Easter Proximity Diagnostics using (reg)ARIMA Residuals

Frank Masci
Version 3.0, 2/12/2005

1. Introduction and Goals

The project goals were outlined in the proposal document: [(Subject: Project proposal: measuring Easter proximity diagnostics from regARIMA residuals; Database: Time Series Analysis WDB; Author: Frank Masci; Created: 04/11/2005; Doc Ref: FMAI-6HSVQ2)]. In a nutshell, the main goals are to devise robust methods for detecting, measuring and correcting for the Easter Proximity (EP) effect under the regARIMA framework, more specifically, using ARIMA model residuals for its detection and regARIMA modelling for its correction. A brief background and summary follows.

The current method for correcting the Easter Proximity (EP) effect in SEASABS (as well as a number of other calendar related effects) involves using the final X11 irregulars from the D13 table. The EP effect can cause the final irregulars to systematically deviate from one, the neutral line of irregulars. The irregulars are regressed on a model parameterized in terms of the fraction of "before" and "during" Easter holiday days that fall in March. A set of coefficients for these two periods are computed and tested for non-zero significance to ascertain whether an EP effect was present. The major drawback of using the D13 irregulars for EP diagnostic tests is the following:

- They have been distorted by the iterative seasonal adjustment process that may or may not be related to Easter.
- The irregulars have not been corrected for outliers prior to testing for the effect. Outliers may lead one to accept an EP effect which would otherwise not be significant if they were appropriately accounted for.
- Statistical tests based on the X11-D13 irregulars are not robust enough to discriminate between regressor models under the regARIMA modelling framework. This makes "regressor fine-tuning" exceedingly difficult.

In lieu of the problems above, we plan to exclusively use ARIMA (and regARIMA) modelling to diagnose and correct for systematics introduced by the EP effect. In particular, we will:

- provide more powerful statistical tests for the presence of an EP effect.
- provide robust criteria on selecting the best EP regressor under regARIMA for any particular series with an EP signature.
- explore the sensitivity from outliers on the significance of an EP effect.
- explore the coupling between ARIMA-specific and regressor-specific parameters and optimise their selection in the fitting procedure. The sensitivity of regARIMA residuals will allow better separation of dependencies on the full parameter space.
- incorporate these improvements in the X12-ARIMA/SEASABS software interface so they can be used in production work by TSA.
- implement an infrastructure for storing EP (and other) regressors in

production/analysis software for flexibility in their selection and optimal usage.

Results will be continuously appended to this document as research proceeds.

2. Nature of the ARIMA Residuals

It would be interesting to compare the distribution of D13 irregulars from the X11 or X12 procedure to that of ARIMA best-model residuals. What is the underlying distribution of these components? Figures 1 and 2 compare the magnitude of the (X12) D13 irregulars with the actual ARIMA best-model residuals *containing no corrections for the EP effect* for two retail series. The neutral lines (mean values) for each distribution are one. The red diagonal lines are lines of equality, with slopes of unity.

The ARIMA model residuals reported by the X12-ARIMA program are on a log scale, defined by:

$$r = \log y_{obs} - \log y_{\text{mod}} = \log\left[\frac{y_{obs}}{y_{\text{mod}}}\right], \qquad (1)$$

where y_{obs} and y_{mod} are the (original) observed and predicted ARIMA-model values respectively. When standardised to have unit-mean as applicable to a multiplicative model, the observed/model residuals can be written:

