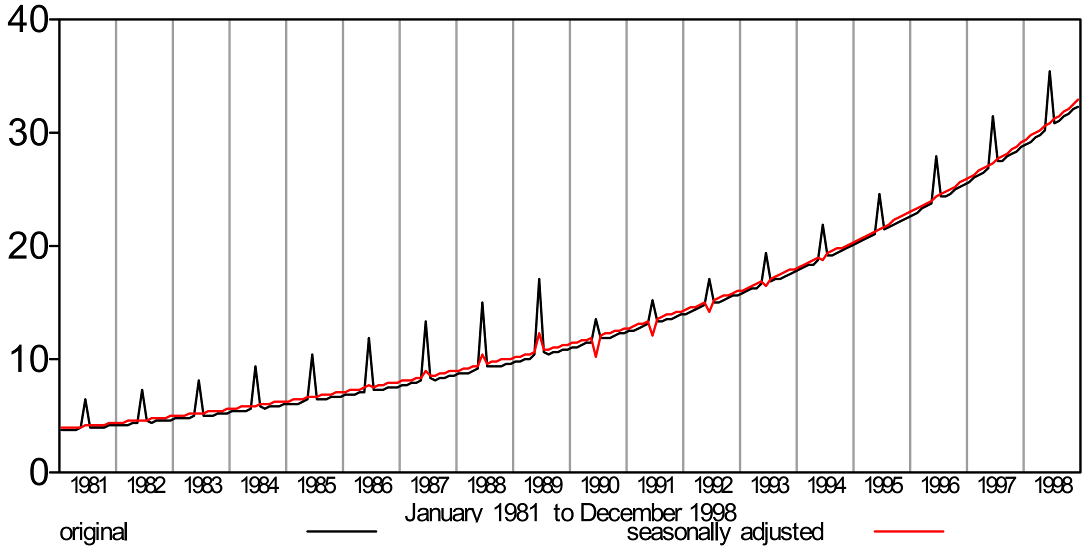
Understanding Seasonal Breaks

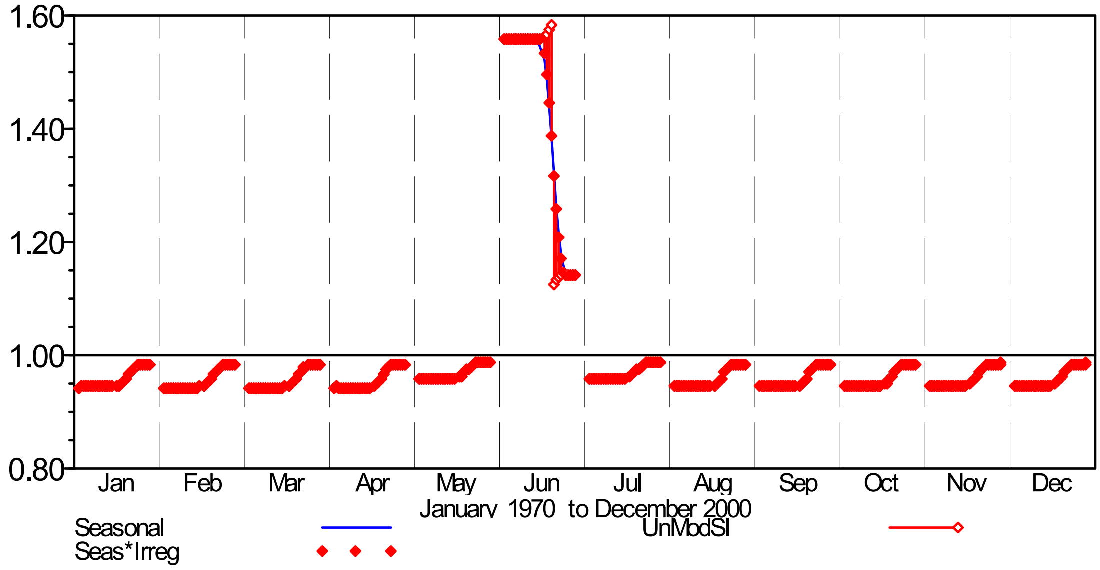
The purpose of this document is to illustrate that there is no unique, clear-cut way to estimate seasonal break (SB) corrections. The SEASABS algorithm sometimes gives non-optimal results. Below we use a simulated seasonal series with a large SB to compare SB factors and locations picked up by the SEASABS algorithm with those computed and fine-tuned manually. Note that there are many other ways to arrive at the same conclusions, however, we must use a method which SEASABS understands [i.e. that can be entered into the SEASABS priors GUI]. In the end, it's the journey that matters, so let's find a way to make it enjoyable.

Simulated Series with SB:

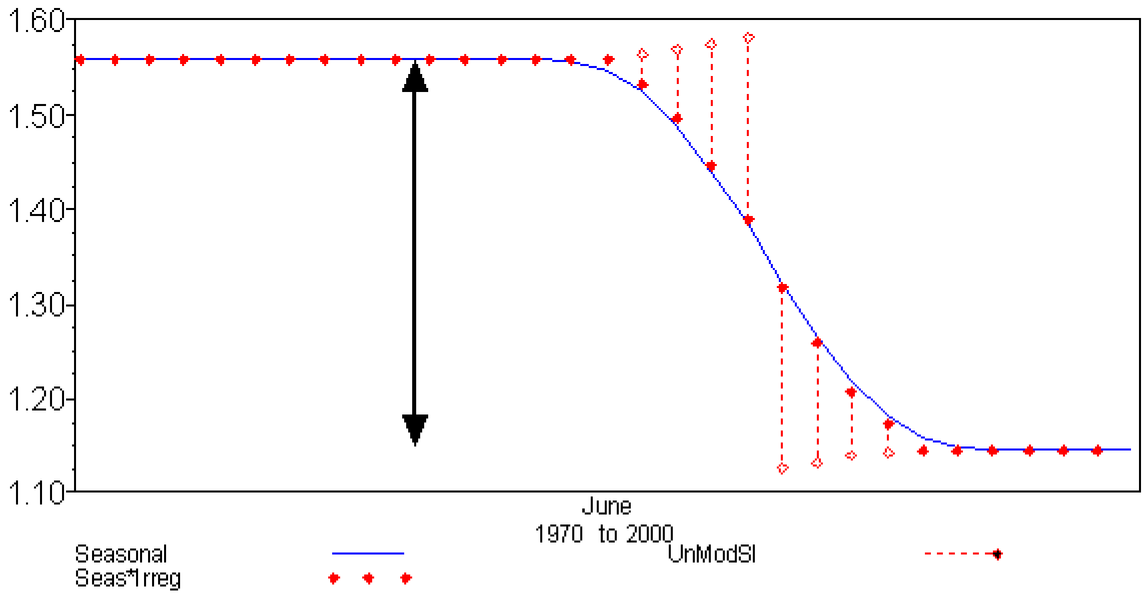
Graph 1: Here's a simulated series [original=black] run through X11 with SA and Trend series superimposed. No SB corrections were made. Note the distorted SA estimates around the seasonal break region. This is residual seasonality due to bad seasonal factor estimates caused by the SB.



