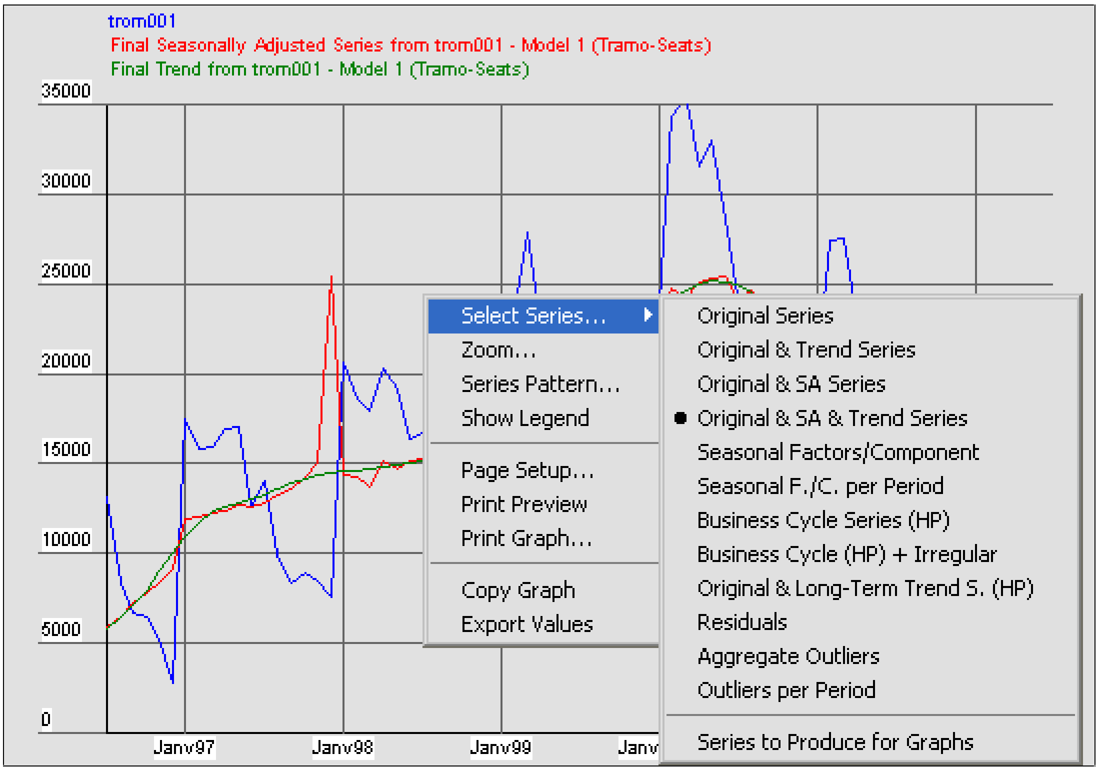
`Error returned from X12-Arima. Skipped! Error in the input parameters.`

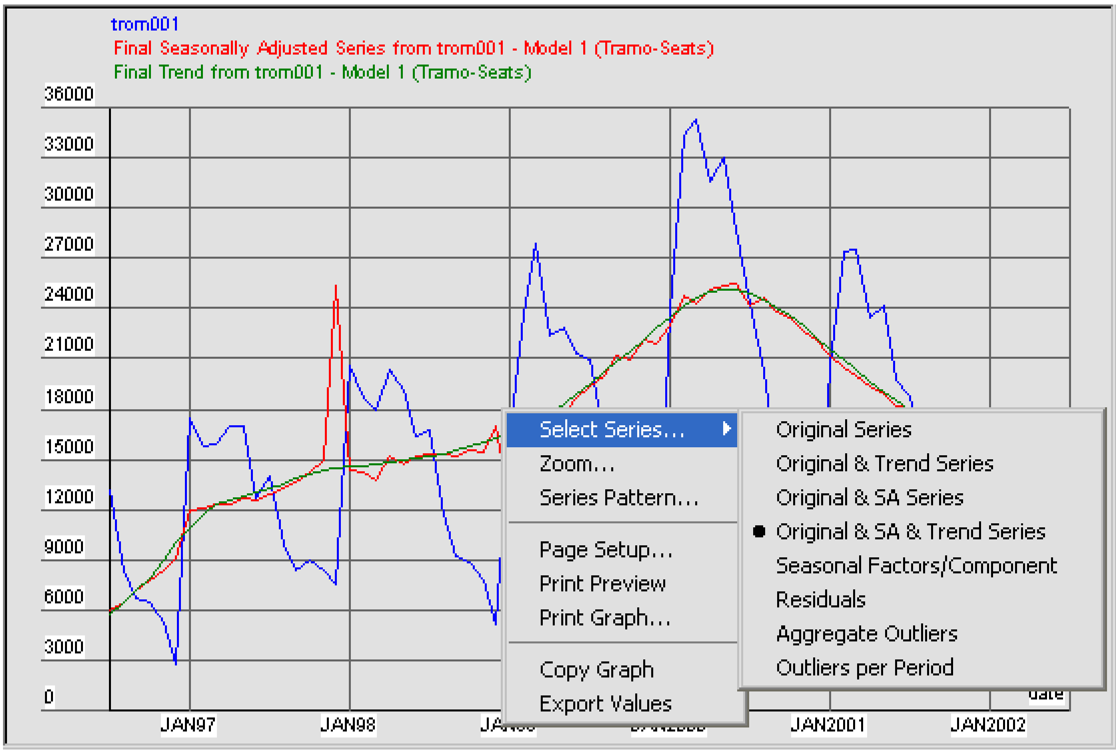
This message is generated during the search phase of an ARIMA model. By exploring the series as shown below, it is probable that the user will discover series that should not figure in the project, thus in different examples, we have met series whose value is constant and equal to 0. X12-Arima cannot process such series. For other series corresponding to real data to be processed, it would be desirable to use the method X12 with the ARIMA part. Unfortuneately, this is not possible, so we suggest then to try using the X12-Arima method on these series by preventing the use of forecasts (Setting to 0 of *Number of Forecasts* in the customised parameters dialogue).



Ø *Exploring the raw data series*

If one wishes to have an overall view of the results, and be able to see the difficulties encountered during the estimation of the model, it is useful to go through the whole series and view the main graphs in the bottom-left part. To do this, they should click on the right-hand button in this part and select the desired curve(s).

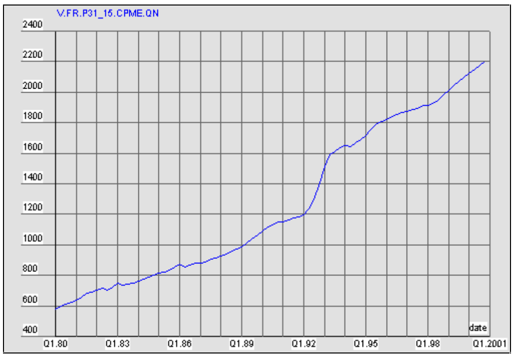


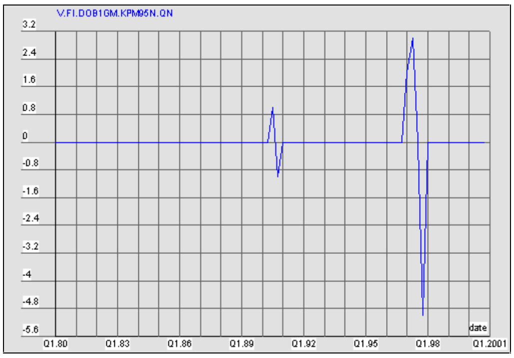


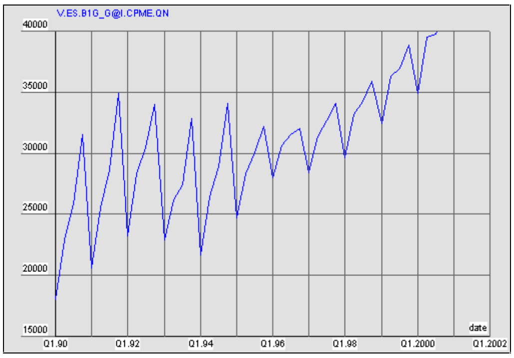
We then return to the list of series by clicking on the first. We can then scroll a list of the series using the up or down arrow keys.

One of the first choices consists in retaining all the three series: original, SA and trend. For example, in the figure shown above, the significant and punctual gap between the seasonally adjusted series and the trend allows the user to locate instantly the presence of an extreme value in December 1997.

Looking at the raw series graphs can provide the opportunity to discover information that could have escaped the cyclical analyst's wisdom: having received several hundreds of series, they have not necessarily had the time to look at them. Scrolling – even rapidly – the graphs authorised by Demetra's interface, risks showing series such as those given below (they are all from an actual process).



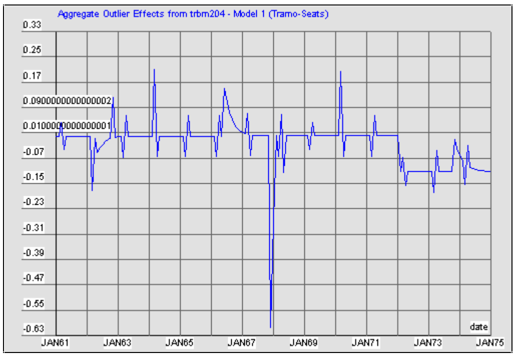


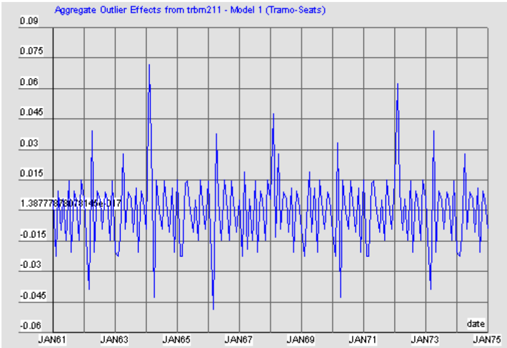


The first hypothetical case corresponds to a series presenting no seasonal characteristic. The second is a very particular series corresponding, probably, to adjustments or to the postponement from one period to the next. The third graph translates the presence of a change in operation mode: in the first quarter of 1995, there was a sudden change in the seasonal component; in this case, it is not easy to take account of this break and it is advised to split the series into two, and deseasonalise the two parts of the series separately. In practice, this means that it is necessary to delete from the project the series points corresponding to the first part.

Ø *Exploring events*

In the same manner, the (rapid) study of the seasonal component, of the residual and *a priori* correction terms (*Aggregate outliers*) is likely to provide useful information on the series studied. Below, are some examples of graph outputs of *a priori* corrections.



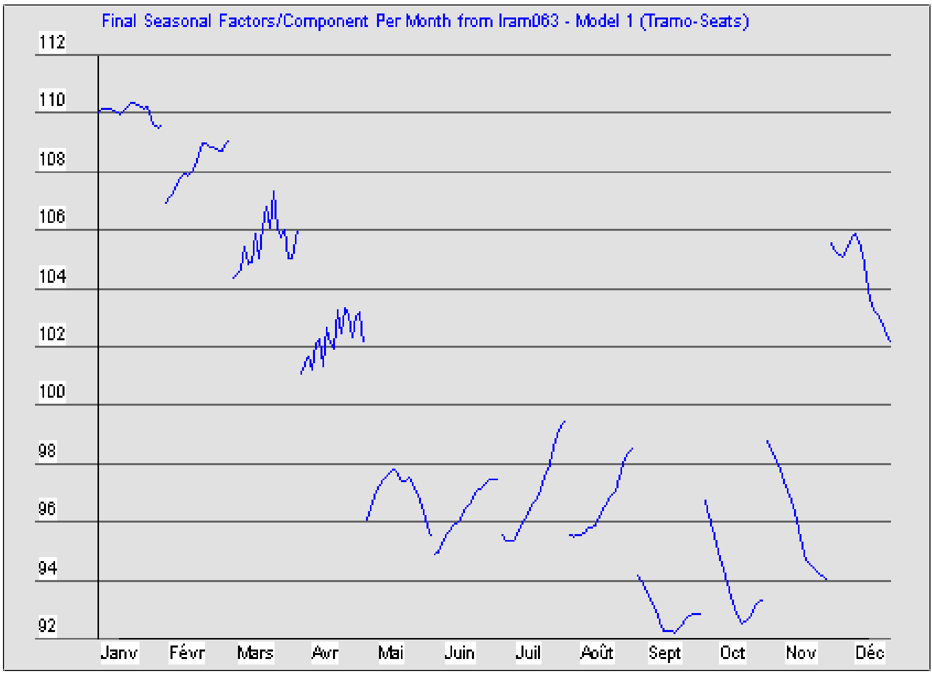


On the left, we can see some events (AO, TC, LS) and, annually, the repetition of a small swing corresponding to the effect of Easter. On the right, the almost random aspect translates the presence of an effect of trading or working days: there is, therefore, a correction term for each month, corresponding to its composition according to the different weekdays. We can identify the presence of extreme points (AO), but it is more difficult to locate that an Easter effect is also present. The aspect of this graphic depends, of course, on the relative importance of the different elements of prior correction. Finally, note that it is possible that the graph does not appear (message: **`The selected series is not available`**), meaning simply that no event has been detected. The current version of Demetra does not enable this to be shown even though other calendar effects exist (correction for trading days, for example).

Ideally, supposing that the cyclical analyst has the time, the different events need to be identified, ie. to establish the socio-economic reasons that provoked the variations in the values of the series.

Ø *Exploring seasonality*

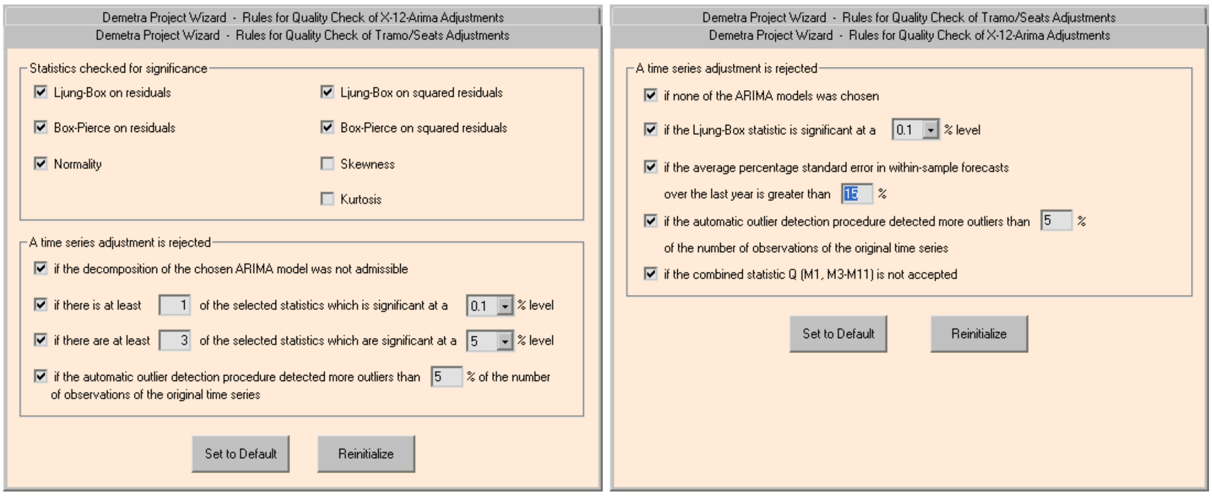
The option F/C per period enables the user to follow the evolution of seasonal coefficients over the whole period studied.



One can therefore judge the deformation of the seasonal structure over a long period. Thus, in the example above (jobseekers in Ireland from 1980 to 1997), one notes that the seasonal trough of summer (June, July and August) tends to diminish while the Autumn trough sunk deeper.

II.3. Improvement of rejected series

What should be done with the rejected series during the first process? First of all, let us immediately eliminate the rapid solution which would mean modifying the acceptance criteria (and therefore rejection) of a model. This can be done, as we have seen earlier, when we define a new project, but can be modified, afterwards, in the *Options/Rules for quality check* menu.



This choice – perhaps tempting – is unreasonable! It implies deliberately avoiding the difficulties met, ignoring the information indicating the nature of these difficulties and, finally, acting as if the models were satisfactory, which they are not.

Similarly, it is of course possible to force the acceptance of a series where the model is rejected. But, this option must only be used as a last resort. In most cases, it is possible to find the reason for the problem and conclude which process to apply, even if one of the possible responses is: not to deseasonalise, because the look of the series does not justify it!