$$R = \exp[r]. \qquad (2)$$

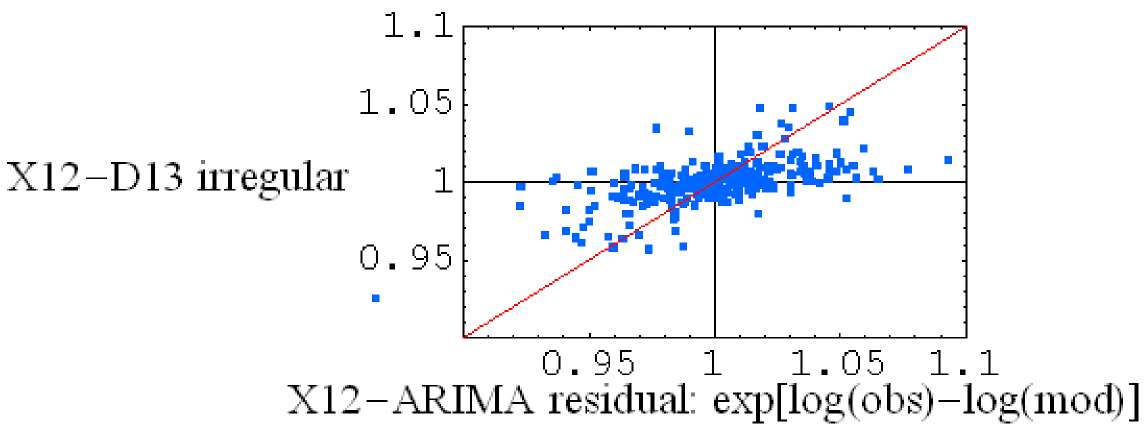


Figure 1: D13 irregulars vs. ARIMA residuals for "liquor state aggregate retail series".

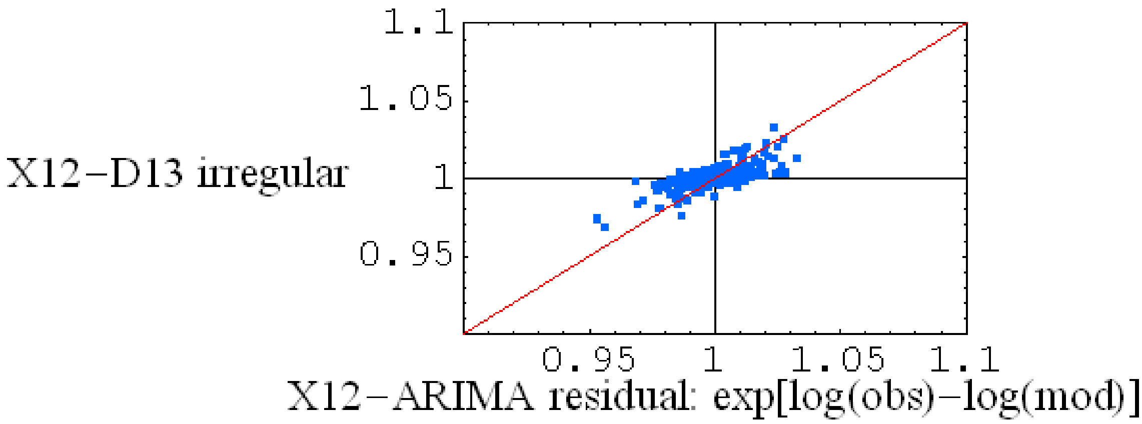


Figure 2: D13 irregulars vs. ARIMA residuals for "supermarket/grocery state aggregate retail series".

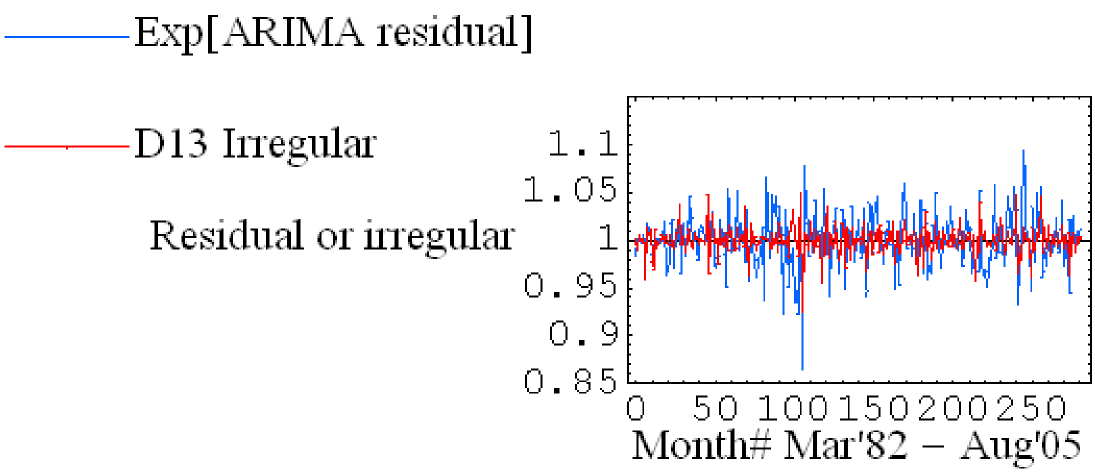


Figure 3: ARIMA residuals or D13 irregulars vs. time for "liquor retail series".