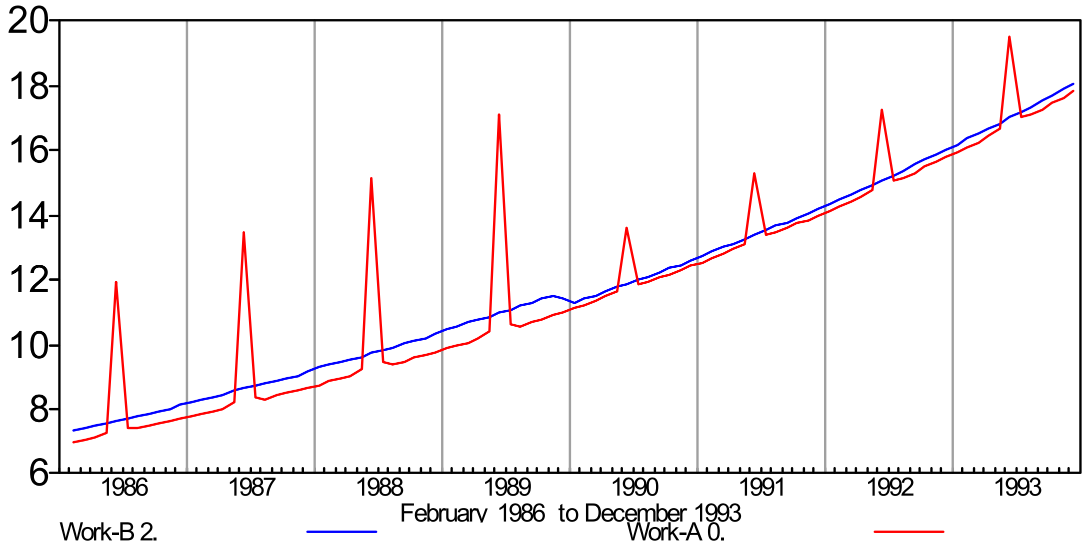
Graph 2: Here's the S*I chart (a SB is clearly evident in June):



Graph 3: Following is a zoom-in of the above for June. Note the internal corrections (dashed red lines) that are used by SEASABS to arrive at the blue seasonal moving average (SMA) line from which seasonal factors are taken. Our goal is to reduce the incidence of these internal corrections, i.e. make the blue line more representatative of the red (real) seasonal data points.

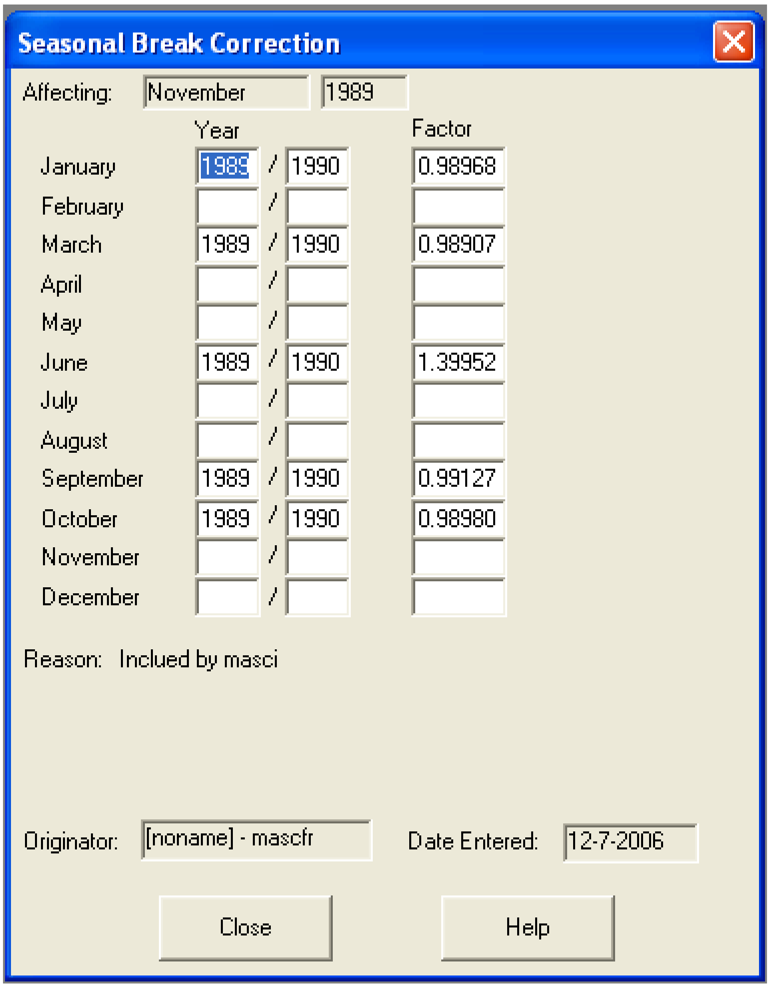


Graph 4: The tiny SBs in all other months in the above S*I chart are a consequence of the initial trend estimate from a centered moving average on the original used to estimate the S*I's, i.e. there is a level shift at 1989/1990 (at the location of our simulated SB; see following figure). A level shift before and after the SB will give a change in mean S*I levels for all months at the same location since S*I = O/T.



SEASABS SB Algorithm:

The prior factors and locations picked by SEASABS are as follows:


Seasonal Break Correction

Affecting: November 1989

	Year		Factor
January	1989	1990	0.98968
February			
March	1989	1990	0.98907
April			
May			
June	1989	1990	1.39952
July			
August			
September	1989	1990	0.99127
October	1989	1990	0.98980
November			
December			