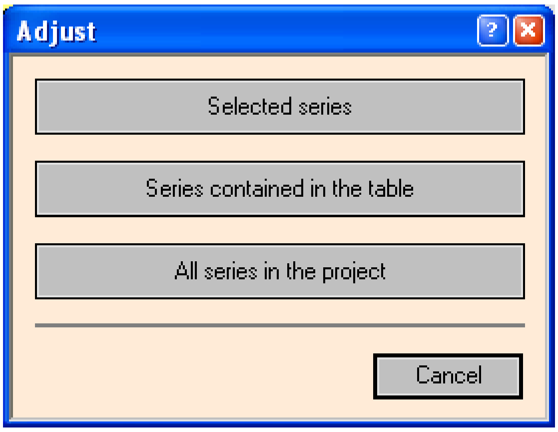
More seriously, to improve the rejected series, Demetra provides two options in its menus:

4 relaunch an adjustment (*Processing/Redo adjustment* menu),

4 or the improvement of rejected series (*Improvement of rejected adjustment/Start improvement of selected series* menu).

These two options are presented in Demetra's tool bar. They neighbour one another and are encircled in the image below.

The option for relaunching an adjustment allows the user to revisualise the initial parameter set-up dialogue for a seasonal adjustment model. Given that we have already made an automatic parcing, for this second phase we now need to opt for a customised adjustment (second dialogue option). This new seasonal model adjustment can be done for several series, according to the selection made in the following dialogue:



The adjustment will be applied:

4 to the series selected,

4 or to all the series shown in the table (taking account of the filters possibly applied on the display)

4 or to all the series of the project (displayed or not in the table). This option corresponds more to a reestimation of all the series rather than the search for our initial estimation.

We will use rather the first option, but after several manipulations that we will see a little further.

II.3.1. Improvement of rejected adjustments

In fact, Demetra proposes a processing method which is still automatic but with some variants regarding the initial standard approach of series, which would have been rejected in the first process. This can be done through the *improvement of rejected adjustments* option. This option is only possible if, among the selected series in the project table, at least one of them has been refused.

Choosing this option in the menu (or the corresponding tool in the tool bar) displays a new window. This one recapitulates in its five sections the information of the project window excluding the section containing the list of the project's series which is replaced by another graph. It is therefore possible to show several complementary graphs describing the series for processing: normally, and by default, the raw series, the trend and the SA series, in the top-left section and the seasonal component in the bottom-left section. Contrary to what happens in the project window, it is now possible to store information concerning several different models, providing that they are all of the same family as the initial model (Tramo-Seats or X12-Arima): in effect, each time that you test an improvement, a new model is added in the right-hand section of the window.

The selected series in the table of the project, the model of which has been rejected, are now going to be proposed to you, one by one. You can improve how they are modelled, with the programme assisting you in the search for a new model. After using these options, you will then be able to:

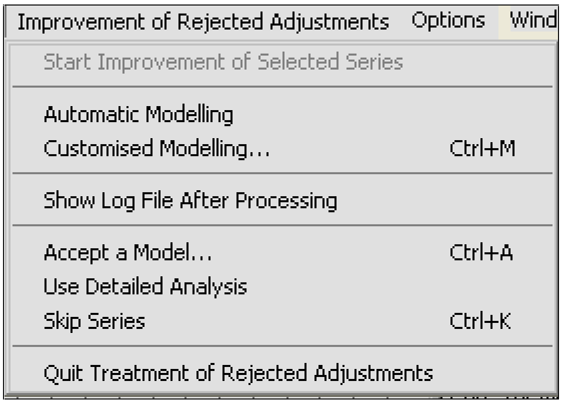
4 accept one of the models tested,

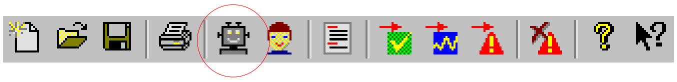
4 switch to the detailed analysis module, or

4 skip the series.

Note that switching to the detailed analysis module makes you leave the improvement mode of rejected series. The two other options make you go to the following series (if there are any left to process); the acceptance updates the project table and the file containing the modelling options. However, if you go to the series, its status (rejected) is not modified in the project.

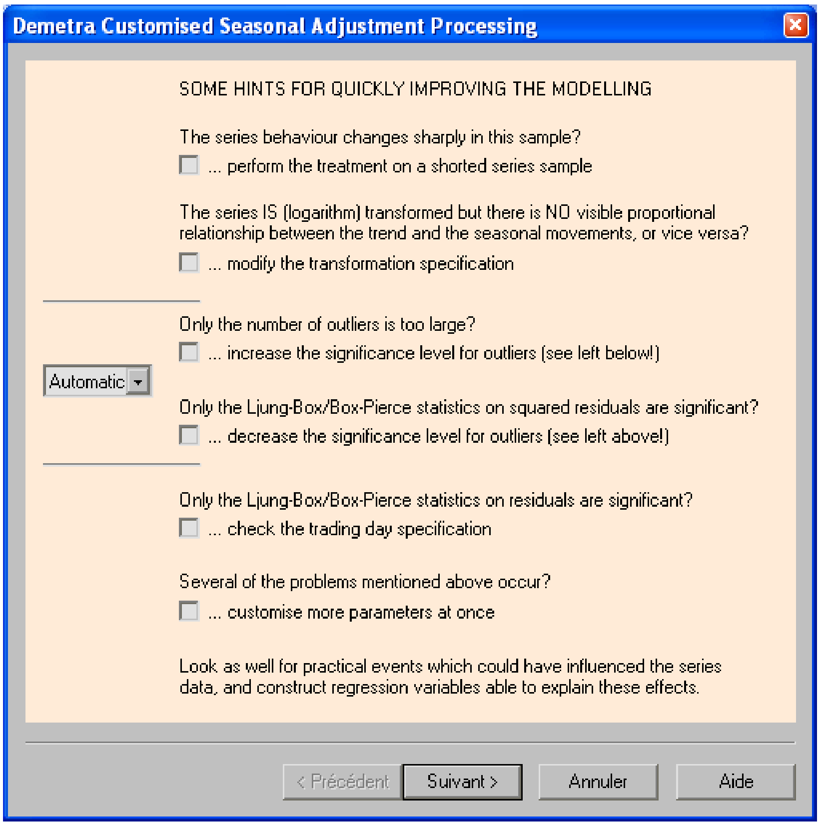
In the improvement assistant, a new menu appears in the menu bar, as well as a new tool bar the elements of which correspond exactly to that of the menu.

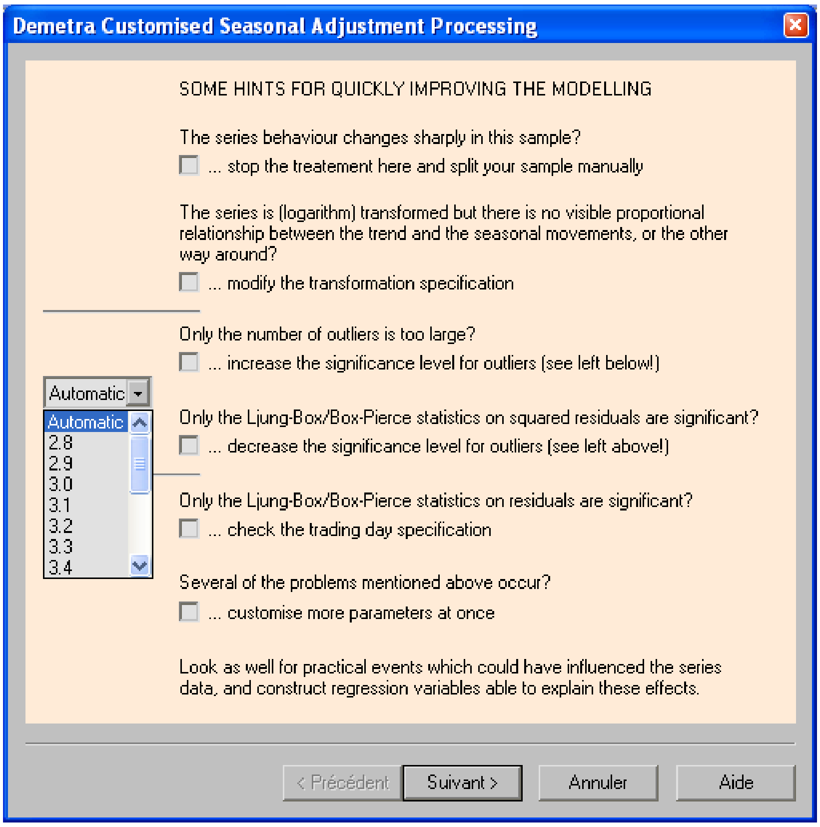




The automatic modelling tool (robot icon) relaunches the process, totally automatically. In fact, some of the model's options are modified, by responding to the questions which are asked in the dialogue box for customised modelling (icon of a human face) according to the values of the statistics obtained in the initial process result. This automatic improvement releases the cyclical analyst from asking themselves the questions that the programme can answer.

If you choose the customised processing tool, you will need to answer yourself the dialogue's questions which are as much hints.

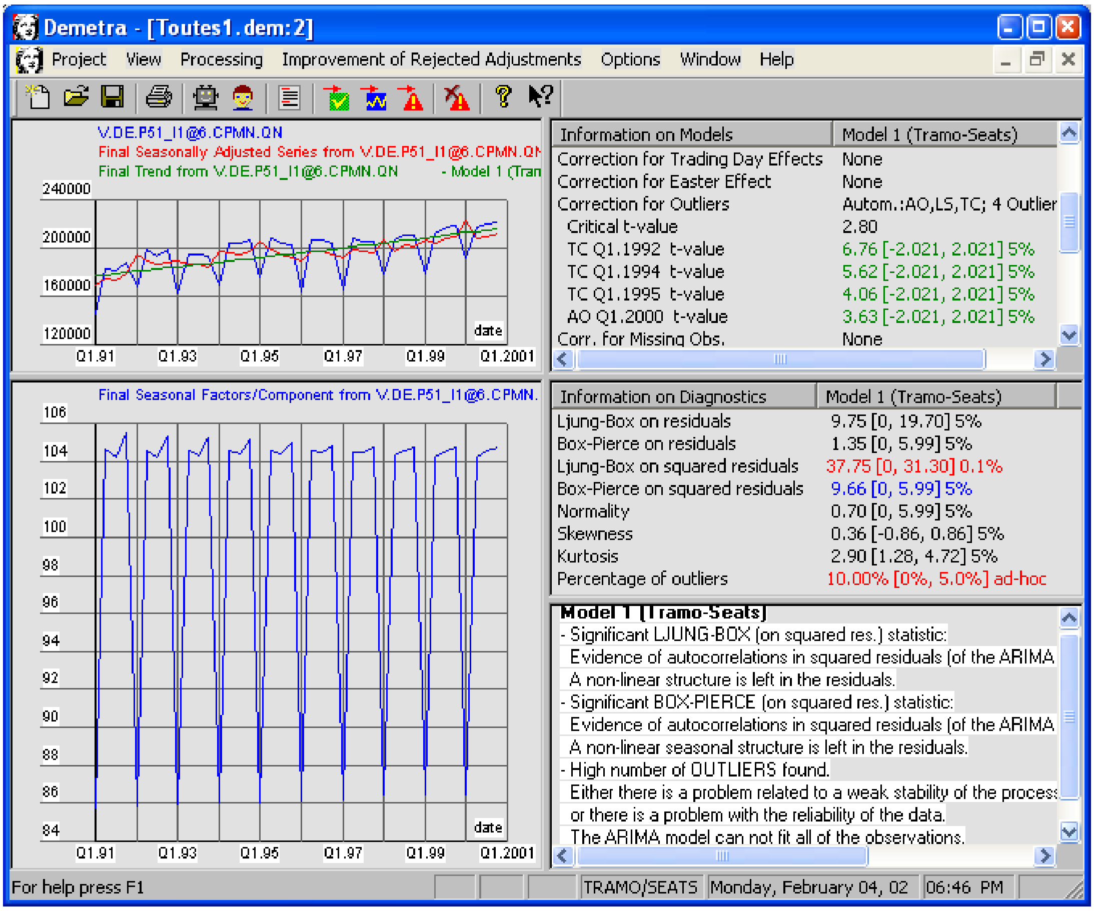

Demetra Customised Seasonal Adjustment Processing
SOME HINTS FOR QUICKLY IMPROVING THE MODELLING
The series behaviour changes sharply in this sample?
... perform the treatment on a shorted series sample
The series IS (logarithm) transformed but there is NO visible proportional relationship between the trend and the seasonal movements, or vice versa?
... modify the transformation specification
Only the number of outliers is too large?
... increase the significance level for outliers (see left below!)
Automatic
Only the Ljung-Box/Box-Pierce statistics on squared residuals are significant?
... decrease the significance level for outliers (see left above!)
Only the Ljung-Box/Box-Pierce statistics on residuals are significant?
... check the trading day specification
Several of the problems mentioned above occur?
... customise more parameters at once
Look as well for practical events which could have influenced the series data, and construct regression variables able to explain these effects.
< Précédent
Suivant >
Annuler
Aide


Demetra Customised Seasonal Adjustment Processing
SOME HINTS FOR QUICKLY IMPROVING THE MODELLING
The series behaviour changes sharply in this sample?
... stop the treatement here and split your sample manually
The series is (logarithm) transformed but there is no visible proportional relationship between the trend and the seasonal movements, or the other way around?
... modify the transformation specification
Only the number of outliers is too large?
... increase the significance level for outliers (see left below!)
Automatic
Automatic
2.8
2.9
3.0
3.1
3.2
3.3
3.4
Only the Ljung-Box/Box-Pierce statistics on squared residuals are significant?
... decrease the significance level for outliers (see left above!)
Only the Ljung-Box/Box-Pierce statistics on residuals are significant?
... check the trading day specification
Several of the problems mentioned above occur?
... customise more parameters at once
Look as well for practical events which could have influenced the series data, and construct regression variables able to explain these effects.
< Précédent
Suivant >
Annuler
Aide

Q.1 Does the series present a sudden break? The response to this question is obtained by looking at the graphs. We discussed this question earlier. The assistant will then propose to modify the modelling period.

Q.2 Has the series been transformed into logarithms (which supposes a multiplicative model, but the graphs do not conform with what one expects from such a model (see above for the presentation of pre-adjustment schemes? We must not then authorise data transformation.

Q.3 Does the series present too many outliers? This is a sign that the algorithms have trouble determining the trend; it is therefore recommended to modify the outlier detection threshold. For this, you can use the scroll menu on the left. This shows by default an 'automatic' selection and proposes values going from 2.8 to 4.2 by steps of 0.1. To reduce the number of outliers, this threshold must be raised. Analysing the statistics in the improvement window is helpful.



On the top-right, you will find the detected events (here, four events, or 10% of observations, value shown in red on the bottom-right) as well as a measure of the importance of their effect. The events are classified in descending order of importance (from 6.76 to 3.63). They have been detected using the automatic threshold selection, *Correction for outliers*, which in this specific case has given the value 2.80. All these numbers are dimensionless (no unit) expressed in standard deviation units from the residual component.

Q.4 Are the statistics on the residual squares (Ljung-Box and Box-Pierce) significant? This indicates that a correlation structure remains in the residuals which does not resemble therefore a white noise. One of the (numerous) ways to obtain an autocorrelation is the presence of extreme point(s) that would not have been detected as such. This explains the advice of lowering the detection threshold for extreme values.

Q.5 Are the same statistics on the residuals significant? Here, too, a structure remains in the residuals. It can result from, among others, the non-correction of the trading days or, to the contrary, a correction made by mistake.

If one of these questions, *and only one*, leads to a positive response, the cyclical analyst must use the options indicated before relaunching the process. For questions 2 and 5, the dialogue corresponding to the option selected follows our suggestion dialogue. For questions 3 and 4, one has to modify the scroll menu located directly on the left. Selection 1 interrupts the processing of the series in progress and goes to the next one.

Q.6 If, on the other hand, the response to more than one of the questions is positive, the suggestions proposed are no longer applicable, the user is therefore left to their own devices again. This choice triggers the customisation dialogue for processing that we will see a little later.

After these various attempts at improving the series, the user is therefore confronted with different competing models. They must therefore decide as to the status of the series in the project underway:

- accept one of the processes,
- transfer the series to the detailed analysis module to continue the analysis, being able to access all the parameters of these models,
- ignore the series, and thus leave it as 'rejected'.

Whatever the decision, the user will then go onto the next series, until all the series selected have been covered and processed by the improvement assistant.

Note that using the adjustment improvement assistant allows the specialised user to obtain the complete outputs of methods used and thus to have all the interpretation and decision tools that are edited for each of the two methods.