<u>From the above plots, we note the following:</u>

- How do we explain slopes of non-unity in Figures 1 and 2, i.e., D13 irregulars not being equal to ARIMA residuals (on the standardised unit scale)? As shown in Figures 3 and 6, this is just a consequence of the ARIMA residuals having a much larger variance. This leaves one to explain why the ARIMA residuals are larger on the scales represented in Figures 1 and 2. The most likely reason is the following:

- The ARIMA model chosen here is not perfect so as to capture all possible systematics, i.e., the residuals are not completely random (note that no regressors were included to correct for EP effects here). Perfectly random residuals may be more difficult to attain in an ARIMA model since it is somewhat limited in what it can model - e.g., outliers and time-varying structure such as seasonality. If these are unaccounted for, their signature would show up in the residuals and thus inflate their variance.

- Note that the D13 irregulars from the X11 process represent a completely different "flavour" of residual than the ARIMA residuals. The ARIMA residuals are differences between the observed data and the best fitting ARIMA model being considered, while the D13 irregulars are effectively residuals of the seasonally adjusted original data from the final trend cycle (usually a Henderson moving average). So in theory, if an ARIMA model captures all possible systematics (including seasonality), it should result in small white-noise residuals. The D13 irregulars however depend strongly on estimates of the final trend cycle through $\hat{I} = T * I / \hat{T}$ (for multiplicative models) or $\hat{I} = T + I - \hat{T}$ (for additive models).

- The D13 irregulars, as estimators of residuals in a seasonally adjusted series from its underlying trend, could actually be biased if the trend cycle traces the underlying noise structure (correlated or not) and other residual systematic patterns (e.g., seaonality and EP effects). In this case the residuals will be underestimated. This is illustrated in Figure 4 where we simulated a series using a non-seasonal ARIMA(0, 2, 1) (pure MA) model with known input variance and a linear trend:

$$(1-B)^2 y_t = \varepsilon_t + \theta\varepsilon_{t-1}\ ; \qquad \varepsilon_t \sim N(0,\sigma^2); \qquad \sigma^2 = 36; \qquad \theta = 0.5$$

$$y_t = at + b \quad \text{(linear trend)} \qquad (a = 1,\ b = 100)$$

$$\text{effective variance of process (ACVF at lag 0):} \qquad \gamma(0)=\sigma^2\left[1+\theta^2\right] = 45$$