Reason: Inclued by masci

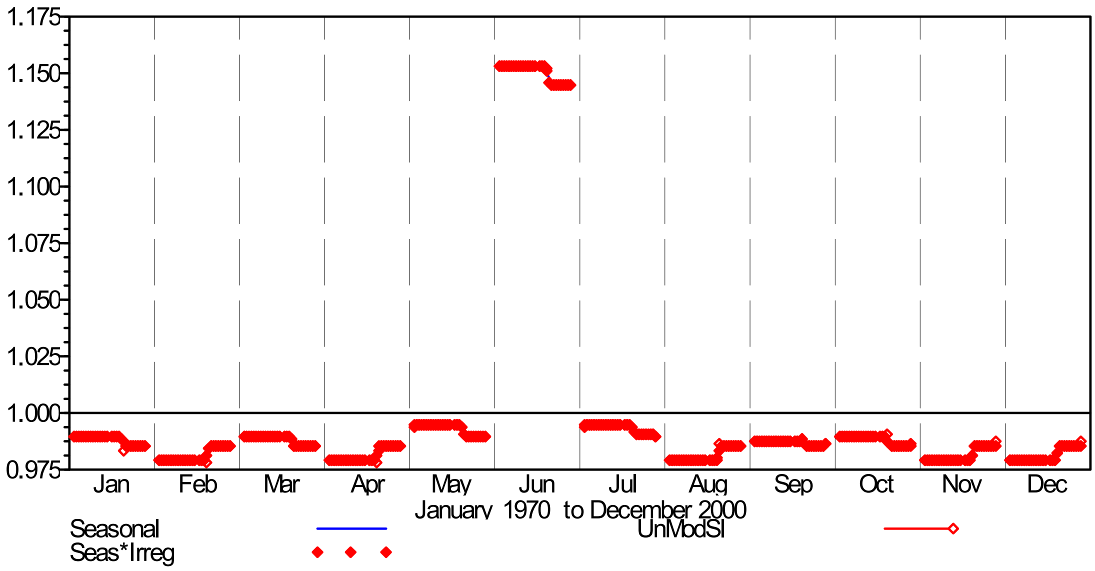
Originator: [noname] - mascfr

Date Entered: 12-7-2006

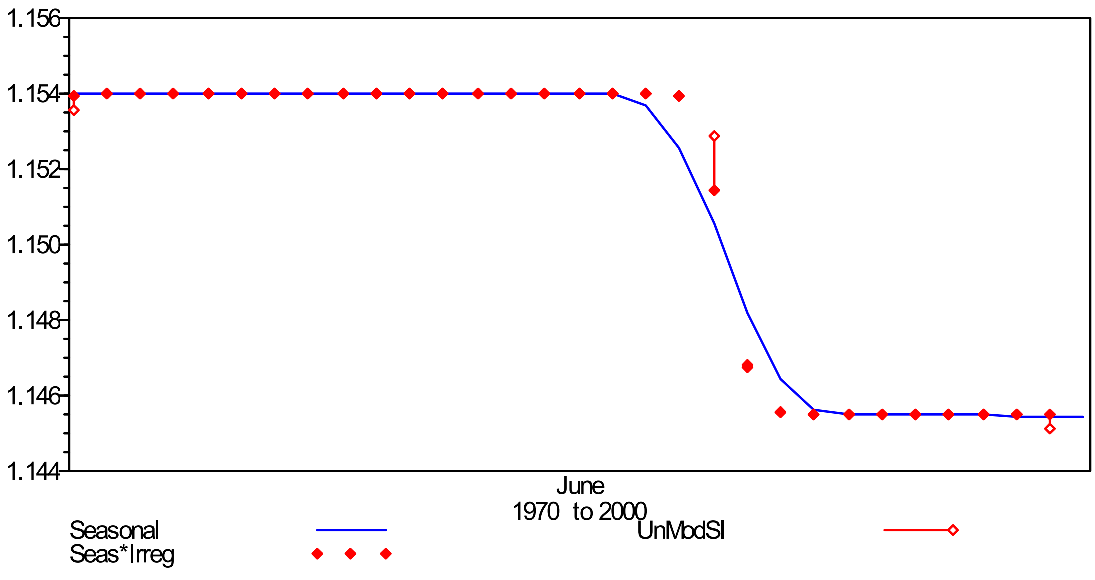
Close Help

Table 1.

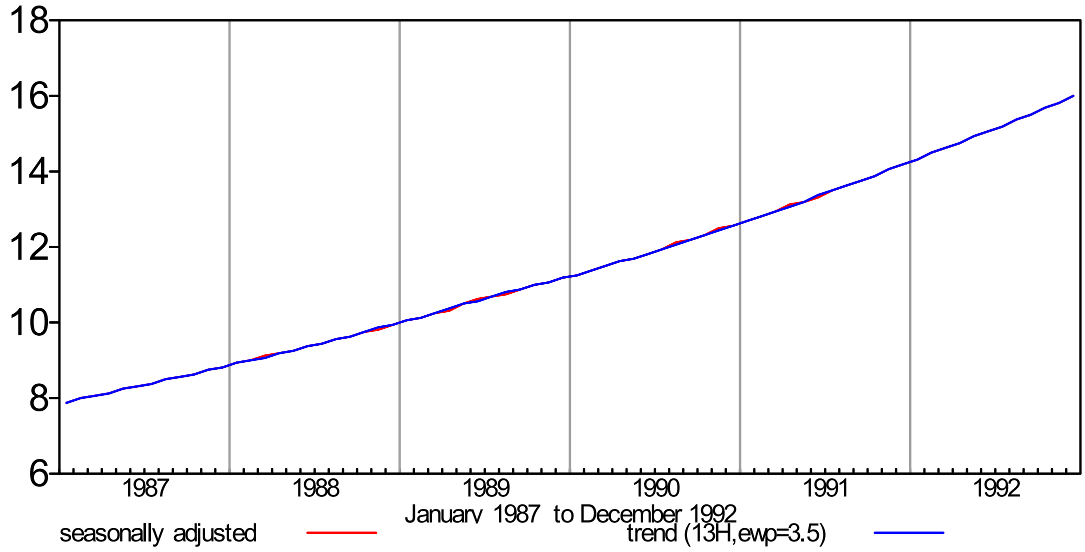
Graph 5: The after SB corrected S*I chart using the SEASABS algorithm:



Graph 6: Following is a zoom-in of the above for June. Note that the internal corrections are minimised compared to the before SB correction S*I chart above (graph 3). This appears adequate, however, the B1 total after SB correction differs by ~1% from the B1 total before SB correction. Can we do better if strict B1 balancing is enforced? See below.



Graph 7: The SA and Trend estimates after correcting for the SB using the SEASABS algorithm. There is no evidence for a residual level shift in the trend (as was apparent before SB correction in graph 4 above). This is a good thing. We can stop right here and not attempt to experiment for ourselves using different SB factors, locations and balancing options. However, what do you do if SEASABS doesn't pick up a SB, or, if you're really adamant and want to manually insert one yourself? This is the subject of the next section.