Log File for Processing 'Model 3 (Tramo-Seats)' of 'V.DE.P51_I1@6.CPMN.QN ' Page 1 / 25

```
 TIME SERIES REGRESSION MODELS WITH ARIMA ERRORS, MISSING VALUES AND OUTLIERS.
                                BETA VERSION (*)

                                      BY

                         VICTOR GOMEZ & AGUSTIN MARAVALL

                  with the programming assistance of G. CAPORELLO

(*) Copyright : V. GOMEZ, A. MARAVALL (1994,1996)

                          TRAMO compilation:  1999-01-28 23:24:15

 SERIES TITLE=V.DE.P51_I1@6.CPMN.QN


  SINCE LONGER FORECAST FUNCTION IS REQUIRED
  BY SEATS, NPRED CHANGED TO ( 8)

 ORIGINAL SERIES

 NUMBER OF OBSERVATIONS:  40

    X  10.003

 YEAR             1         2         3         4

 1991          145.700   182.660   181.290   188.330
 1992          167.520   198.370   193.990   198.620
 1993          163.020   195.140   194.070   192.990
 1994          169.980   204.610   203.450   207.160
 1995          176.260   208.290   203.810   202.210
 1996          163.530   206.680   206.570   203.710
 1997          165.490   207.770   205.930   205.380
 1998          178.690   206.490   211.170   210.540
 1999          179.280   212.140   217.060   218.560
 2000          192.380   217.430   219.490   221.370
DATES OF EASTER DURING THE REQUESTED TIME SPAN

 YEAR MONTH  DAY

   1991   MARCH   31
   1992   APRIL   19
   1993   APRIL   11
   1994   APRIL    3
   1995   APRIL   16
   1996   APRIL    7
   1997   MARCH   30
   1998   APRIL   12
   1999   APRIL    4
```

Output (log) Tramo-Seats (page 1 of 25)

```
                                         U. S. Department of Commerce, U. S. Census Bureau

                                          X-12-ARIMA monthly seasonal adjustment Method,
                                                       Release Version 0.2.8



                                   This method modifies the X-11 variant of Census Method II
                                   by J. Shiskin A.H. Young and J.C. Musgrave of February, 1967.
                                  and the X-11-ARIMA program based on the methodological research
                                   developed by Estela Bee Dagum, Chief of the Seasonal Adjustment
                                    and Time Series Staff of Statistics Canada, September, 1979.

 Primary Programmers: Brian Monsell, Mark Otto


Series Title- CE,M,C9I000V11B,VAL
Series Name- CE,M,C9I000V11B,
Mon Feb  4 22:28:02 2002

        -Period covered-  1st month,1972 to  5th month,2001
        -Type of run - additive seasonal adjustment

        -Sigma limits for graduating extreme values are 1.5 and 2.5 .
        -3x3 moving average used in section 1 of each iteration,
         3x5 moving average in section 2 of iterations B and C,
         moving average for final seasonal factors chosen by Global MSR.
        -Holiday adjustment factors applied directly to the final seasonally adjusted series
        -Spectral plots generated for selected series
        -Spectral plots generated for series starting in 1993.Jun
    CE,M,C9I000V11B,VAL                                                                  PAGE 1/28, SERIES CE,M,C
```

Output (log) X12-Arima (page 1 of 28)

With the adjustment improvement tool, we therefore have a means, more or less automatically, to continue processing the series to be deseasonalised. It provides a variety of tools to help the user in their decisions. However, it presents the disadvantage of obliging the cyclical analyst to decide on *each* series individually. We will now propose a working method where it is possible to continue benefiting from mass processing.

II.3.2. Improving models by batch processing

Our method is based on being able to filter series to show only the rejected series, but above all on the possibility of sorting the project series table according to whichever column of the table. To illustrate, a first example uses the last column of this table: the *Seasonality* column. By clicking on the heading of this column, we sort the table; it can contain three distinct values:

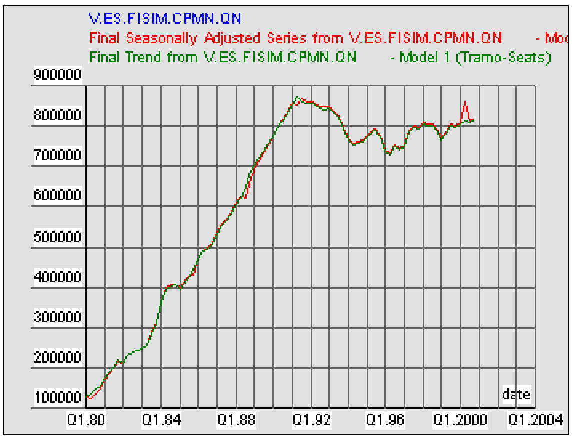
- for Tramo-Seats: *non-seasonal model used, seasonal model used, to be checked*
- or for X12-Arima: *not significant, probably present, significant*.

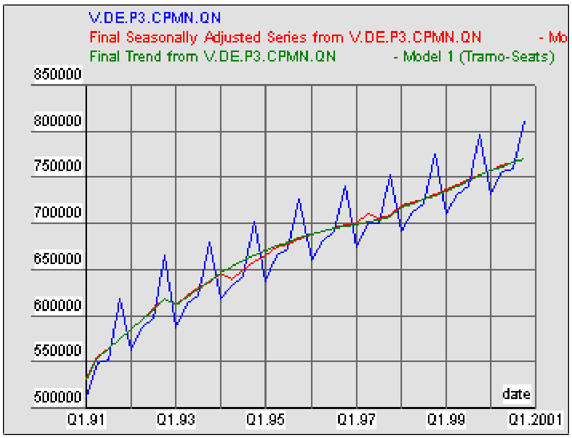
By exploring then the series graphically, it is easy to locate those which, in effect, do not need seasonal adjustment.

Ø *Too many outliers*

Let us return to our series to be improved and reuse several of the hints provided by the adjustment improvement assistant. Let us start by the too frequent presence of events or outliers. If we follow what was said above, it is easy to sort the series according to the *Percentage of outliers* column. The problematic series can therefore be found at the end of the table.

Let us remember again that graphically exploring the series thus grouped together can prove worth the effort. Here are, for example, two series presenting respectively 13% et 15% of outliers and thus rejected by Demetra.





In the first case, the series is not a seasonal one. In the second, the series is extremely regular, almost periodical. The model describes it very well; the residual component is very small and thus its standard deviation equally small. Any point deviating a little from the model finds itself in fact at a significant distance when one evaluates it as a standard deviation unit.

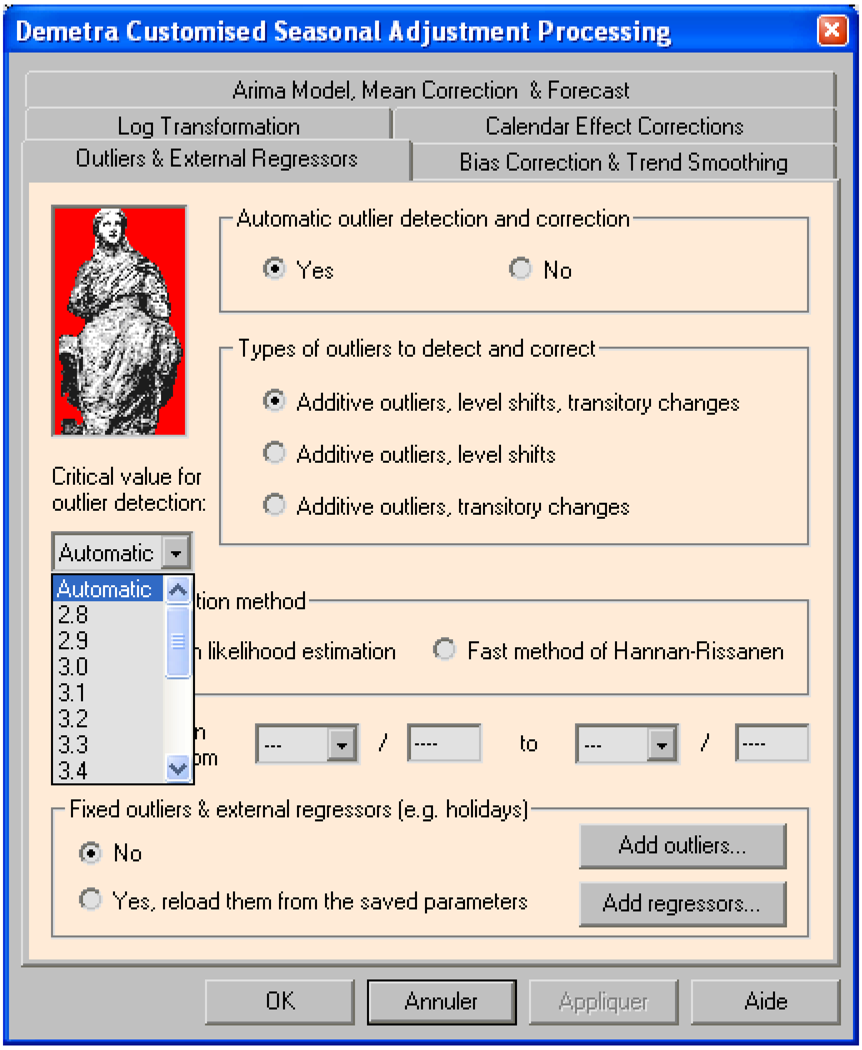
Having sorted our series according to the number of *outliers*, we can select them and effect, globally, a personalised process adapted to these series, that it is to say by modifying the detection and the processing of *outliers* on the model.

In the customised processing dialogue, we select the corresponding tab where we can manipulate the options concerning the events. Let us process them in a different order to that in the dialogue. First of all, we find the scroll menu for fixing the detection threshold of outliers; it is the same as the one we met with the improvement assistant. It functions in the same way: increasing the threshold for reducing the number of events detected. We can thus progressively increase the threshold by setting it at 3.2, then 3.7 and 4.2.

Then it is possible to limit the type of events to be detected automatically. In particular, the option without the transitory changes (TC) can be interesting to test when there are many of them.

We can also modify the detection method. The rapid method, as its name suggests, should only be used for cases where the calculations must absolutely be improved.

It is possible to limit the period in which Demetra looks for events. Note that in a batch process, this restriction will apply to all series of the batch.



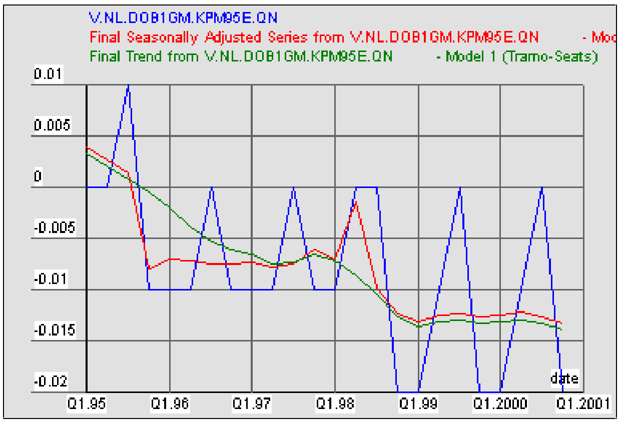
Finally, in serious cases, and it is the first option of the dialogue, it is possible to completely disactivate the detection and the processing of outliers. This choice is to be used with caution. Indeed, if it makes it possible to completely eliminate the outliers and thus respect the quality criteria, it implies that the possible extreme points have an influence in determining the tendency. In this case, it is therefore useful to specify the extreme points oneself, or to introduce an external regressor (last zone of the dialogue).

By progressively modifying the options in the order shown above and by reexecuting the adjustment each time on the subset of rejected series because of an excessive number of events, there are strong chances of gradually increasing the number of models accepted.

Ø *Ljung-Box statistics on squared residuals*

Let us continue in the logic of the improvement assistant and consider the series for which the Ljung-Box statistic on residual squares is too high (significant). But beware, this statistic is only calculated by the Tramo-Seats method; X12-Arima does not produce it.

We sort therefore our series table according to this criterion; we explore the series graphically where the statistic appears in red (because of rejecting the adjustment). This operation often allows the user to highlight particular series.



Let us select now the series where *only* the Ljung-Box statistic for the residual squares is significant, in line with the hints of the improvement assistant and restart the customised adjustment by lowering the detection threshold of outliers, for example by fixing it at the minimum value of 2.8.

Ø ***Ljung-Box statistic on residuals***

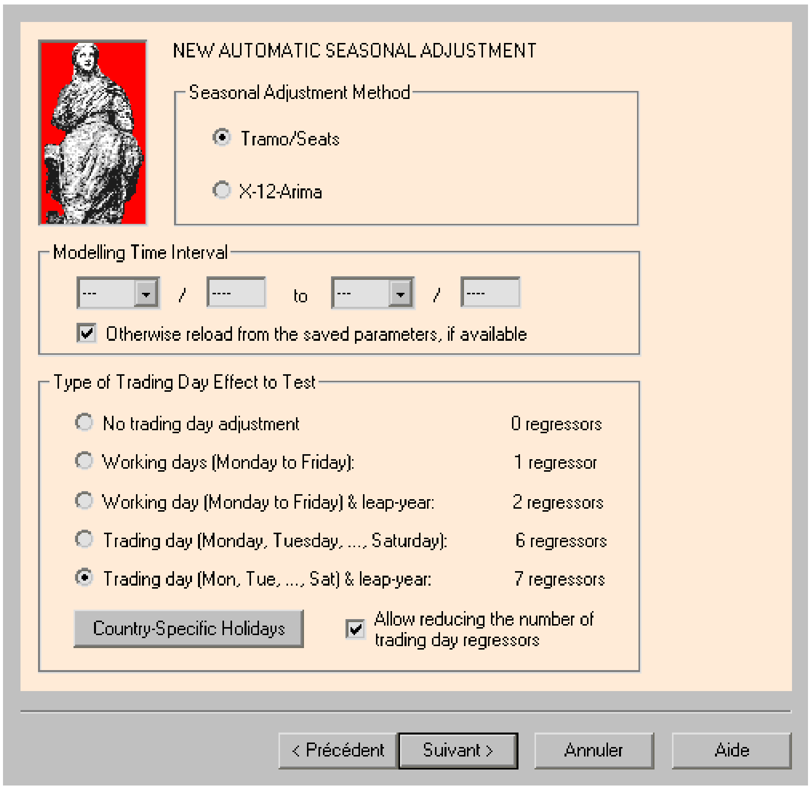
To finish in the logic of the improvement assistant, our last task is to process the cases where only the Ljung-Box statistic is significant. We should not forget that this means that the information subsists in the residuals. Among the very many possibilities, the assistant suggests modifying the parameters for correcting trading days.

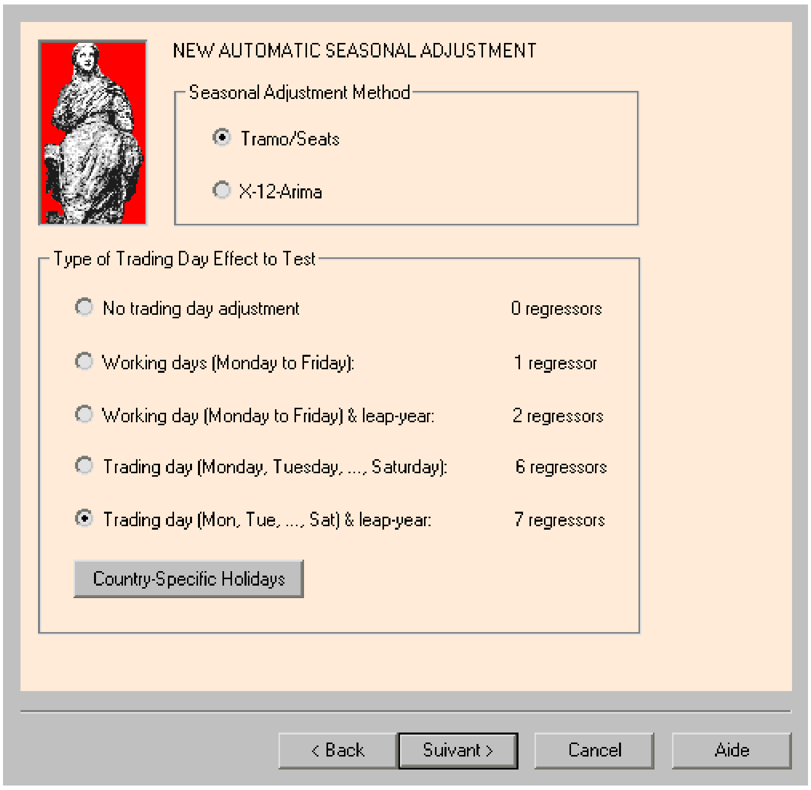
To take this suggestion into account in a batch process, the following is required:

4 filter the presentation to retain only the rejected series,

4 sort the table according to the Ljung-Box statistic column,

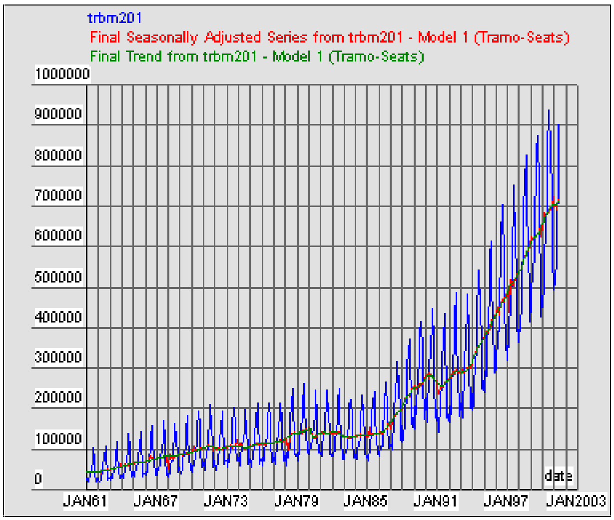
4 select among the series in the bottom of the table, the statistics shown in red, those which share the same correction option for trading days (CTD). If it is necessary to perform a discontinuous selection (Ctrl-click), a new batch process (*default parameter for a new processing*) can be started by choosing in the following screen another CTD.