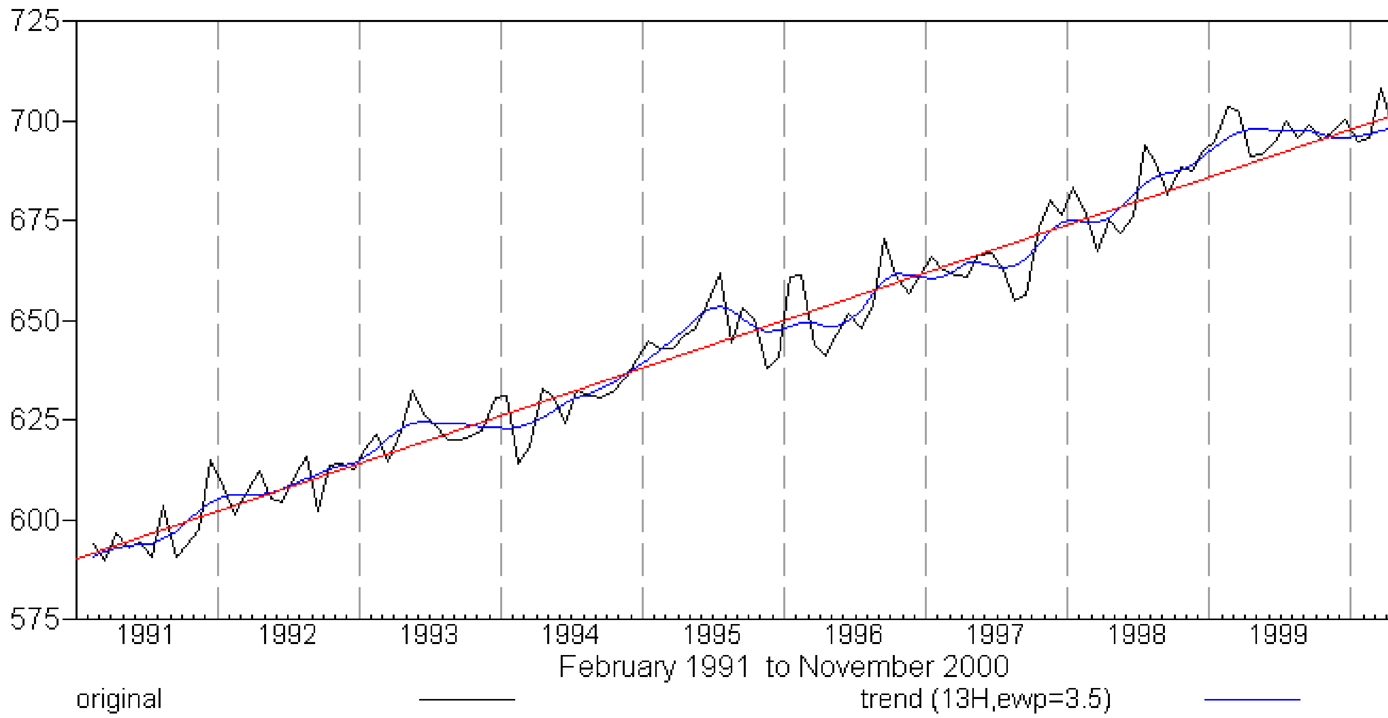


Figure 4: Simulated series (black line) using the correlated-noise model above. The red line is the true (simulated) trend, the blue line is the X11-derived trend from a 13 term Henderson moving average. Note that the D13 irregulars ("residuals") derived from the black (original seasonally adjusted series) and the blue X11 trend will be underestimated and misrepresent the underlying process variance. The expected variance from this process is 45, while the variance in D13 irregulars is ~ 20.

- The main conclusion is that usage of D13 irregulars to diagnose systematic calendar-related effects (as residuals from underlying X11 trend cycles) is not robust. This is a general problem inherent with the Henderson filters. The X11 derived trends are likely to masquerade the real underlying noise structure and systematic patterns sought for.

- ARIMA model residuals appear to be more sensitive at detecting systematic, non-normal effects in data. Their larger variances relative to D13 irregulars for the above two series is consistent with the models not capturing all possible systematics. This is reflected by the non-Gaussian nature of the ARIMA residuals in Figure 5.