Manual Method 1: equalising S*I levels

As a first test, this method approximates SB factors from the sole criterion of equalising seasonal factor levels in all periods where a seasonal break can be justified. The main steps used to arrive at SB corrections (with reference to the simulated example above) are as follows:

1. Having identified the "break period of interest" (June), we now need to identify other periods in which to redistribute the "excess" seasonality to be 'removed' from June at years <=1989. This is so that we can balance the B1 totals later, i.e. so that they remain unchanged after application of all SB priors. There are two ways in which to pick these "balancing periods", depending on whether they are compensating or non-compensating:
(i) compensating periods are those where we can "gently" throw-in/remove a little more seasonality in order to compensate for the reduction/addition of seasonality in our main break period. In other words, these are periods where there is a hint of a break that goes more-or-less in the "opposite" direction to the break in our period of interest. We want to ensure that inclusion of seasonal break corrections are justified. Note that if one has a large seasonal break for one period (e.g. example above) then tiny breaks in other periods are inevitable (why? graph 4 may give you a clue). The example above obviously falls under this category. A rule of thumb when picking balancing periods is to pick those with the "noisiest" S*I values, or, deviate most from the blue SMA lines. I.e. basically those periods where inclusion of a SB might alleviate the discrepancy between S*I values and smoothed (seasonal factor) estimates from a SMA.
(ii) non-compensating periods are those which show no evidence of seasonal breaks going in the opposite sense that may be used for B1 balancing. Inclusion of a break may not be appropriately justified. However, we can still perform the B1 balancing by apportioning the excess seasonality removed from our main break equally into all other periods (i.e. spread it out thinly so that we avoid introducing any new undesired breaks). More weight should be given (that is, larger SB factors can be allowed) to periods with the "noisiest" S*I values, or, deviate most strongly from the blue SMA lines.

2. We then attempt to estimate initial SB factors for the periods identified in (1). A good approximation is the ratio of mean S*I level before/after the break years in each period. In the example above, this is before/after the 1989/1990 break years. The arrow in graph 3 above indicates what we're after: the required factor for June from the uncorrected S*I chart is f ~ 1.58 / 1.16 ~ 1.36. One can arrive at more accurate measures by using the uncorrected D10 (seasonal factors) table. We do the same for every other period where a break can be justified. In general, when estimating SB factors f_j for some period j , there are three cases to consider:

   (i) if there is no evidence for a break, then explicity f_j =1 (under a multiplicative decomposition model of course).

   (ii) if seasonal factor levels are ~ constant before/after the break year for period j "SByr(j)" (=1989 in above example),

$$f_j \approx \frac{[\text{mean S*I at years} \leq SByr(j)]}{[\text{mean S*I at years} > SByr(j)]}.$$

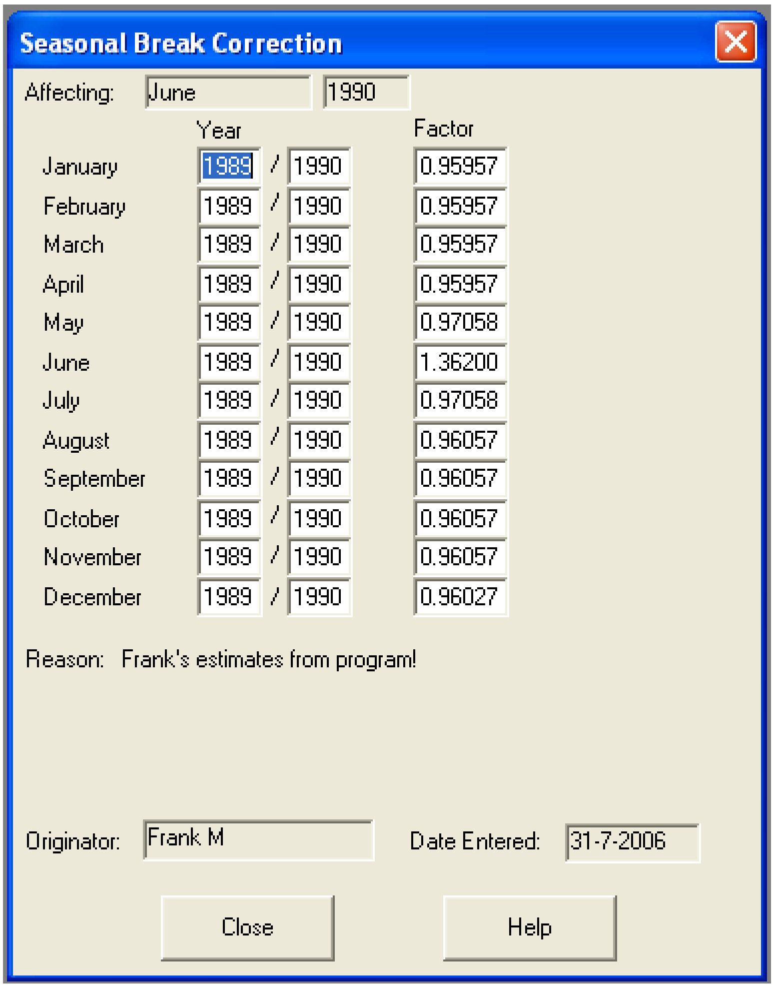
   The simulated case above is an example of this.

   (iii) if there is moving seasonality, the f_j is really the size of the "jump", i.e. the size of the discontinuity at the break years: SByr(j) to Sbyr(j) + 1 for period j:

$$f_j \approx \frac{(\text{S*I})_{SByr(j)}}{(\text{S*I})_{SByr(j)+1}}.$$

3. Having estimated your initial SB factors, you now need to ensure they give "balanced" B1 totals. In other words, these 'approximate' factors need to be rescaled such that the original B1 total is still equal to the B1 total with these rescaled factors applied to each period. This can be performed by entering these, together with B1 series from SEASABS into the *R* program described in Appendix I. This program performs exact balancing, i.e. such that the B1 total remains the same (within machine precision) after all SB factors are applied. Recall that SEASABS currently allows a 2% variation in the B1 total. This does not seem methodologically correct. It may have to do with the inner workings of the automated SEASABS algorithm.

4. The rescaled (or "B1-balanced") SB factors from step (3) can now be entered into the SEASABS "Prior Corrections" GUI (e.g. Table 2 below). You are now ready to perform a "Research Analysis" run. Note that if at any time you tweak/readjust the SB factors in the SEASABS GUI, you must rebalance them. This can be done using either the *R* program or, by just looking at the B1 table total in SEASABS an ensuring this is appropriately close enough to the original total (i.e. the pre-SB corrected total).



Seasonal Break Correction

Affecting: June 1990

	Year		Factor
January	1989	1990	0.95957
February	1989	1990	0.95957
March	1989	1990	0.95957
April	1989	1990	0.95957
May	1989	1990	0.97058
June	1989	1990	1.36200
July	1989	1990	0.97058
August	1989	1990	0.96057
September	1989	1990	0.96057
October	1989	1990	0.96057
November	1989	1990	0.96057
December	1989	1990	0.96027

Reason: Frank's estimates from program!