NEW AUTOMATIC SEASONAL ADJUSTMENT
Seasonal Adjustment Method
Tramo/Seats
X-12-Arima
Modelling Time Interval
to
Otherwise reload from the saved parameters, if available
Type of Trading Day Effect to Test
No trading day adjustment
0 regressors
Working days (Monday to Friday):
1 regressor
Working day (Monday to Friday) & leap-year:
2 regressors
Trading day (Monday, Tuesday, ..., Saturday):
6 regressors
Trading day (Mon, Tue, ..., Sat) & leap-year:
7 regressors
Country-Specific Holidays
Allow reducing the number of trading day regressors
< Précédent
Suivant >
Annuler
Aide


NEW AUTOMATIC SEASONAL ADJUSTMENT
Seasonal Adjustment Method
Tramo/Seats
X-12-Arima
Type of Trading Day Effect to Test
No trading day adjustment
0 regressors
Working days (Monday to Friday):
1 regressor
Working day (Monday to Friday) & leap-year:
2 regressors
Trading day (Monday, Tuesday, ..., Saturday):
6 regressors
Trading day (Mon, Tue, ..., Sat) & leap-year:
7 regressors
Country-Specific Holidays
< Back
Suivant >
Cancel
Aide

With rapidity in mind, it is also possible to limit oneself to the two extreme choices: no CTD for an adjustment with 7 regressors.

Nevertheless, in the examples shown, it appears that these tests do not allow us to significantly increase the number of series where modelling is accepted. Here again, using a graph shows that among those series where the statistics on residuals are significant, there are often some with breaks: either a trend break or a seasonality break.



Information on Diagnostics	Model 1 (Tramo-Seats)
Ljung-Box on residuals	51.33 [0, 46.80] 0.1%
Box-Pierce on residuals	0.67 [0, 5.99] 5%
Ljung-Box on squared residuals	22.47 [0, 32.70] 5%
Box-Pierce on squared residuals	3.56 [0, 5.99] 5%
Normality	0.81 [0, 5.99] 5%
Skewness	-0.02 [-0.22, 0.22] 5%
Kurtosis	3.20 [2.55, 3.45] 5%
Percentage of outliers	3.09% [0%, 5.0%] ad-hoc

Model 1 (Tramo-Seats)
- Significant LJUNG-BOX (on res.) statistic:
Evidence of autocorrelations in residuals (of the ARIMA model fitting).
A linear structure is left in the residuals.
-> Model failed some diagnostic tests.

Thus, in the example above, we can clearly see a trend break towards 1986. We can remark too that the series under analysis is particularly long (40 years) for a seasonal adjustment. By cutting it down to start in January 1986 (still 20 years), the series is accepted by the default process of Tramo/Seats.

Ø *Other systematic paths for the project table*

To conclude this part which supposes the processing of (very) many series, we can keep in mind that it can be instructive to go through the table by visualising the series graphically and putting the items into order according to the different columns of statistical results. We remind readers of the list in the table below, as well as the methods which generate them.

We have looked at statistics on residuals. We can also use statistics on **the distribution** of residuals (normality, skewness, kurtosis). The values of these statistics are high when the distribution of residuals deviates from the ideal distribution given by modelling (normal law). By exploring them, we can note the possible disturbances, breaks or structural change for which these values are high.

List of statistics in the project table

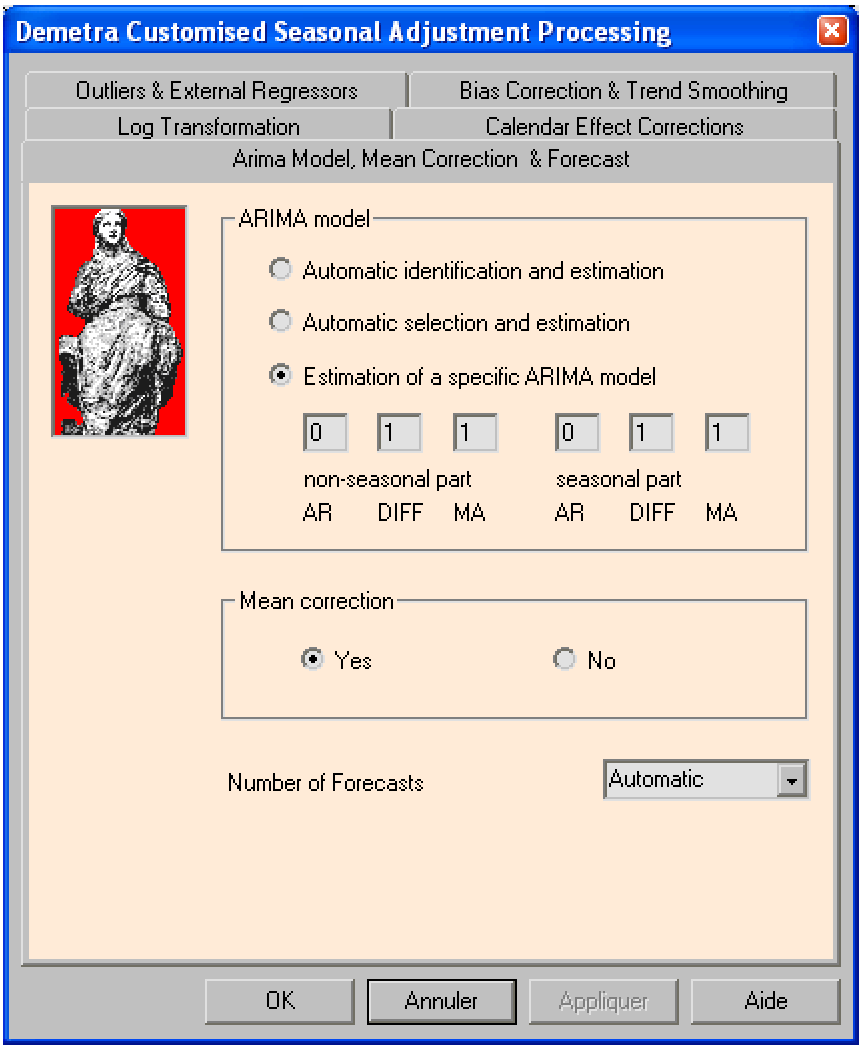
	Tramo-Seats	X12-Arima
Number	x	x
Name	x	x
Status of adjustment	x	x
Time span (n° of obs.)	x	x
Last time saved results	x	x
Arima model	x	x
Ljung-Box on residuals	x	x
Ljung-Box on squared residuals	x	
Box-Pierce on residuals	x	
Box-Pierce on squared residuals	x	
Normality	x	
Skewness	x	
Kurtosis	x	x
Forecast error		x
Percentage of outliers	x	x
Combined statistic Q		x
Transformation	x	x
Mean correction	x	x
Trading day effect	x	x
Easter effect	x	x
Outliers	x	x
Missing observations	x	x
Other regression effects	x	x
ARIMA pre-adjustment	x	
X-11 pre-adjustment		x
X-11 seasonal filter		x
X-11 trend filter		x
Seasonality	x	x

The **ARIMA model** column makes it possible to sort series according to the order of the ARIMA models selected. When the Tramo-Seats method is used, it is worth focussing on the models:

4 of (0 0 0) type for the non-seasonal part (first triplet),

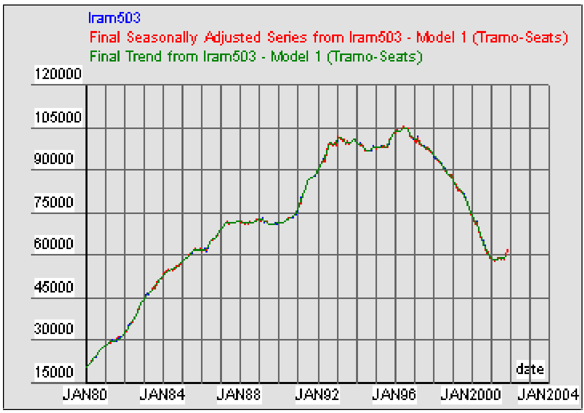
4 where, to the contrary, the order is high, which indicates that the algorithm has had difficulty in determining the trend.

For series whose models have been rejected, we can make a last attempt for ARIMA modelling by imposing the most classical model called AIRLINE. This is the one Demetra proposes by default when you select the option *Estimation of a specific ARIMA model*. Its order is (0 1 1)(0 1 1) and is reputed to describe well many current time series.

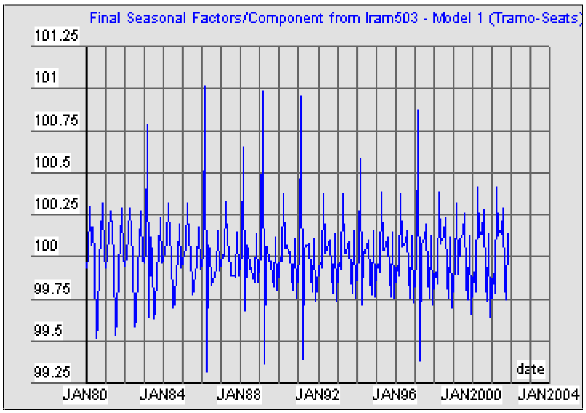


Finally, the last column in the table (**seasonality**) makes it possible to group those series which, according to Demetra, do not present any seasonal component. We can check that this is the case by looking ourselves. Inversely, certain series processed by Demetra as seasonal may not present any very apparent seasonal component.

Raw series, trend and SA



Seasonal component



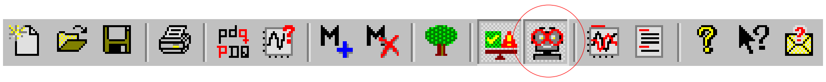
In the example above, the seasonal component (multiplicative) is of the order ± 1%.

II.3.3. Expert system

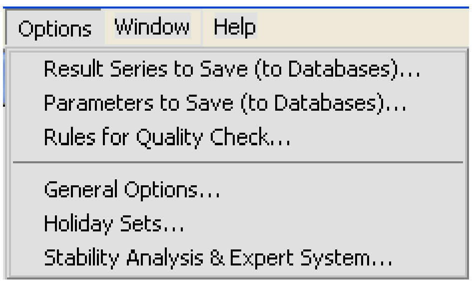
The Demetra version of Mars 2002 (version 2.0, Service Pack 1) has introduced an expert system to automate the improvement of rejected adjustments. When a series is rejected and the option is activated, the expert system will reexecute a whole adjustment series by modifying, one by one, the different parameters of the model. The user guide lists the options which are modified.

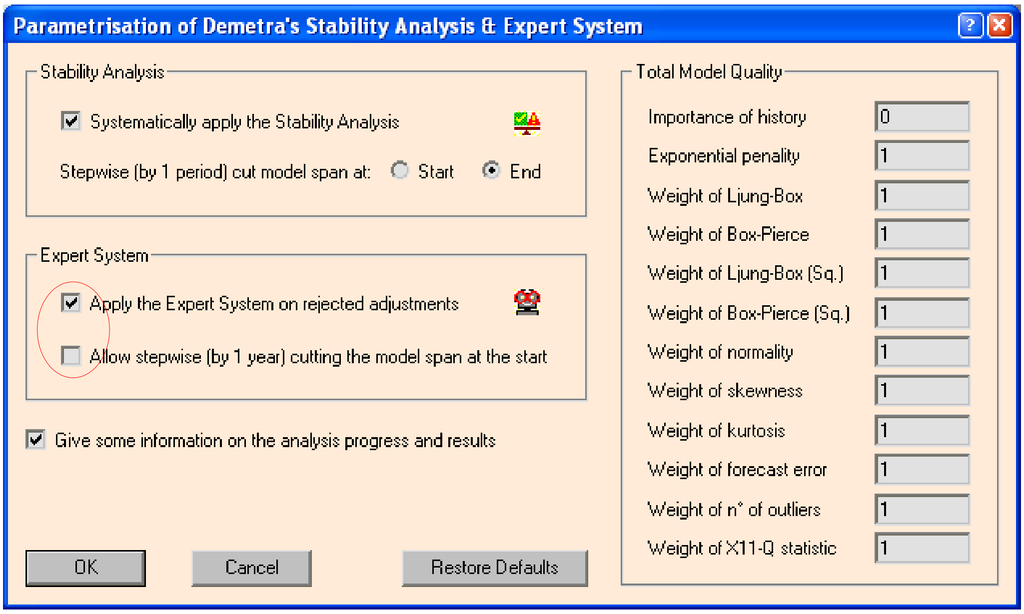
It is also possible to authorise Demetra to cut the series by steps of one year, after a minimum four-year period of analysis.

The expert system can be activated by clicking on the corresponding icon in the tool bar (a red robot with glasses sitting on a pair of scales.



or by ticking the box in the parameters dialogue of the expert system. The automatic reduction can also be activated in this dialogue.



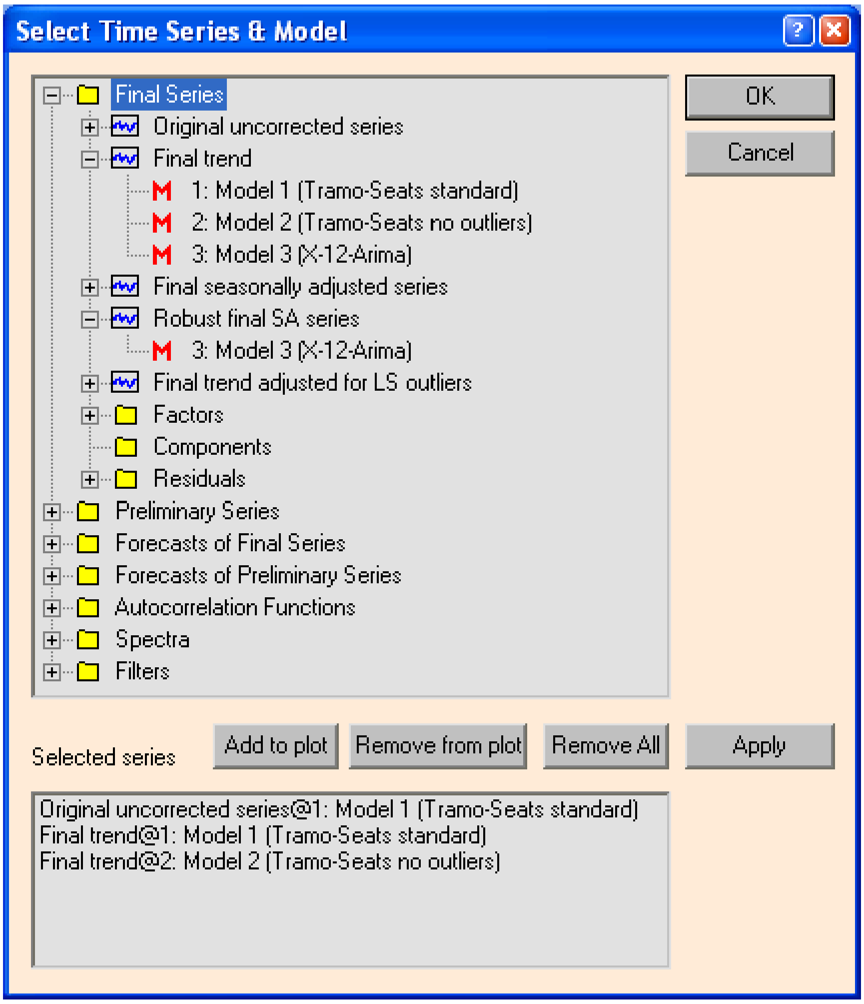


II.3.4. Processing in the detailed analysis module

If you have the necessary time available, going into the detailed analysis module can be worthwhile. It allows the user to stock the parameters and the results of different model tests on the same series. In particular, it enables the user to test the two main families of methods (Tramo-Seats and X12-ARIMA), which the improvement assistant does not do for rejected series. It gives access to all the parameters of these two methods: the options already accessible in the automatic process appear again, but within a totally different interface, and other even more technical parameters can be accessed.

Finally, and it is maybe one of its chief interests, the detailed analysis module proposes an extremely practical graphic comparison tool. The menu is activated only when at least one model has been made in the detailed analysis module. The graphic comparison tool shows a window in four sections, in each of which it is possible to define a graph composed of one or several series. The series represented in the same section must be compatible together (same scale).