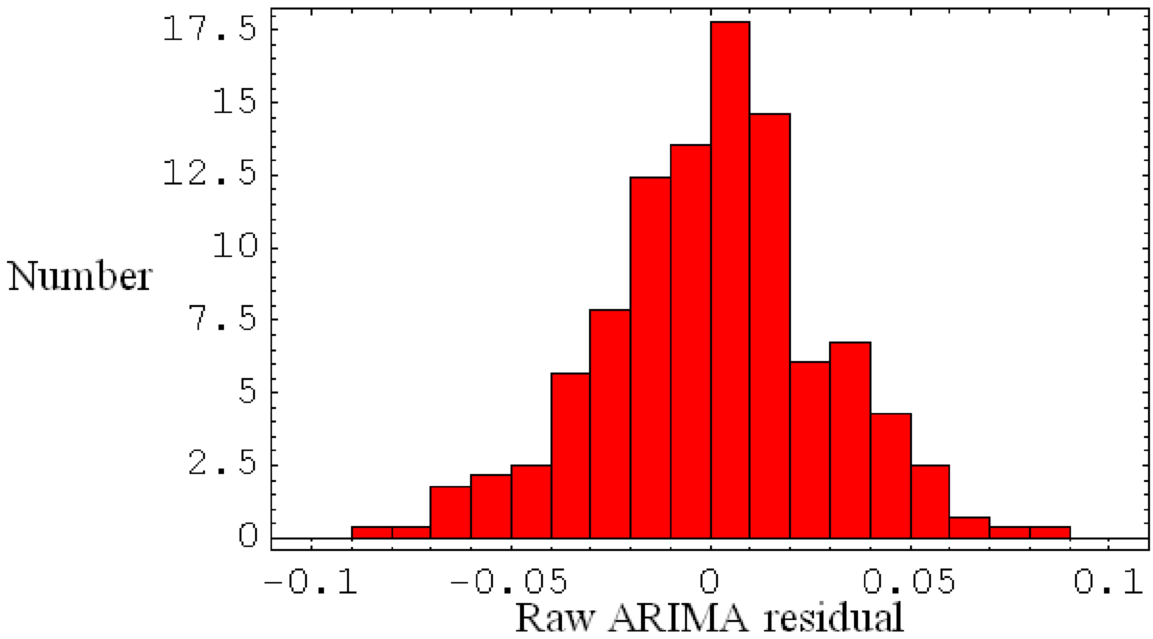


Figure 5: Raw ARIMA residuals (equation 1) for "liquor retail series". Theoretically should be Gaussian, but kurtosis is ~4.7. This indicates that systematics have not been fully captured by the model.

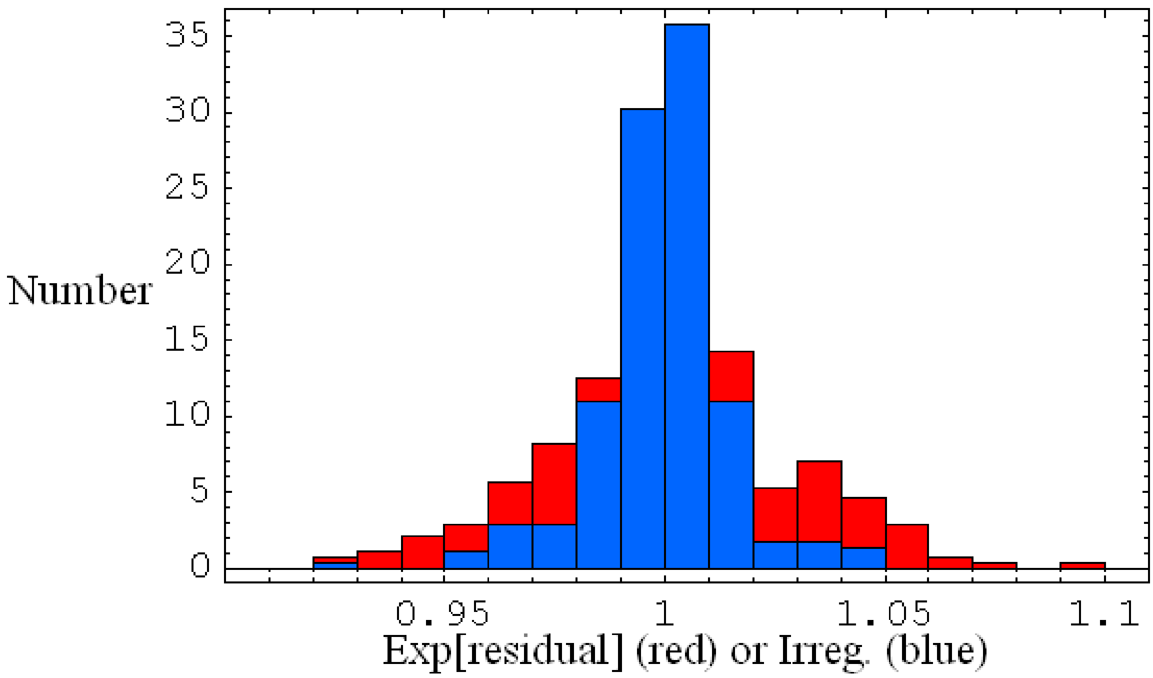


Figure 6: Histograms of Exp[ARIMA residuals] (red; equation 2) and D13-irregulars (blue) for "liquor retail series".