Originator: Frank M

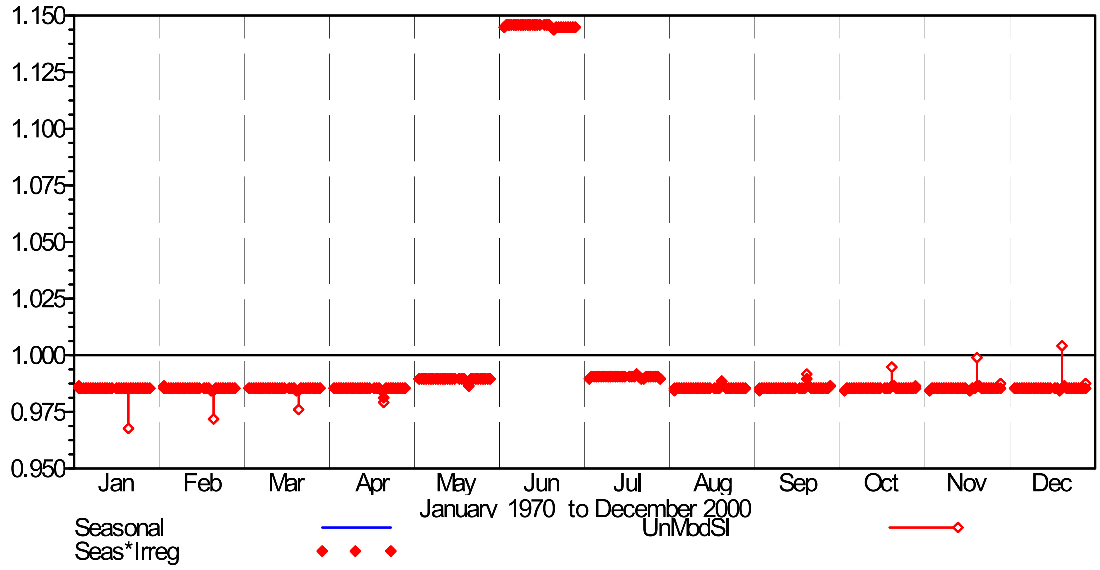
Date Entered: 31-7-2006

Close Help

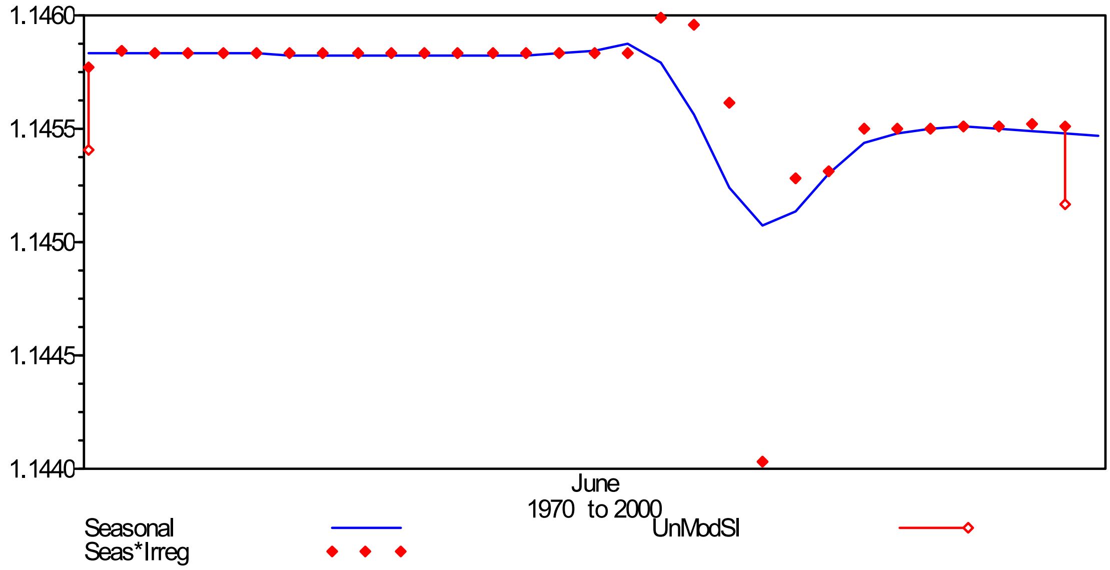
Table 2.

5. You can now perform a few sanity checks after correcting for SBs:
(i) peruse the S*I chart to ensure that all prior SBs were appropriately minimised and seasonal factor estimates are reliable for all periods (i.e. internal corrections to SMA derived seasonal factors are minimised);
(ii) ensure the B1 table total is ~ the same before and after SB correction.
(iii) peruse the D10 (seasonal factors) table and ensure they average to ~ 1 in every year.
(iv) examine graph of final seasonally adjusted overlayed with final trend cycle to ensure no systematic level shifts (Trend Breaks) greater than the inherent volatility were introduced (e.g. see graph 10 below). If so, then it is recommended you read through the section below: " *Manual Method 2: preventing/minimising potential level shifts* ".

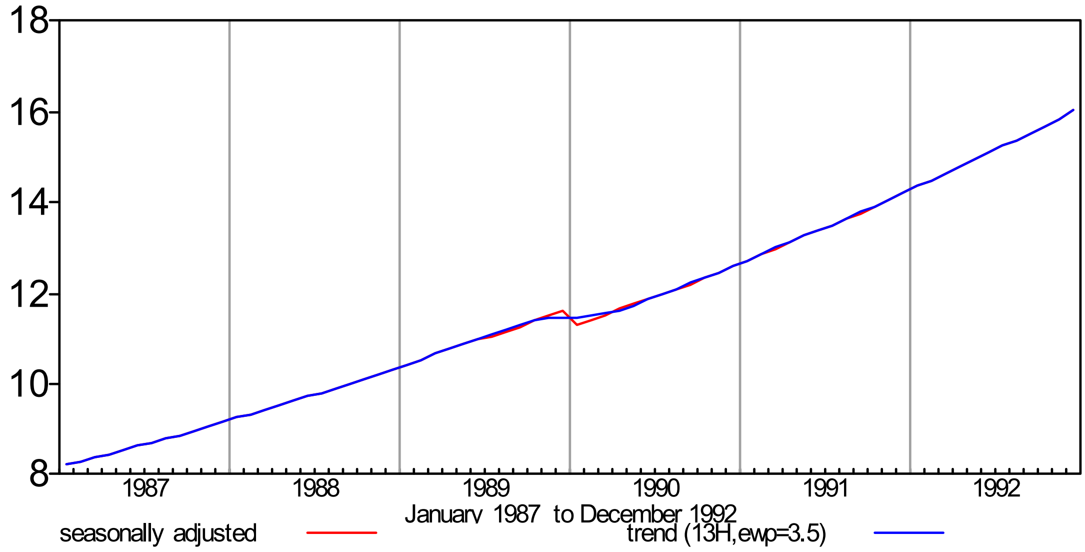
Graph 8: The after SB-corrected S*I chart using manual method 1 which attempts to equalise mean S*I levels before and after the break in all months:



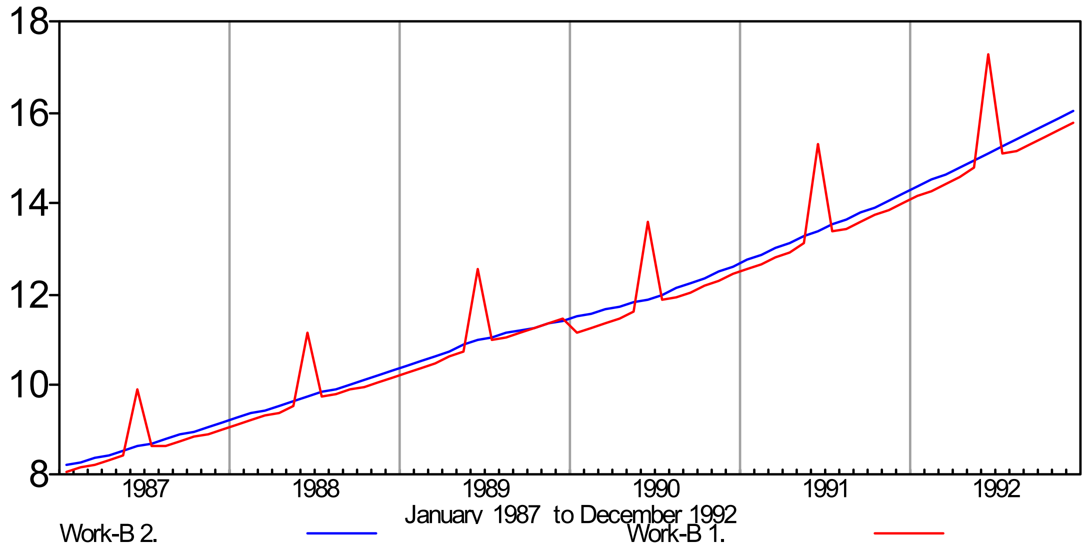
Graph 9: Zoom-in of the above for June. Note that we did a 'better' job than SEASABS at smoothing the break, however, see graph 10.



Graph 10: The following shows the SA and Trend estimates after correcting for the SB using manual method 1. Note the level shift around Dec/1989 - Jan/1990! This is a consequence of our criterion of equalising S*I levels for <u>every</u> month. In a nutshell, this is because in order to keep the original total the same before and after SB insertion, there must be a compensating increase in the trend before the seasonal break. I.e. given the multiplicative model, O = T.S.I, enforcing the <u>same</u> "S" before and after the break years 1989/1990 <u>in all months</u> means that "T" must be correspondingly higher before 1989 to account for the 'excess seasonality' (graph 4) removed from June and hence conserve "O". See also graph 11.



Graph 11: The following is the B1 (O corrected for SB priors = O/SB) with its first iteration trend cycle (B2) overlayed. Note that S*I ~ B1 / B2 where B2 = MA(B1) [i.e. a centered moving average on the B1]. The resulting S*I chart was in graph 8 above. So, just to reiterate the discussion in graph 10, the level shift in the final trend estimate is a consequence of a systematic shift in the initial B1 series, simply brought about by the principle of "conservation in the original".



Manual Method 2: preventing/minimising 'potential' level shifts

If a series is substantially volatile (which the above simulated series is not), then you may be able to get away in introducing a noticeable level shift from the criterion of strictly equalising S*I levels, as assumed in 'Manual method 1' above. I.e. it might get hidden amongst the volatility and become insignificant. A good working practice is to always check your SA and trend estimates for level shifts around the SB periods. If you see one, then how do you minimise it, or better still (but near impossible) remove it? One way is to relax the assumption of equalising S*I levels across all months. There are two key points:

- The goal is to ensure that the derived seasonal factors, S = SMA[S*I], are the best possible representation of the actual S*I's. That is, such that variation in I = S*I/SMA[S*I] = T*I/TMA[T*I] is minimised. We do not need to try too hard at "smoothing" seasonal breaks [i.e. by equalising all S*I levels], we just need to ensure that the SMA[S*I] curves "fit" well, while at all times preserving the B1 total. If we are too perfect in equalising S*I levels, then nusance level shifts can be introduced to compensate for (or that attempt to undo) our "removal" (and hence "redistribution") of seasonality from the original.

- In essence, we want to redistribute total 'power' in the original such that: **(i)** the resulting seasonal factor estimates from S=SMA(S*I) are reliable and representative enough in both the "period of interest" (i.e. that contains the main break to correct) and other periods used for "B1-balancing".
(ii) no systematic level shifts are introduced.

With these in mind, here's a general method, which is essentially an extension to 'Manual Method 1':

1. It is assumed you have first used 'Manual Method 1' above to estimate and apply SB corrections, and that you do indeed see a noticeable (significant) level shift around the SB year(s).

2. From your full set of SB factors (from 'Manual Method 1'), the procedure is to remove one or more factors (i.e. set them explicitly to 1) on either side of the "main break period" of interest (June in above example).

3. Rescale these factors to ensure B1 totals get balanced after application (using either the *R* program described in Appendix I, or, by comparing B1 table totals before/after the SB corrections are applied).

4. Enter these factors in the SEASABS "Prior Corrections" GUI and run a "Research".

5. Examine the quality of the seasonal factors "S" from the S=SMA(S*I) operation, i.e. that they are representative enough of the actual S*I values given the volatility. Also, examine the strength of any level shift: is this smaller/larger than the level shift that was present when you started with step (1)? If still noticeable (significant), go to step 6.

6. Continue iterating steps (2) - (5) by resetting/removing/tweaking different combinations of SB factors (in some "justifiable" manner of course) from your initial set. This is to be done until level shifts are appropriately minimised (or removed), but at the same time, reliability in seasonal factors is maintained. The trick is, how much "redistributing of seasonality" can we perform before we start to mess up seasonal factor estimates, i.e. introduce too many internal corrections on the S*I charts.

For our simulated case above, the optimal SB factors from this iterative procedure are in Table 3. Due to the unrealistic nature of this series (i.e. noiseless), we found we had to redistributed all the excess seasonality (removed from June) into the earlier months: Jan-May. This was in order to "soften" the incidence of a level shift at the 1989/1990 boundary (graph 10/11), but at the same time maintaining reliable seasonal factor estimates (e.g. S*I chart in graph 12) and B1 totals perfectly balanced.