The dialogue for defining a graph appears as shown below. Each file icon contains one or two series if the folder containing it shows a plus sign. The series is represented by a small blue line chart. By double clicking on this icon, it will activate the curves corresponding to the different models tested in the detailed analysis module.

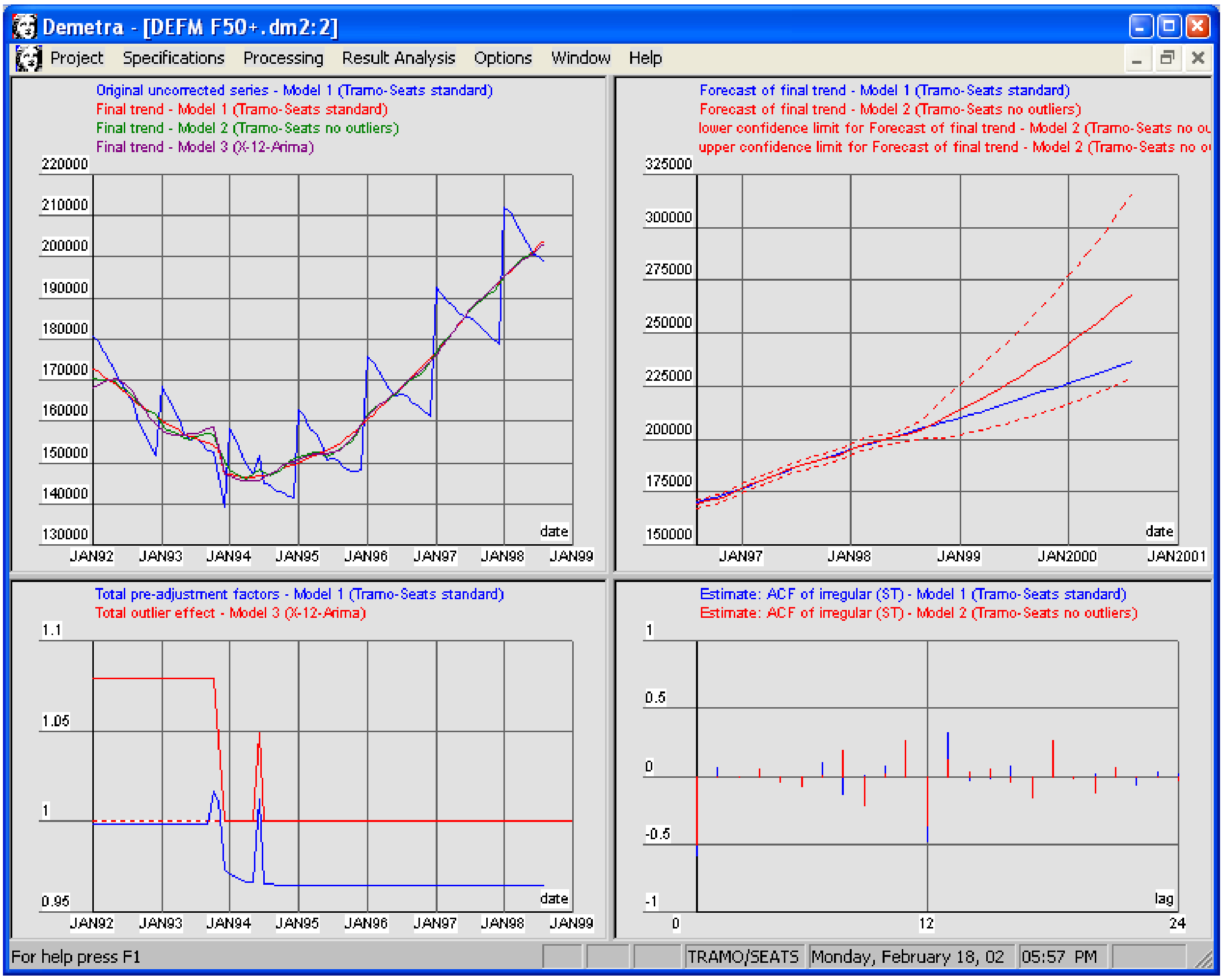


Note that Tramo-Seats and X12-ARIMA do not exactly produce the same groups of series. Not every model may be listed in the choices possible for representing one of the types of series.

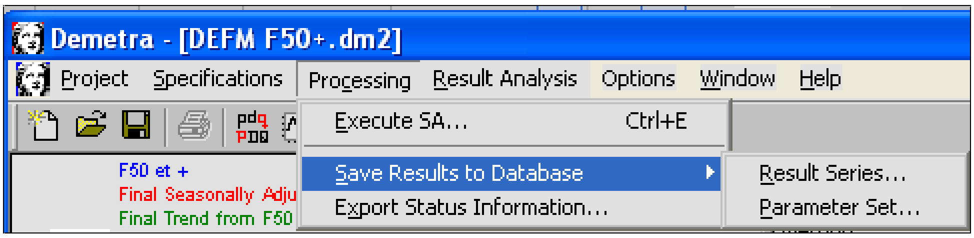
In the example below, on the top left, we have represented the original series as well as the trends estimated by three different models, and on the right, the trend forecasts provided by the two Tramo-Seats models.

On the bottom left, we can compare the events detected by the two models where we have activated the automatic detection of outliers. We can remark that Tramo-Seats and X12-ARIMA do not run their adjustment in the same way, but detect the same events excluding those of type TC, the search of which is not foreseen in the X12-ARIMA algorithm.

Finally, on the bottom right, we find the representation of the autocorrelation function of the residuals of Tramo-Seats models. This graph is one of the key tools for those wishing to perfect the ARIMA modelling.



Before finishing, let us take note that it is possible to transfer the results of one of the models analysed in the detailed module to the project containing the series thus examined. This also holds for the parameters selected in this model.



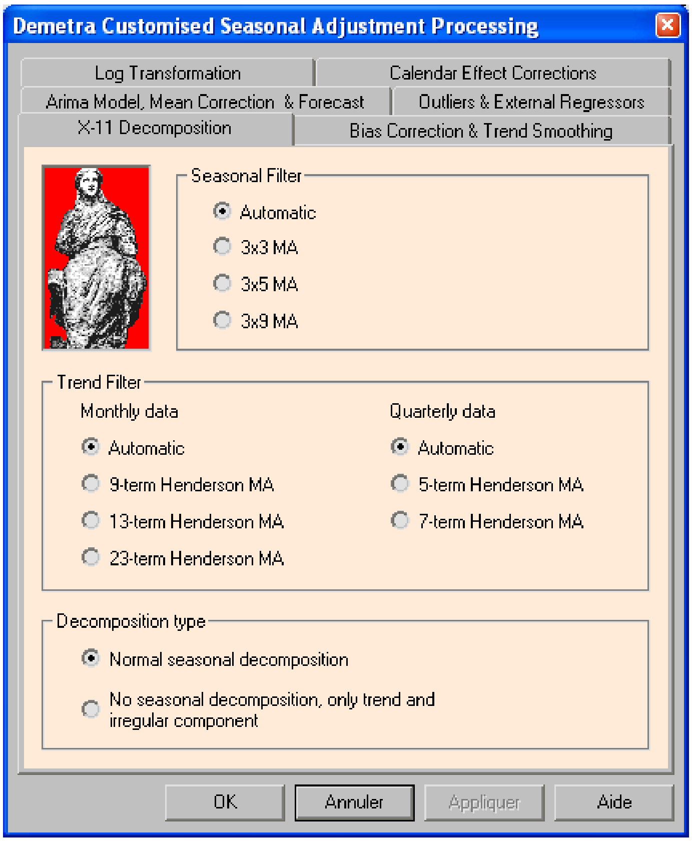
This enables the user to store the information of the model selected using the global detailed analysis module in the database from which the series originates. Thus, in Excel, the tabs of the folder containing the source will be updated or created to keep the components produced by the model, as defined in the storage options.

II.4. Specific characteristics of X12-Arima

II.4.1. Parameterisation of X12-Arima

If you have chosen the X12-Arima method, you will obtain a dialogue for customising the adjustment that is largely similar to that for the Tramo-Seats method. Some differences should nevertheless be noted: some of the options in Tramo-Seats are not available. Note, for example, that it is impossible to introduce an intervention variable made up of a sequence of ones (*Add outliers* button in the tab *Outliers & External regressors*) or still, to limit choice for adjusting the bias of the SA series.

However, a supplementary tab (*X11 Pre-adjustment*) appears in the customisation dialogue and concerns essentially selecting the orders of different moving averages on which the X12 is based.



Ø *Trend filter*

You can choose the order, ie. the number of terms used for estimating the trend. This concerns the so-called Henderson moving averages that present good smoothing properties. The higher the number of terms (9, 13, 23 months or 5 or 7 quarters), the more the trend will be smooth. By default, X12 automatically selects the order according to the relative share of the irregular component in the variations of the series.

Ø *Seasonal filter*

This concerns the order of moving averages estimating the seasonal coefficients using deviations from the trend. This parameter controls the possibilities of evolution, over time, for seasonal coefficients. Its smallest value, 3x3, an average of deviations from the trend using the same month of 5 consecutive years to estimate the seasonal coefficient of the month of the median year, corresponds to greater flexibility for the coefficients. However, 3x5 (7 years) and 3x9 (11 years) impose a slower evolution of coefficients, ie. limit the possibilities of deforming the seasonality over the whole period.

Of course, the higher the order of this moving average, the more serious the problem of series extremities and therefore the necessity of having a 'good' Arima model for completing the extremities.

By default, X12 selects the option that it considers the most suitable.

Ø *Pre-adjustment type*

The choice is given to disactivate seasonal adjustment by X12: it must be used on non-seasonal or already deseasonalised series. It allows the user in this case to obtain an estimation of a trend as well as diverse statistics calculated by X12.

II.4.2. Exploring results in the case of X12-Arima

If you use X12-Arima, the list of statistics available in the project table differs from those given by Tramo-Seats. Fewer statistics are calculated on the residuals. Similarly, the list of Arima models tested is more limited; in this column, the message *none of the models were chosen* can appear, when none of the possible models were selected. This situation leads to the model's rejection; this is one of the default options of the adjustment quality control.

However, in the columns completed by X12, a *forecast error* column appears. This concerns, in fact, backcasts: X12 uses the Arima model selected to generate forecasts of the values observed during the last year using data from previous years. The deviation is measured as a percentage; if it is too significant, the method will not use extrapolation of the series by the Arima model to complete the data. This information is highlighted in the *X11-pre-adjustment* column by the message *without ARIMA forecasts*.

The *X-11 seasonal filter* and *X-11 trend filter* columns indicate the order of moving averages retained (see previous paragraph).

Finally, the *combined statistic Q* is an indicator combining various particular statistics linked essentially to the stability of the model's different components over the period under analysis. Its value is between 0 and 3 (dimension-less) and should not exceed 1.

II.5. Updating and correction strategy

In the previous sections, we have described the use of Demetra for the initial setting up of seasonal adjustment process. The approach presented, essentially based on the automatic analysis tool and its capacity to process by batch has enabled us to judge all the series in the project. If the cyclical analyst has not modified any of the option boxes in the dialogue for launching processing, the definition of the models retained has been stored in the project file; this includes the choice of the method applied (Tramo-Seats or X12-Arima), the specification of preparatory processing tasks (outliers, data transformation …), the identification of Arima models used (the orders of the different parts) and the estimation of the coefficients of these models. Finally, using this model has produced the different series resulting from seasonal pre-adjustment (trend, SA, irregular, etc.).

Afterwards, new data are going to be periodically generated, for example on a monthly basis. The cyclical analyst must then provide seasonally adjusted data and, possibly, propose a commentary on the recent evolution. They must therefore deseasonalise again the series they have, completed by points recently acquired. Do they have to go through the initial steps again?

The reply to this question is not simple. It is directly linked to how the user organises their work, resulting from, on the one hand, personal choice, but also from their environment, above all the choices of their institution. Moreover, the simple application of the same model to slightly different data (yet another observation!) can lead to a modification of the past component values. Thus, taking an example using X12-Arima, the appearance of an additional point modifies (slightly) the predictions of the observations used for estimating the trend; and this, even after having changed nothing in the model's specification. The trend being slightly different, the accounting equation of the pre-adjustment model implies that the totality of the other components can also vary.

Thus, the arrival of data at date *t* makes it possible to calculate the SA value at date *t*, but it can also lead to the SA data needing adjustment concerning dates *t - 1*, *t - 2* …

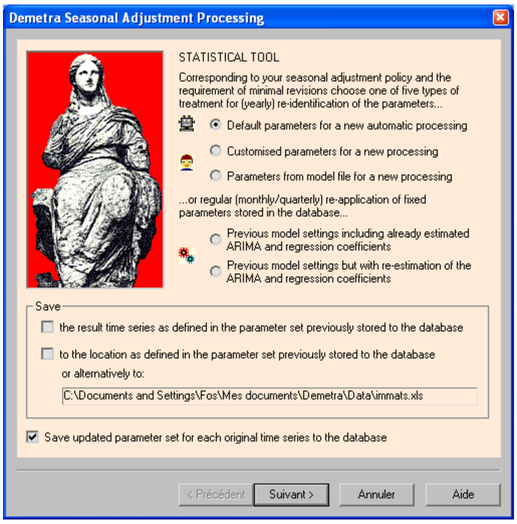
The scale of possible adjustments depends, of course, on the modifications made to the model (reestimation of coefficients, model reidentification, modification of preparatory processing, even change of method!). They are always present, even in the absence of any modification.

In the organisation of deseasonalisation tasks, it can be useful to fix a deadline, beyond which the published series will no longer be retrospectively corrected.

The following table summarises how seasonal adjustment tasks could possibly be organised. It corresponds to a periodical application of the model, without reestimating coefficients, to produce the components of the seasonal model. This organisation is commonplace, but must not be considered as normative.

		Policy of the institution	**First adjustment**	**New data**	**Revision**	**data reintegration**
	Frequency		*Once*	*Monthly*	*Annual?*	
Decisions	Choice of method (TS or X12)	x ?				
	Choice of model					
	pre-processing options		x			x
	identification of model		x			x
	estimation of coefficients		x		x	x
	Application of model		x	x	x	x
	Adjustment options		1, 2, (3 ?)	4	5	1, 2, (3 ?)

The adjustment options shown at the bottom of this table correspond to the different radio buttons of the dialogue for launching the seasonal adjustment process in Demetra.



Another possibility, still sometimes used, consisted in estimating a seasonal adjustment model only once a year in order to produce, not SA data, but predictions for seasonal coefficients for the 12 months to come. Afterwards, it was enough to apply these coefficients to the raw data to obtain, using only one arithmetic operation, the seasonally adjusted value.

II.6. Some typical situations

II.6.1. Changing seasonality over the period under analysis

The methods used by Demetra authorises a slow modification of the seasonal effect over the whole period under analysis, ie. a progressive deformation of the structure of the model's seasonal coefficients. However, certain series can present a sudden break in seasonality, linked for example to a modification of the definitions of the variable retained or to its calculation mode or still to a regulatory or legislative change. Demetra does not detect such breaks and the models retained for series containing such breaks can be accepted without the problem being detected. This has significant consequences for the reliability of the pre-adjustment proposed, in particular around the point of the break.

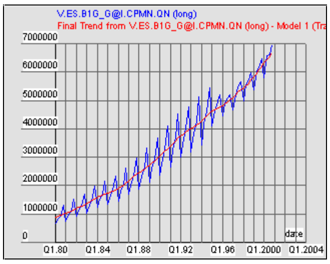
It is therefore up to the user to seek out such situations in their visual exploration of the data.

Ø *Modifying the period under analysis*

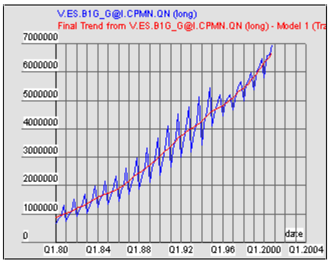
Here is a series affected by a sudden seasonality change (in the first quarter of 1995). We are going to study it over three different periods: the whole series (21 years, 84 observations), over 11 years and finally over the last period after the seasonality break at the start of 1995 (6 years, 24 observations). All the results shown are done so using the standard Tramo-Seats options for a new process.

Raw series and trend

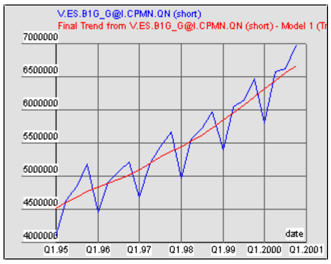
1980-2000	1990-2000	1995-2000



Rejected

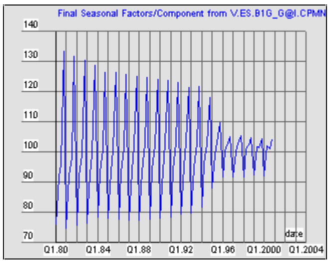


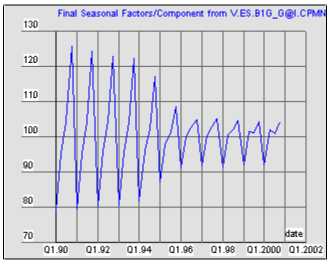
Accepted

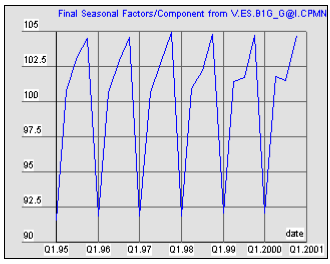


Accepted

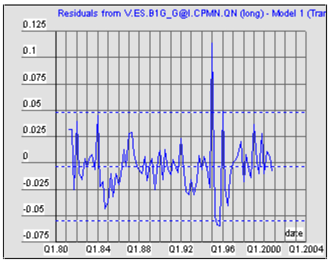
Seasonal component

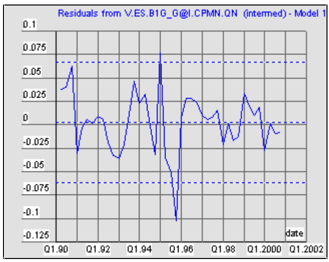


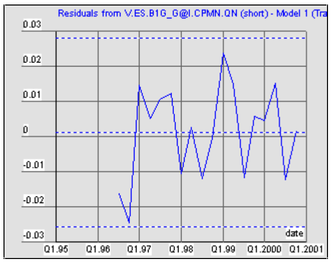




Residuals







In the second option, a statistic on residuals is significant at 5%, linked most certainly to how the residuals were configured in the transition year, 1995. But this is not enough to reject the model.

Which is the best choice? In other words, what is the interest of storing information from the past, or even distant past.

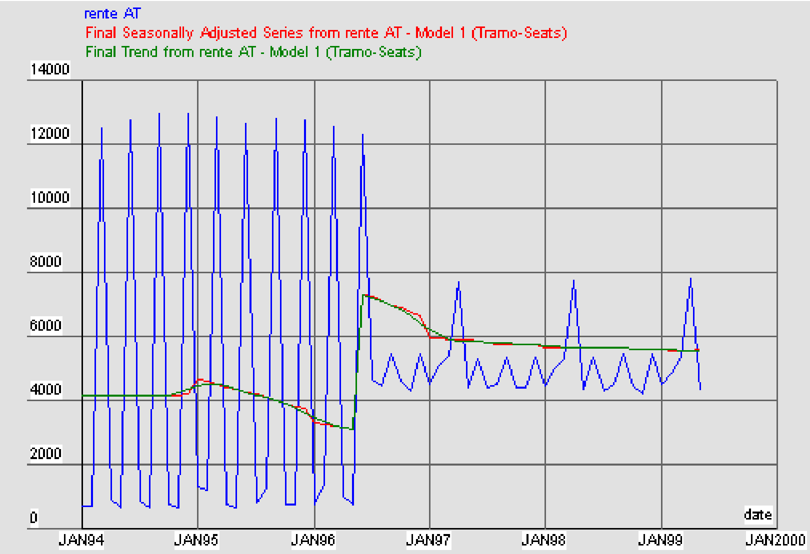
Besides, the orders of the non-seasonal part of the ARIMA models are the following: (1 1 3) for the longest period under analysis and (0 1 1) for the two others. The parsimonious argument favours, with equivalent results, those models which are the least complex. Therefore, here again, even if we could improve the model of the long series to make it more acceptable, the game wouldn't be worth the candle.

Therefore, following a pragmatic approach, it would seem that the third choice, the shortest period, is the one that allows the user to produce current estimations of the trend. The behaviour of the series has clearly changed and, what is more, we have enough observations under the new system to be able to estimate a model.

In this method, we will produce the seasonal pre-adjustments separately for the two periods:

- 1980-1995,
- the period after the break in 1995.

Ø *Integrating seasonality change into the model*



Looking at the graphical representation, it would seem that there was a modification of the seasonal behaviour of the series in July 1996. Despite this, the default model is accepted by Demetra: all the statistics on the quality of the adjustment are good! Such an observation means that exploring the data – even rapidly – is a 'must'. The user will be able to see immediately if there is a serious problem in the series; the slightly more detailed analysis of the results will rapidly persuade them of the unsuitability of the model retained, just by the transformation into logarithms alone.

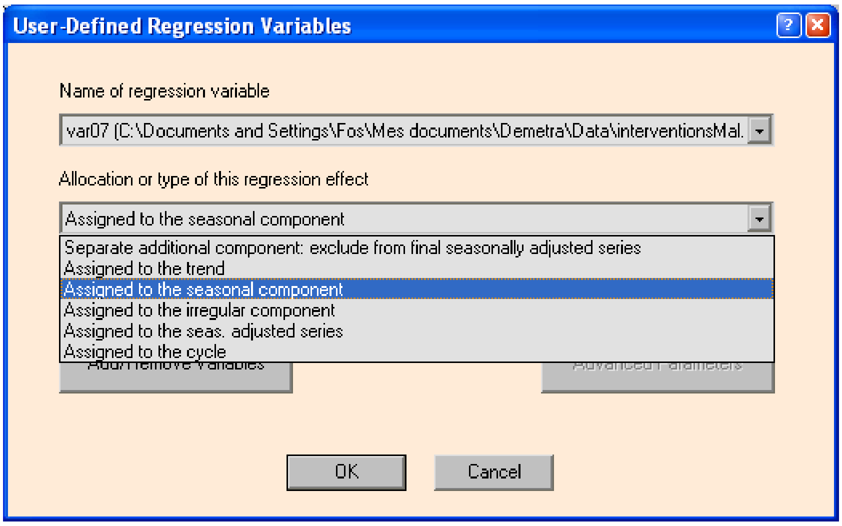
In cases of a similar nature, according to the scale of the change made and the respective length of sequences before and after the change of seasonality, the default model can be accepted.

In the example presented here, which comes from real data, we know exactly what modification was made: a particular payment made initially at the end of each quarter. From the middle of 96, it was decided to make a monthly advance payment and to correct the amounts paid in April each year.

Having identified the difficulty, what solutions can we find?

The simplest and the most radical solution consists in splitting the series into two parts. This supposes, of course, that the second part, after the break has sufficient observations for estimating a model.

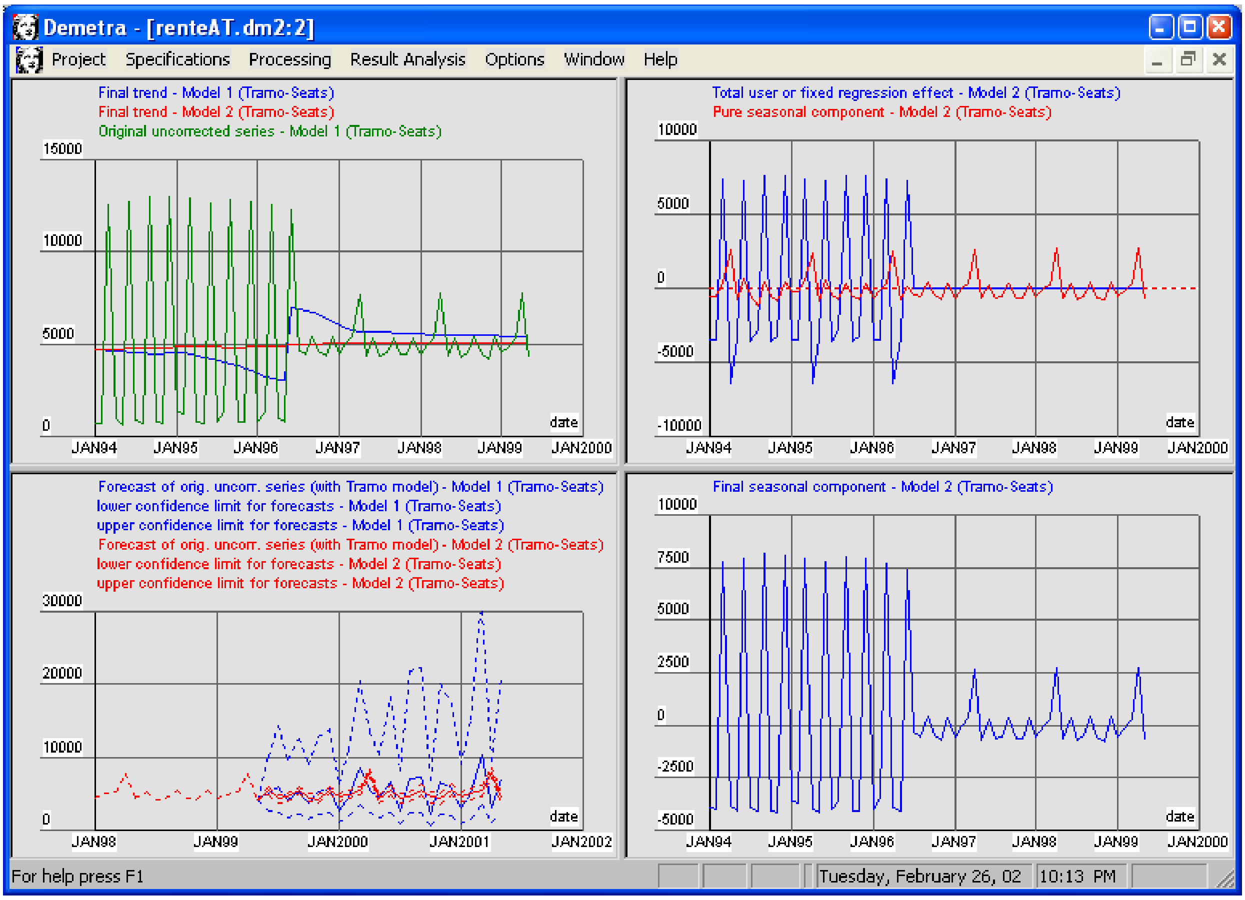
Another solution, much more difficult to implement, consists in introducing external variables which would affect the seasonal component. We have to create 11 intervention variables, for as many seasons in the year minus one. The values taken by these variables are 0 up until 1996 inclusive. From this date, they can take three values: +1, 0 and -1. Thus, we are going to create a first variable, for example var01 corresponding to January; it will be +1 for all the months of January after June 1996, -1 for the months of December in 1996 and the following years, and 0 everywhere else. We also create in the same way variables var02…, var11 corresponding to the months of February to November. There is no var12 variable for the month of December, indeed the very particular structure of our eleven variables guarantees that the sum of seasonal variations provoked by the break is zero.



We will then introduce these variables as external regressors, using the method introduced in the previous part. We will assign all these variables to the seasonal component. Each of them provides the estimation of the seasonal coefficient variation resulting from the break in seasonality. In order to produce a forecast of the modelled series, it is necessary to extend the structure that we have just described over the whole period desired.

In the set of parameters that we have just proposed, the reference situation is made up by the part preceding the break. With a view to seasonal adjustment for short-term or forecast purposes, it is perhaps preferable to do the opposite: consider the current situation (after the break) as the reference and reparameterise our external variables so that they estimate the deviation from the past compared with the present. In this approach, the past values of the regressors, +1, 0 or -1; the recent values and, above all, future ones are all equal to 0. This second approach makes it possible, eventually, when there are sufficient points, to abandon the oldest past of the series without having to set up the parameters again.

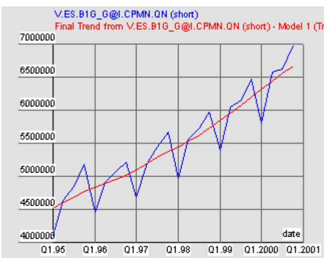
This is the choice we present in the following figure, obtained using the graphical comparison tool. In the top-left part, there is the representation of the raw series and the two trends calculated, one by the default model and the other after introducing the retrogressors; at the bottom, the forecasts given by these two models. The data up until May 1999 are observed data, the true forecasts start afterwards; they are accompanied by confidence intervals on the forecasts given. Noteworthy is that the forecasts provided by this model with regressors are not only in line with what we expect of a series whose behaviour is so regular, but are also much more precise.

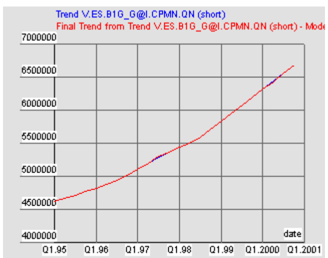


The right-hand section describes the seasonal component as it is estimated by the model with a break in seasonality: at the bottom, the ‘final’ component, at the top, the way of obtaining it. The series in red represents the ‘pure’ seasonal component, ie. without taking account of the regressors; with the parameter set retained, it is thus estimated on the recent part of the series. The blue series summarises the impact of the 11 regressors, its value is zero over the end of the period. The final component is the sum of these two series.

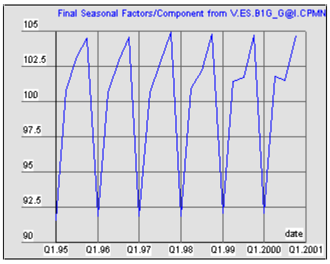
II.6.2. Non-idempotence: seasonality in the trend

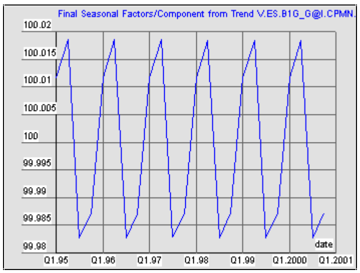
The term idempotence, little used, means that the application of a transformation to data, to which the transformation has already been applied once, gives an identical result to the first transformation. In our case, if we deseasonalise the trend obtained after a seasonal adjustment, we should obtain the same series. Nothing of the kind. From the previous example, by retaining the shortest period, we can deseasonalise the trend (graph shown below on the right) or the SA series (not shown), we find different results.





Seasonal component





In particular, we find, very strangely, a seasonal component in the trend! Take note, however, that the automatic scaling of graphs can lead to confusion: on the left, the seasonal variations go from -7.5% to +5%; on the right, it is ±0.02%! But the residuals of the right-hand model are so weak (between ±0.002%) that the seasonality is considered as significant.

We can also remark that the ARIMA models retained are very different. In particular, the order of the autoregressive part of the second model is much higher than in the first. The forecasts provided by the two models risk diverging, in particular if the forecast horizon is long.

This situation is theoretically bothersome, but practically of no consequence. It happens notably when the events have been detected and taken into account in the initial series, which is the case here. Indeed, the 'correction for outliers' used in Demetra introduces non-linearities.

Information on Models	Model 1 (Tramo-Seats)
Time Span (n° of obs.)	Q1.1995 - Q4.2000 (24)
Method	Tramo/Seats
PRE-ADJUSTMENT	
Transformation	Logarithm
Mean Correction	None
Correction for Trading Day Effects	None
Correction for Easter Effect	None
Correction for Outliers	Autom.:AO,LS,TC; 1 Outlier(...
Critical t-value	2.80
AO Q4.1995 t-value	3.66 [-2.021, 2.021] 5%
Corr. for Missing Obs.	None
Corr. for Other Regr. Effects	None
Specif. of the ARIMA model	(0 1 1)(0 1 1) (fixed)
Non-seas. MA (lag 1) value	-0.38
Non-seas. MA (lag 1) t-value	-1.55 [-2.021, 2.021] 5%
Seasonal MA (lag 4) value	-0.40
Seasonal MA (lag 4) t-value	-1.33 [-2.021, 2.021] 5%
Method of Estimation	Exact Maximum Likelihood
DECOMPOSITION	
ARIMA Decomposition	Exact
Seasonality	Seasonal model used

Information on Diagnostics	Model 1 (Tramo-Seats)
Ljung-Box on residuals	5.50 [0, 18.30] 5%
Box-Pierce on residuals	2.43 [0, 5.99] 5%
Ljung-Box on squared residuals	9.28 [0, 18.30] 5%
Box-Pierce on squared residuals	0.23 [0, 5.99] 5%
Normality	0.79 [0, 5.99] 5%
Skewness	-0.26 [-1.13, 1.13] 5%
Kurtosis	2.12 [0.74, 5.26] 5%
Percentage of outliers	4.17% [0%, 5.0%] ad-hoc

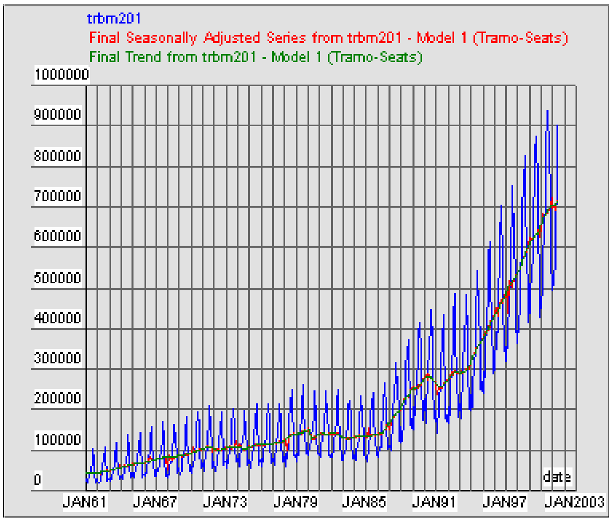
Model 1 (Tramo-Seats)
-> Model passes all diagnostic tests.