Seasonal Break Correction

Affecting: June 1990

	Year		Factor
January	1989	1990	0.91742
February	1989	1990	0.91742
March	1989	1990	0.91742
April	1989	1990	0.91742
May	1989	1990	0.91742
June	1989	1990	1.36200
July	1989	1990	1.00000
August	1989	1990	1.00000
September	1989	1990	1.00000
October	1989	1990	1.00000
November	1989	1990	1.00000
December	1989	1990	1.00000

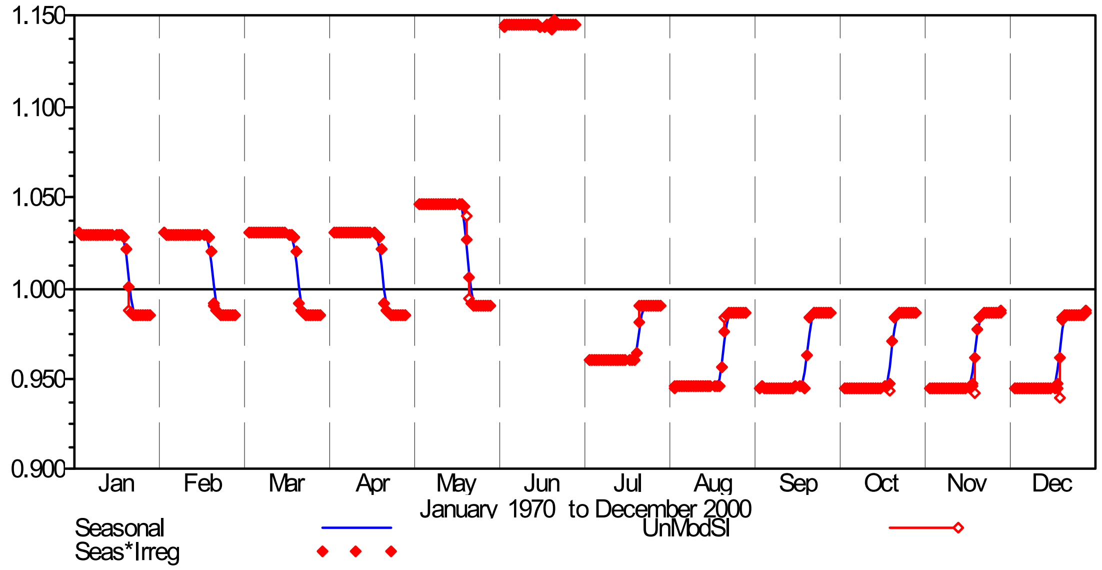
Reason: Frank's estimates from program!

Originator: Frank M   Date Entered: 2-8-2006

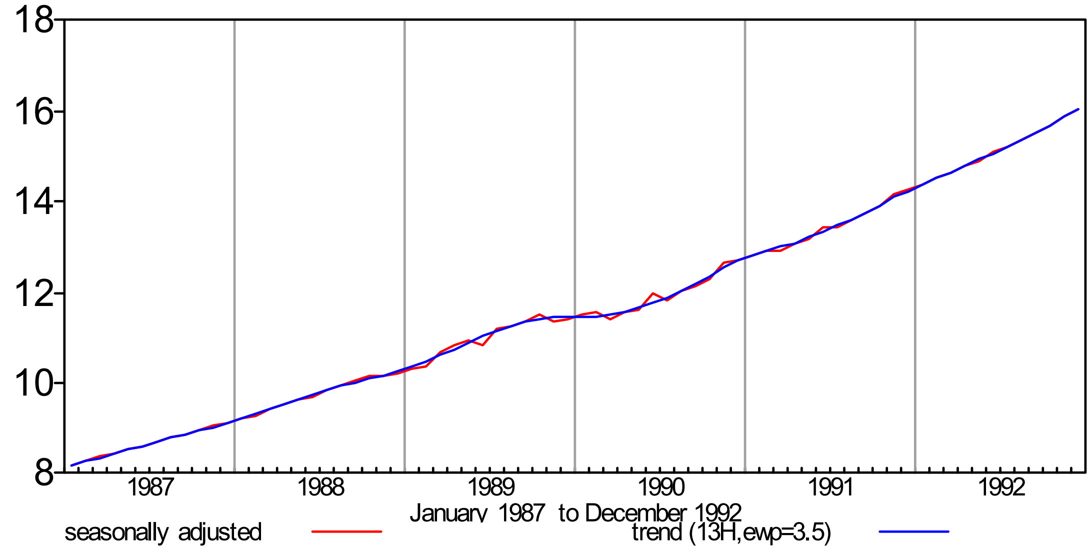
Close   Help

Table 3.

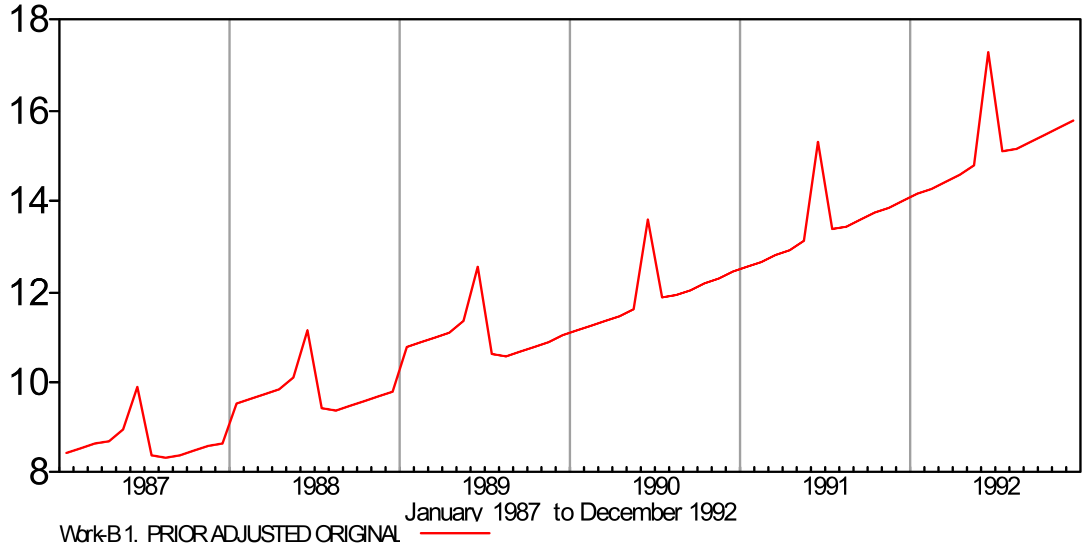
Graph 12: The S*I chart after inserting these SB corrections is as follows. Note that breaks are still allowed in the S*I chart, provided the seasonal factor estimates do not get too messed up.