Information on Models	Model 1 (Tramo-Seats)
Time Span (n° of obs.)	Q1.1995 - Q4.2000 (24)
Method	Tramo/Seats
PRE-ADJUSTMENT	
Transformation	Logarithm
Mean Correction	Yes
Mean t-value	2.65 [-2.021, 2.021] 5%
Correction for Trading Day Effects	None
Correction for Easter Effect	None
Correction for Outliers	Autom.:AO,LS,TC
Critical t-value	2.80
Corr. for Missing Obs.	None
Corr. for Other Regr. Effects	None
Specif. of the ARIMA model	(3 1 0)(0 1 1) (fixed)
Non-seas. AR (lag 1) value	-1.66
Non-seas. AR (lag 1) t-value	-8.73 [-2.021, 2.021] 5%
Non-seas. AR (lag 2) value	1.29
Non-seas. AR (lag 2) t-value	4.95 [-2.021, 2.021] 5%
Non-seas. AR (lag 3) value	-0.42
Non-seas. AR (lag 3) t-value	-2.30 [-2.021, 2.021] 5%
Seasonal MA (lag 4) value	-0.69
Seasonal MA (lag 4) t-value	-2.24 [-2.021, 2.021] 5%
Method of Estimation	Exact Maximum Likelihood
DECOMPOSITION	
ARIMA Decomposition	Approximated
Seasonality	Seasonal model used

Information on Diagnostics	Model 1 (Tramo-Seats)
Ljung-Box on residuals	6.06 [0, 15.50] 5%
Box-Pierce on residuals	1.39 [0, 5.99] 5%
Ljung-Box on squared residuals	9.62 [0, 15.50] 5%
Box-Pierce on squared residuals	1.48 [0, 5.99] 5%
Normality	1.42 [0, 5.99] 5%
Skewness	-0.59 [-1.13, 1.13] 5%
Kurtosis	3.72 [0.74, 5.26] 5%
Percentage of outliers	0.00% [0%, 5.0%] ad-hoc

Model 1 (Tramo-Seats)
-> Model passes all diagnostic tests.

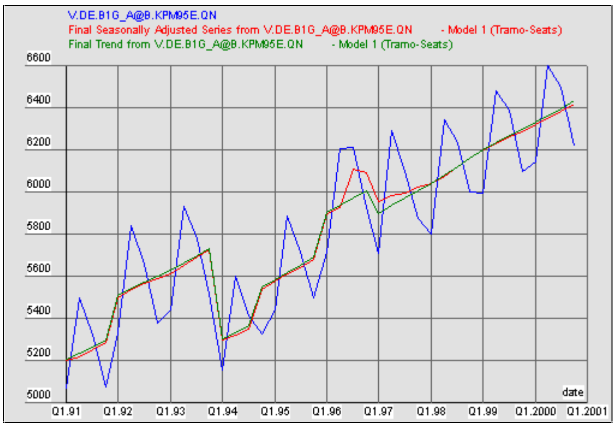
II.6.3. Significant Ljung-Box: structure of residuals



Trend break, very long series: cut down the series. The process must be relaunched by modifying the period for analysis in the definition dialogue of a new adjustment.

One can also delete the unretained points in the project data file. This allows to retain a single start date for the whole project. Indeed, if in the project definition, a start date is specified, this applies to all the series of the table. Of course, Demetra will process the series over the period given.

II.6.4. Too many outliers



The series is too irregular: we must raise the threshold for detecting the extreme points, even delete the automatic detection (see above, improvement by batch). In addition, note that we have 10 years of quarterly data and only 5% of outliers – the value for the rejection of an adjustment in the default options for the quality of models – limits us to a number of events strictly below two, be it a maximum of one!

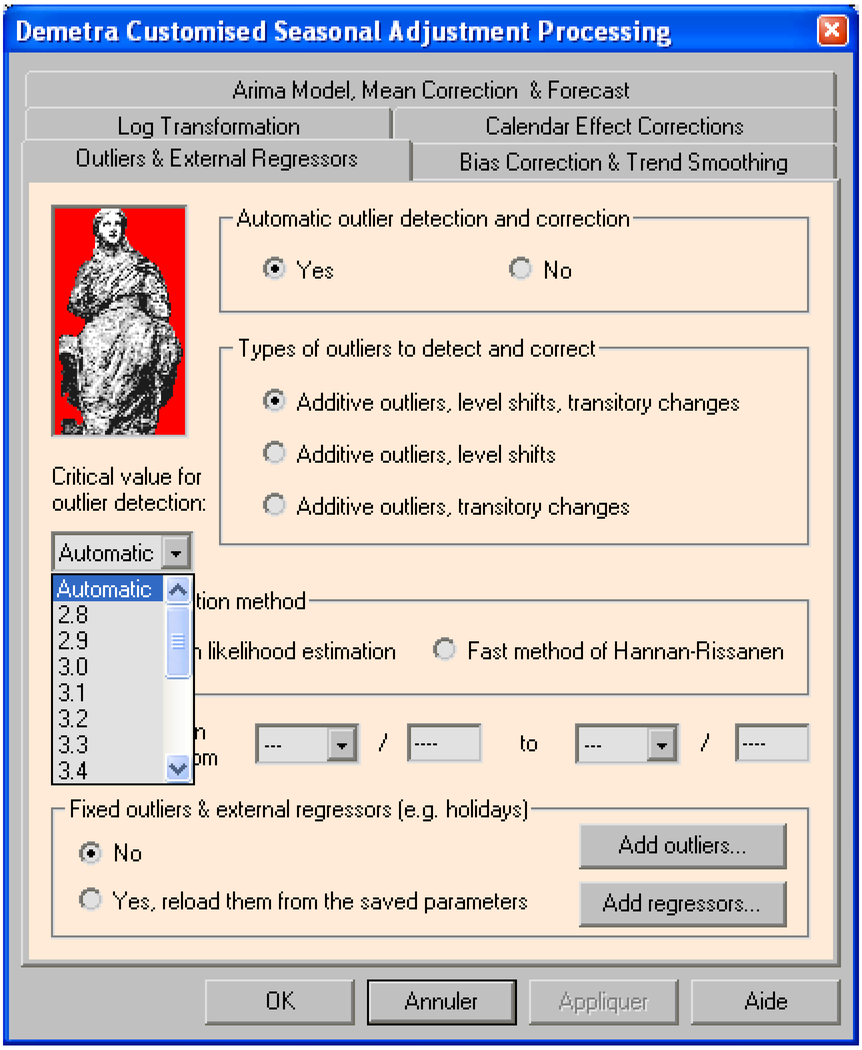
II.6.5. Incorporating *a priori* information

In the previous graph, we can locate a depression in the series over the two years 1994 and 1995. Let us suppose that the cause of this decrease has been identified, we could then imagine 'helping' Demetra in its search for *outliers* by explicitly telling it that this event exists. We are going to present two possible methods:

- introduce two events of LS type (over Q1.94 and Q1.96), or
- introduce an intervention variable, indicating that something happened over the two years.

Note that these two methods are not equivalent: in the first case, we ask Demetra to look for two rungs, two step values, independent from one another in theory. In the second, we consider that there was an effect over the two years (lag) and a return to 'the normal' after these two years have elapsed; the value of the second step is the opposite of that of the second.

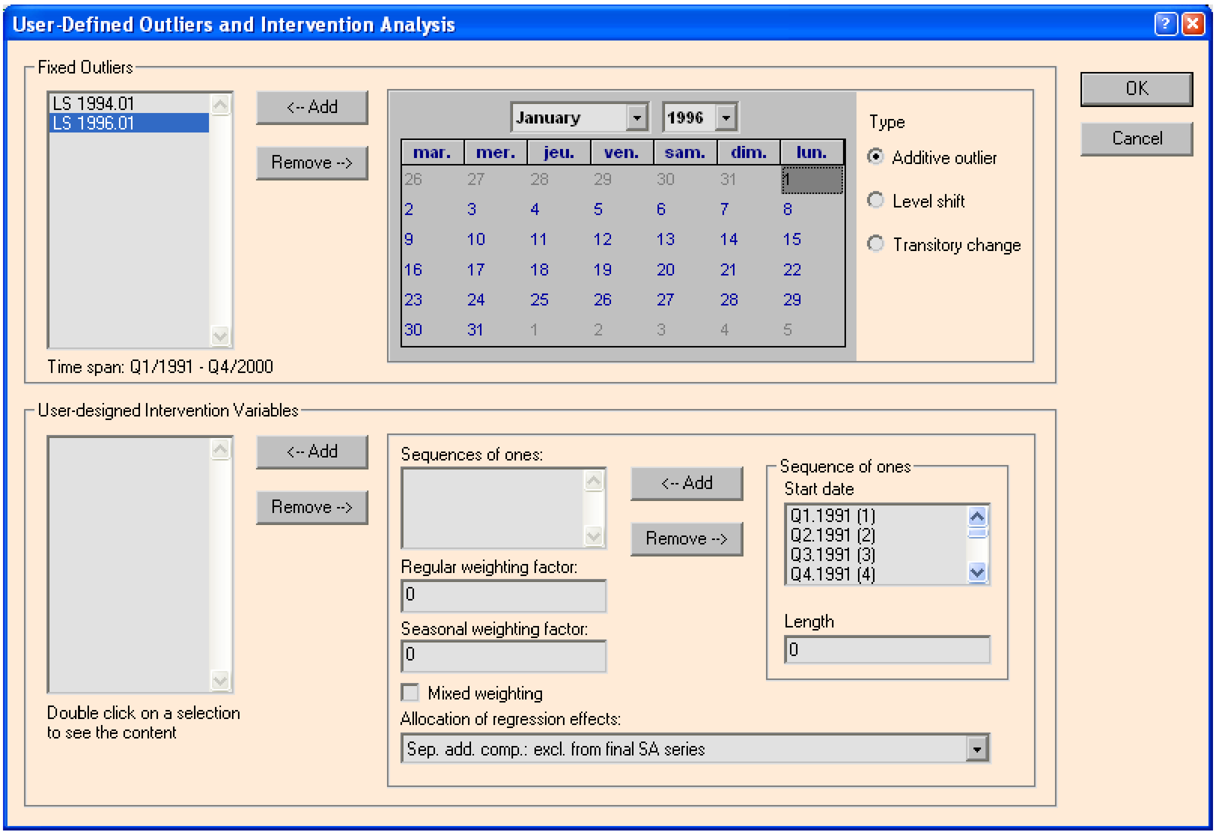
We will present, finally, the possibility of introducing a supplementary variable, chosen or defined by the analyst at the outset, and which they suppose exerts an influence on the level of the series to model. We will then speak of an external regressor.



Ø *Manual introduction of outliers*

In the batch analysis dialogue, let us choose the option *Customised parameters for a new processing*. In the following dialogue, at the bottom of the bookmark *Outliers & external regressors*, you can find the buttons, *Add outliers*… and *Add regressors*…

The first leads us to the following dialogue.

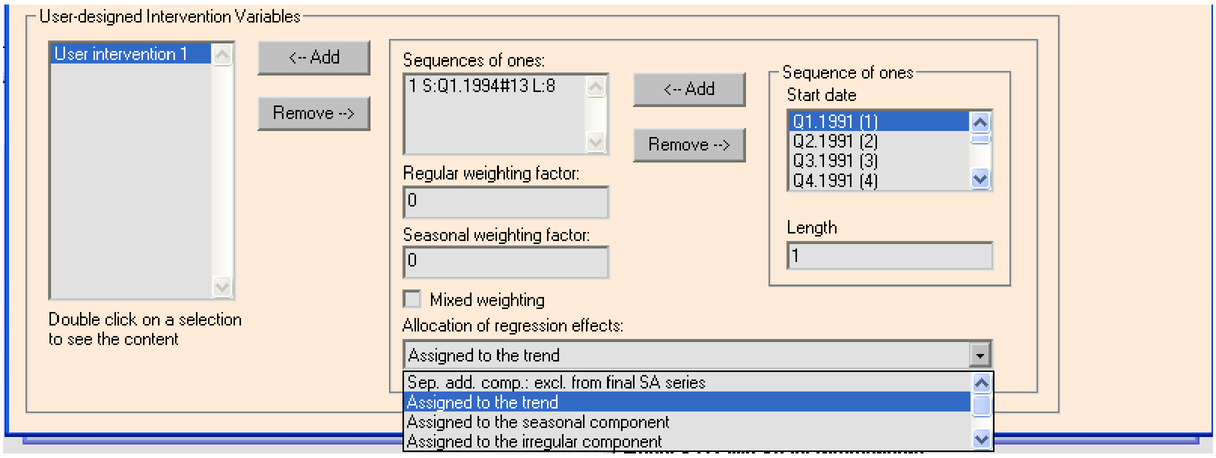


The top part concerns the events. The calendar makes it possible to choose the dates of events known by the analyst at the outset. Do not forget to give the day of the month (Demetra will remind this to you) and above all give the type of event that you wish to introduce using the three right-hand radio buttons.

Note that in our example, we introduce two events for a series whose data go from 1991 to 2000, be it 40 quarters. We are therefore already at the 5% ceiling for outliers fixed in compliance with quality rules. We will have to then forcibly accept this model if the other criteria are acceptable.

Ø *Introducing an intervention variable*

This second method is only possible if one uses the Tramo-Seats algorithm, X12-Arima not offering the same flexibility and necessitating the use of a regressor explicitly defined. This method does not seem accessible from the batch process (probably following a bug). Indeed, introducing such a variable, represented by a series of 1, is done by the bottom part of the dialogue shown above. However, when it is accessed from the batch process, the bottom right-hand *Sequence of ones / Start date* scroll list is empty or contains only the unformatted start date. It should contain the list of dates observed for the series. In order to test this option, we need to launch the improvement module for rejected series or shift the series into detailed analysis mode. We can then introduce a sequence of 1s, beginning in Q1.94 and lasting for 8 quarters.



We shouldn't forget to indicate in the bottom scroll box, *Allocation of regression effects,* the component of the model that must be modified by the intervention variable for which we have just chosen the period. In our case, we will assign it to the trend.

Tramo-Seats

Separate additional component: exclude from final seasonally adjusted series
Assigned to the trend
Assigned to the seasonal component
Assigned to the irregular component
Assigned to the seas. adjusted series
Assigned to the cycle

Tramo-Seats allows to assign the effect of a variable to any of the model's components (trend, season, irregular, SA or cycle). In practice, this means that the step, the height of which will be estimated by Demetra, will be considered as an *a priori* correction to assign to the selected component.

The first choice assigns the intervention effect to a particular component which will be removed from the series before any modelling, and whose impact will not show in the SA series.

Note that you can check the pertinence of introducing *a priori* information. We can see this in the information provided on the models.

Information on Models	Model 3 (Tramo-Seats fixed Outliers)	Model 4 (Tramo-Seats Intervention)
Time Span (n° of obs.)	Q1.1991 - Q4.2000 (40)	Q1.1991 - Q4.2000 (40)
Method	Tramo/Seats	Tramo/Seats
PRE-ADJUSTMENT		
Transformation	None	None
Mean Correction	Yes	None
Mean t-value	4.58 [-2.021, 2.021] 5%	--
Correction for Trading Day Effects	None	None
Correction for Easter Effect	None	None
Correction for Outliers	2 Outlier(s) fixed	None
Critical t-value	3.80	3.20
LS Q1.1994 t-value	-6.89 [-2.021, 2.021] 5%	--
LS Q1.1996 t-value	3.89 [-2.021, 2.021] 5%	--
Corr. for Missing Obs.	None	None
Corr. for Other Regr. Effects	None	1 Regressor(s)
User0 t-value	--	-8.08 [-2.021, 2.021] 5%
Specif. of the ARIMA model	(1 0 0)(0 1 0) (fixed)	(0 1 1)(0 1 0) (fixed)
Non-seas. AR (lag 1) value	-0.55	--
Non-seas. AR (lag 1) t-value	-- [-2.021, 2.021] 5%	--
Non-seas. MA (lag 1) value	--	-0.99
Non-seas. MA (lag 1) t-value	--	-- [-2.021, 2.021] 5%
Method of Estimation	Exact Maximum Likelihood	Exact Maximum Likelihood
DECOMPOSITION		
ARIMA Decomposition	Exact	Exact
Seasonality	Seasonal model imposed	Seasonal model imposed

Information on Diagnostics	Model 3 (Tramo-Seats fixed Outliers)	Model 4 (Tramo-Seats Intervention)
STATISTICS ON RESIDUALS		
Ljung-Box on residuals	4.44 [0, 25.00] 5%	29.81 [0, 25.00] 5%
Box-Pierce on residuals	0.95 [0, 5.99] 5%	0.72 [0, 5.99] 5%
Ljung-Box on squared residuals	16.74 [0, 25.00] 5%	10.11 [0, 25.00] 5%
Box-Pierce on squared residuals	2.70 [0, 5.99] 5%	0.45 [0, 5.99] 5%
DESCRIPTION OF RESIDUALS		
Normality	2.43 [0, 5.99] 5%	1.34 [0, 5.99] 5%
Skewness	-0.37 [-0.84, 0.84] 5%	-0.18 [-0.82, 0.82] 5%
Kurtosis	4.10 [1.33, 4.67] 5%	2.10 [1.35, 4.65] 5%
OUTLIERS		
Percentage of outliers	5.00% [0%, 5.0%] ad-hoc	0.00% [0%, 5.0%] ad-hoc

The method – consisting in introducing two steps – provides, in effect, two significant values (-6.89 and +3.89) which vary in absolute terms. The second method – supposing a temporary depression of the trend – also provides a largely significant value (-8.08). Be careful, these two sets of values are not directly comparable: the numbers are expressed in standard deviation units from the residual component. However, these two models are largely different, as are the respective sizes of residuals (their standard deviations). Choosing between the two methods essentially involves the analyst's interpretation of events.

Pay attention not to implement these two methods at the same time to process the same event. In the copy of the screen below, we have *simultaneously* imposed the two events *and* the modification of the trend.

Information on Models	Model 1 (Tramo-Seats)
Time Span (n° of obs.)	Q1.1991 - Q4.2000 (40)
Method	Tramo/Seats
PRE-ADJUSTMENT	
Transformation	None
Mean Correction	None
Correction for Trading Day Effects	None
Correction for Easter Effect	None
Correction for Outliers	2 Outlier(s) fixed
Critical t-value	2.80
LS Q1.1994 t-value	-195013152.00 [-2.021, 2.0...
LS Q1.1996 t-value	195013152.00 [-2.021, 2.02...
Corr. for Missing Obs.	None
Corr. for Other Regr. Effects	1 Regressor(s)
User0 t-value	195013152.00 [-2.021, 2.02...
Specif. of the ARIMA model	(0 1 1)(0 1 0) (fixed)
Non-seas. MA (lag 1) value	-0.99
Non-seas. MA (lag 1) t-value	-- [-2.021, 2.021] 5%
Method of Estimation	Exact Maximum Likelihood
DECOMPOSITION	
ARIMA Decomposition	Exact
Seasonality	Seasonal model imposed

Information on Diagnostics	Model 1 (Tramo-Seats)
STATISTICS ON RESIDUALS	
Ljung-Box on residuals	18.51 [0, 25.00] 5%
Box-Pierce on residuals	2.67 [0, 5.99] 5%
Ljung-Box on squared residuals	9.29 [0, 25.00] 5%
Box-Pierce on squared residuals	0.79 [0, 5.99] 5%
DESCRIPTION OF RESIDUALS	
Normality	2.53 [0, 5.99] 5%
Skewness	-0.43 [-0.85, 0.85] 5%
Kurtosis	1.92 [1.30, 4.70] 5%
OUTLIERS	
Percentage of outliers	5.00% [0%, 5.0%] ad-hoc

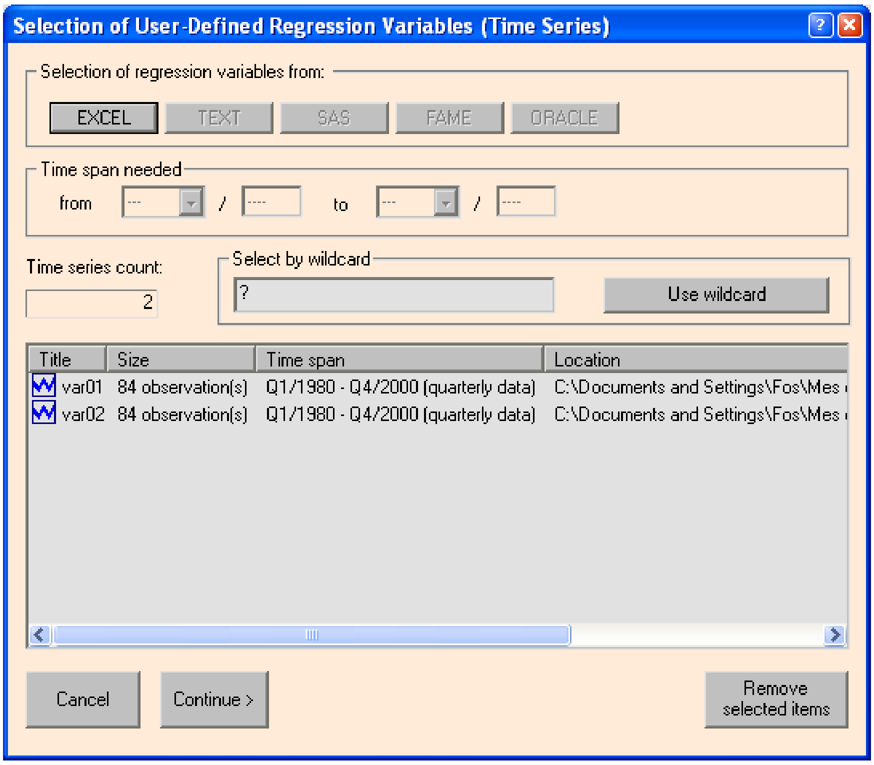
The result shows solely the problem of an over-large number of events. But there is a serious problem: look at the coefficients of events and of the intervention variable; they are gigantic ($2\ 10^8$…) and equal in the opposite direction. There is clearly a problem of colinearity between the two methods selected.

Ø *Introducing an external regressor*

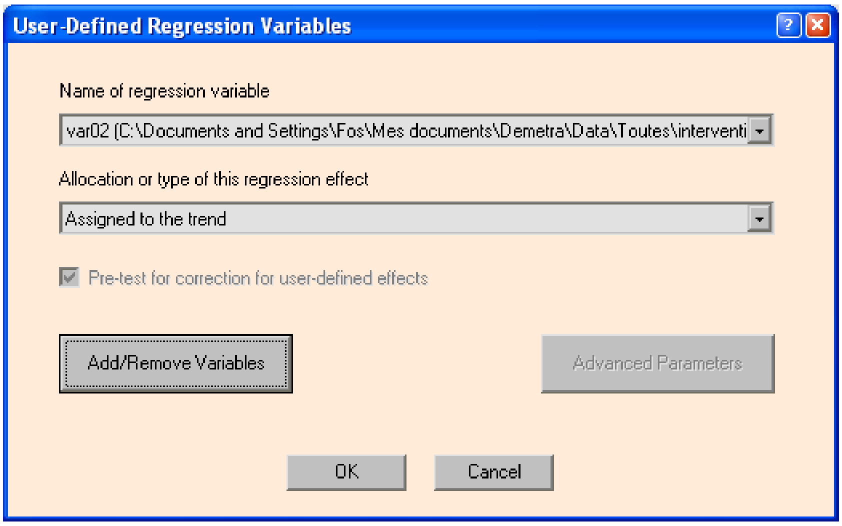
The third option – authorising the addition of external information known beforehand allows to introduce a regressor. This concerns one or more variables exerting an influence on the level of the series to deseasonalise. This option can be accessed by clicking on *Add regressors*… of the *Outliers & external regressors* tab. The option is accessible in the two large families of methods and whatever the work method (batch processing, improvement or detailed analysis). We will describe the way it functions in batch processing.

The *Add regressors* button calls up a set of two dialogues: the first allows the user to choose the file containing the external variable(s). Note that this file is necessarily of the same type as that containing the series for process: only the button corresponding to this type is activated in the dialogue.

On the model of the dialogue for selecting series for analysis, you must indicate the external variables that you are going to use for the seasonal adjustment model. Pay attention to select only those which will be effectively used.



Having selected the external variables, we then go to the following dialogue using the *Continue* button. Here, we will find in the first scroll list the list of variables selected in the previous phase.



For each of them, we need to assign a type of influence to the series to deseasonalise. The choice of the allocation depends on the method used. For Tramo-Seats, the proposed options for introducing intervention variables can be found higher up.

Tramo-Seats	X12-Arima
Separate additional component: exclude from final seasonally adjusted series Assigned to the trend Assigned to the seasonal component Assigned to the irregular component Assigned to the seas. adjusted series Assigned to the cycle	User Defined Constant User Defined Seasonal Trading Day Length-of-Month Length-of-Quarter Leap Year Stock Tradind Day Stock Length-of-Month Easter Labor Day Thanksgiving Day Additive Outlier Level Shift Ramp Effect User Defined Holiday Statistics Canada Easter Temporary Change

For X12-Arima, the list, which is longer, makes it possible to assign the impact of the regression to different parameters of the model. The first three elements of the list are equivalent to their available counterparts for the Tramo-Seats method. In the list you have the possibility of processing different types of events (AO, LS, TC and Ramp) or the different types of calendar effects.

To conclude, note that the management of outliers and regressors can seem quite cumbersome: for each new attempt, you need to reintroduce the events and the sequences of ones or select the file containing the variables. It is also interesting to use the option *Reload from the saved parameters* proposed in the dialogue. You can also keep the prior information definitions while testing the various modelling options.

Ø Reusing a model

When you process a large number of series by batch in the automated module, some of them can be similar or, at least, closely connected. Therefore, if you deseasonalise statistics of jobseekers, you will probably have to work at different levels of information groupings: analysis by geographical zone, sex, age group. In this context, diverse series often show very similar characteristics.

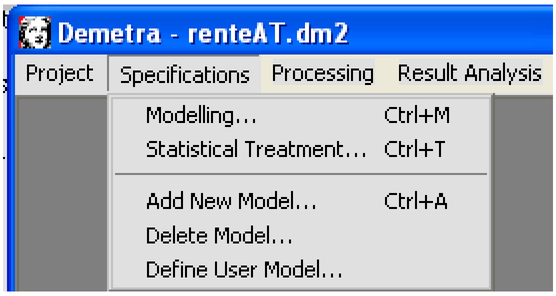
However, it is possible that similar series do not produce similar seasonal adjustment models, or even that while one of the models is accepted, the other is rejected. It can therefore be interesting to select the model retained for one of the series and apply it to the other. To do this, there are three possibilities:

- manual recuperation of the definition of the model, or
- copying and pasting in the project definition file,
- creation of a file storing the definition of the model (file "**.mdl**").

The first option can be used while staying within the automated processing module: on the basis of the results obtained after processing the model series, you can select the parameters that you wish to reuse for studying the second series. You will then choose a customised processing of the latter and go through the tabs of the dialogue to fix the corresponding choices yourself.

The second option involves going into the project definition file (in Excel, it is the *Demetra parameters* tab of the file containing the data. Look for the model series and you will find the definition, in text form, of all the option specified by the user. Copy this text and paste it into the corresponding space for each of the series to be parameterised. This done, don't forget in Demetra to produce an update of the series (in the project series table, right click and select the menu *Time Series -> Update …*

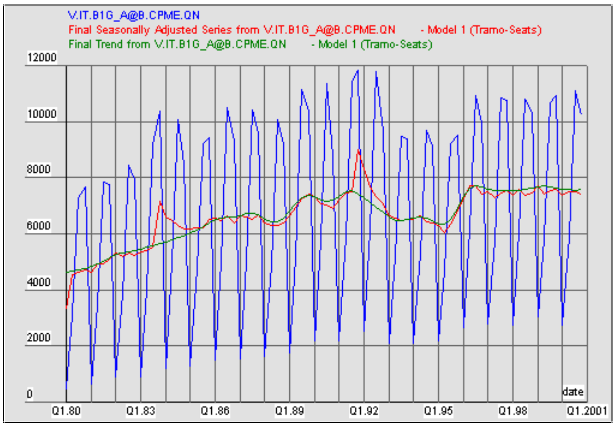
The third solution supposes passing by the detailed analysis module. You are going to create a file to study the model series. You can improve the model to a great detail. When you are satisfied with the model obtained, you can export it by selecting *Define User Model* in the *Specifications* menu.



The dialogue for selecting the model in the detailed processing will allow you to designate the model to be retained, then the dialogue for recording a file will allow you to create a model file (of type MS-DOS **.mdl**).

Afterwards, you will use this model by selecting the third option of the dialogue for selecting the process (*Parameters from model file for a new processing*) which will allow you to designate the file to use as a model.

II.6.6. Three statistics on residuals significant at 5%



The visible events (Q4.83 and Q4.91, both of TC type, look surprising. The plateau of the 1993-1995 period is not detected as an event and it seems that there is a change in the trend starting in 1993.

The data are transformed into logarithms, which does not seem justified at all (evolution of the raw data in a channel with parallel sides).

Several solutions are possible:

4 increase the threshold of event detection,

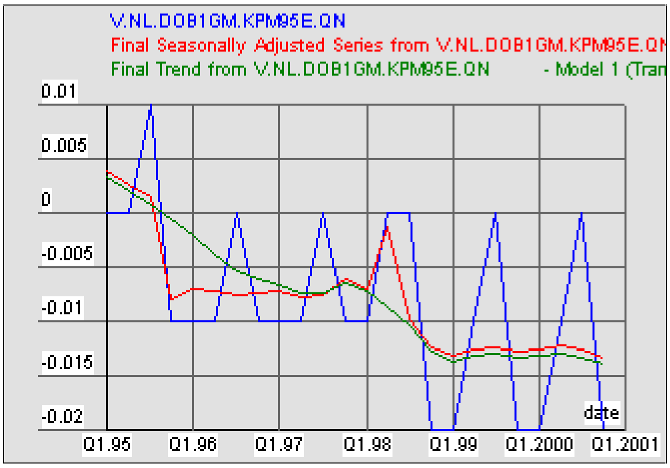
4 do not detect type TC events, without changing the threshold,

4 block transformation by logarithms.

Each of these actions allows the model thus obtained to be accepted.

II.6.7. Non-seasonal data

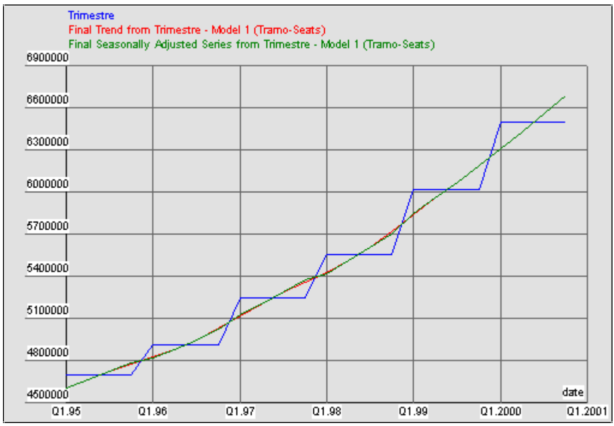
Ø *Discretised data*



The model is rejected: too many outliers and a significant residual result. Looking at the data, we can see rounded values: out of the 24 values observed, we only have 4 distinct values (-2%, -1%, 0% and +1%). It is probably because of a growth rate, for example, of the price index or GDP, where the value is given as a percentage, without a decimal. For such data, discretised, it is imprudent to use models which explicitly process real variables.

In the particular case of growth rates, we prefer to work directly with the original variable (price index or GDP). Moreover, we can remark that the modelling method retained for processing the original series integrates the possibility of having a model for the variations of the series. This is the meaning of the 'I' in ARIMA.

Ø *Annual series broken down into quarters*



In the example, the default options of Tramo-Seats detect two residual statistics significant at 5% (in blue) and a high number of outliers, which leads to the model's rejection. The assistant's automatic improvement retains the excess of outliers and keeps other residual statistics significant at 5%.

By opting for the automatic non-detection of outliers, the two Box-Pierce tests on residual statistics become significant, the distribution of residuals is asymmetric and the model is rejected.

Clearly, there is a structure in the residuals, it is evident on examination of the curve: the four quarters possess the same value. These data certainly come from annual and *'quarterised'* values which were used to produce a pseudo quarterly data series.

The series is not, in reality, a series with an infra-annual periodicity of observation but an annual series. There is therefore no need to deseasonalise. Either it must be established whether real data exist or the project series must be deleted.

II.7. To conclude

We should not forget that the problem of seasonal pre-adjustment of a time series gives rise to an infinite number of possible answers. Among these is: can one be sure to have found the 'good' seasonal adjustment model, if not the 'best'?

Unfortunately, the answer is no! And besides, what would the criteria be to judge whether one model is 'better' than another?

We have seen that we have numerous indicators at hand which, if they do not highlight the right model, sometimes allow the user to locate the bad ones. We have also seen that this palette of indicators can, occasionally, let a bad model slip through. This is therefore a last chance to recall the extreme importance of visually checking the series produced: it would be a great pity not to take advantage of the capacities of analysis and of graphical representation that the human brain enjoys!