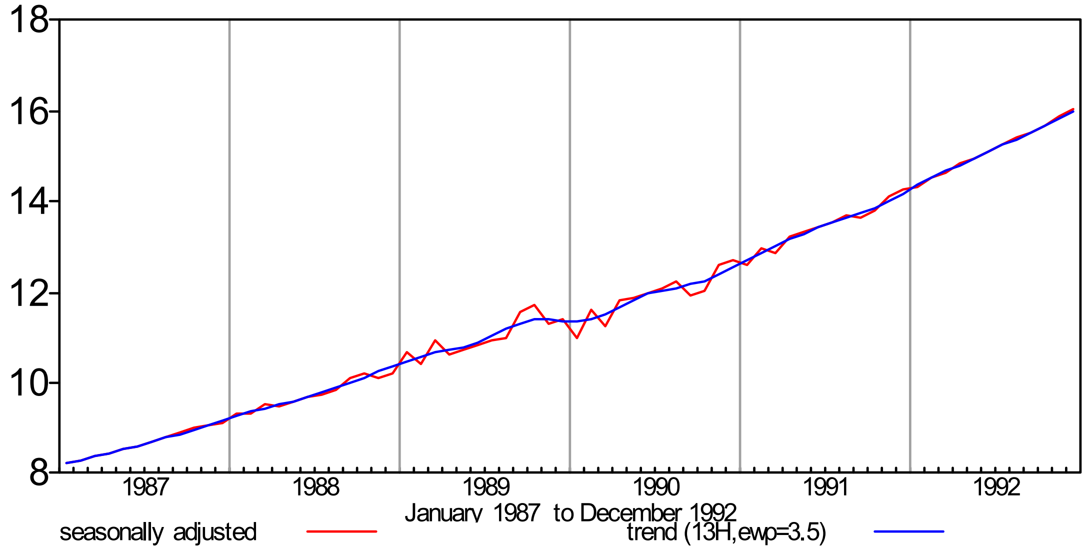
Graph 13: The following shows the SA and Trend estimates after correcting for SBs using manual method 2. Note the level shift around Dec/1989 - Jan/1990 is less apparent than that introduced by simply equalising all S*I levels (graph 10).



Graph 14: The following is the B1 (O corrected for SB priors = O/SB). The systematic level shift in B1 is very much reduced compared to that in graph 11.



Graph 15: In closing, you may have noticed that the SEASABS SB algorithm appears to have done a better job at estimating seasonal factors and minimising level shifts (e.g. compare graph 7 with 13). This is because SEASABS has the greater flexibility in allowing B1 totals to vary by up to 2% before/after SB corrections. This allows one to tweak the SB factors and locations even further than is possible by enforcing the requirement that B1 totals should remain fixed. It would be interesting to see how the SEASABS algorithm would fair if it enforced exact balancing of B1 totals. As a quick test, we took the SB factors estimated by the SEASABS algorithm (Table 1) and rescaled them such that B1 totals remain fixed after SB corrections are applied. The resulting SA and trend graphs are below. This time, SEASABS performed badly!



Appendix I: a program to ensure "balanced" B1 series

I have written a 'quick and dirty' *R* function which reads as input a set of initial (approximate) SB factors estimated from a S*I chart and a B1 series from SEASABS. It then rescales the factors such that the B1 total remains unchanged after they are applied. Note that this program performs exact balancing, i.e. such that the B1 total remains the same (within machine precision) after the final SB factors are applied. Recall that SEASABS currently allows a 2% variation in the B1 total. This does not seem methodologically correct. It may have to do with the inner workings of the automated SEASABS algorithm.

Rescaling of input (approximate) SB factors to ensure "balanced" B1 totals is performed using the following procedure. Note that this only applies under a multiplicative decomposition model. The *R* program needs to be modified to handle series described by additive models.

$B1$ totals before SBs inserted $=$ $B1$ totals after SBs inserted:

$$\sum_{period\ j}^{N} \sum_{year\ i}^{SByr(j)} B1_{ij} = \left[f_1 \sum_{year\ i=1}^{SByr(1)} B1_{i1} + \sum_{i=SByr(1)+1}^{i\max(1)} B1_{i1} \right] + \left[f_2 \sum_{year\ i=1}^{SByr(2)} B1_{i2} + \sum_{i=SByr(2)+1}^{i\max(2)} B1_{i2} \right] + \cdots + \left[f_N \sum_{year\ i=1}^{SByr(N)} B1_{iN} + \sum_{i=SByr(N)+1}^{i\max(N)} B1_{iN} \right], \qquad (1)$$

where:

$N =$ number of periods (months or qtrs in a year).

$f_j =$ approximate prior SB factor for period j (see step 2 under "Manual Method 1.." for cases).

$i\max(j) =$ the maximum year for period j.

$SByr(j) =$ the 'seasonal break' year for period j [= last year in period before which S*I level changes abruptly, i.e. break location]

To ensure that the B1 total before application of our approximate seasonal break factors = B1 total after their application, we compute the residual (either excess or deficit) in this equality:

$$D = B1(total\ before) - B1(total\ after), \qquad (2)$$

and then equally apportion this difference amongst all those periods which required a seasonal break correction. If we denote the number of periods requiring a SB correction as $N_{SB} \le N,$ then we assume that the "amount of residual B1" that needs to be thrown back into any particular period j is $\approx D / N_{SB}$. If we take any period j from the right hand side of equation (1), our goal is to determine a factor g_j such that:

$$f_j \sum_{year\ i=1}^{SByr(j)} B1_{ij} + \sum_{i=SByr(j)+1}^{i\max(j)} B1_{ij} + \frac{D}{N_{SB}} = g_j f_j \sum_{year\ i=1}^{SByr(j)} B1_{ij} + \sum_{i=SByr(j)+1}^{i\max(j)} B1_{ij}. \qquad (3)$$

I.e. g_j is the factor by which an initial f_j needs to be multiplied in order to make the B1 totals balance when g_j is computed for all periods j that require a seasonal break. Simplifying and rearranging equation (3), we have:

$$g_j = \frac{D}{N_{SB}} \left[f_j \sum_{year\ i=1}^{SByr(j)} B1_{ij} \right]^{-1} + 1. \qquad (4)$$

Therefore, each period with a SB prior gets its own g_j "fudge" factor. The final "effective" SB factors that ensure "exact" balancing of the B1 totals are then given by:

$$SB_j = g_j f_j.$$

The R function that implements the above is located in the file **"S:\R\TSAlib\compute_SBfactors.r"**. Instructions and an example on how to call this function are at the top of this file.

Appendix II: Further Work/Investigation

- Explore inner workings of SEASABS algorithm further. Past documents? A rudimentary entry exists in the SEASABS help.
- Maybe optimise/improve SEASABS algorithm: e.g. enforce 'stricter' B1 balancing; enable one to specify a SB location explicity and SEASABS takes care of the rest; better output diagnostics?
- Refine above manual methods (maybe not general enough).
- Explore X12ARIMA/regARIMA modelling for SB corrections and compare to (or use to improve) SEASABS algorithm.
- Resurrect BPG on SBs, expand on manual methods.
- Submit/update relevant SRs.