By hypothesis, the ARIMA residuals *r* (equation 1) for a "good model fit" are expected to be normally distributed:

~ N(0, var). It is no surprise that the underlying distribution function of the exponential of a normally distributed quantity (R in equation 2) is the standard lognormal distribution. In other words, a random variable is lognormally distributed if its logarithm is normally distributed. The lognormal distribution is defined as:

$$P(R) = \frac{1}{R\sqrt{2\pi\sigma_r^2}} \exp\left[-\frac{(\log R - \mu_r)^2}{2\sigma_r^2}\right]; \quad R > 0, \qquad (3)$$

where σ_r and μ_r pertain to the normally distributed quantity $r = \log R$.

The mean and variance of R when distributed as $P(R)$ are respectively:

$$E(R) = \exp\left[\mu_r + \frac{\sigma_r^2}{2}\right]; \qquad \mathrm{var}(R) = \left(\exp\left[\sigma_r^2\right] - 1\right)\exp\left[2\mu_r + \sigma_r^2\right]$$

If we assume $\mu_r = 0$ (e.g. Fig. 4), it is interesting to note that in the limit $\sigma_r \to 0: R \sim 1 \Rightarrow \log R \sim R - 1$ and he

$\sigma_r^2 \sim \mathrm{var}(R)$. This implies that the lognormal distribution (eq. 3) approaches a normal distribution with $E(R)$

$$P(R) \sim \frac{1}{\sqrt{2\pi\sigma_r^2}} \exp\left[-\frac{(R-1)^2}{2\sigma_r^2}\right]; \quad -\infty < R < \infty \qquad (4)$$

This last statement justfies the use of Gaussian (parametric) test statistics on the D13 irregulars in TSA production work to explore the signifance of EP (and other systematic) effects. It's important to note that this is only true in the limit of small variance in the residuals and/or irregular component. The degree to which the normal distribution approximates the lognormal distribution is quantified in Figure 7.

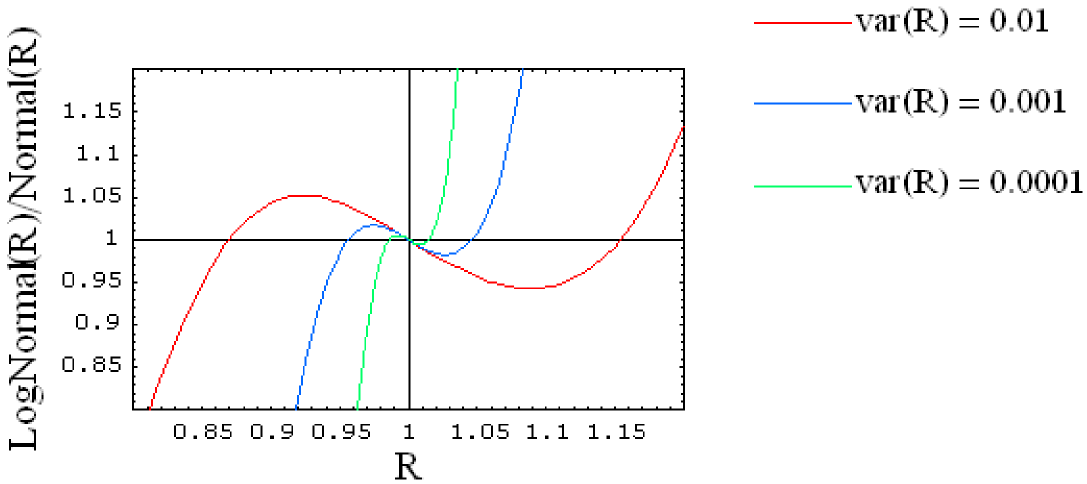


Figure 7: Ratio of *LogNormal* (eq. 3) to *Normal* (eq. 4) approximation prediction as a function of the residual parameter R for three variances. Ratios close to 1 indicate the normal approximation is a good one. For each of these variances, only those regions which fall within ~2 standard deviations of the mean <R>=1 are those which matter in

practice. For each variance, these ranges are approximately:

$$0.8 \le R_{\text{var}=0.01} \le 1.2$$

$$0.94 \le R_{\text{var}=0.001} \le 1.06$$

$$0.98 \le R_{\text{var}=0.0001} \le 1.02$$

The largest discrepancies from the normal approximation are therefore for the largest variances. Conservatively speaking, we usually see variances no larger than 0.001 (sd ~0.03), which corresponds to discrepancies no larger than 5% (blue curve) over the R values of interest.

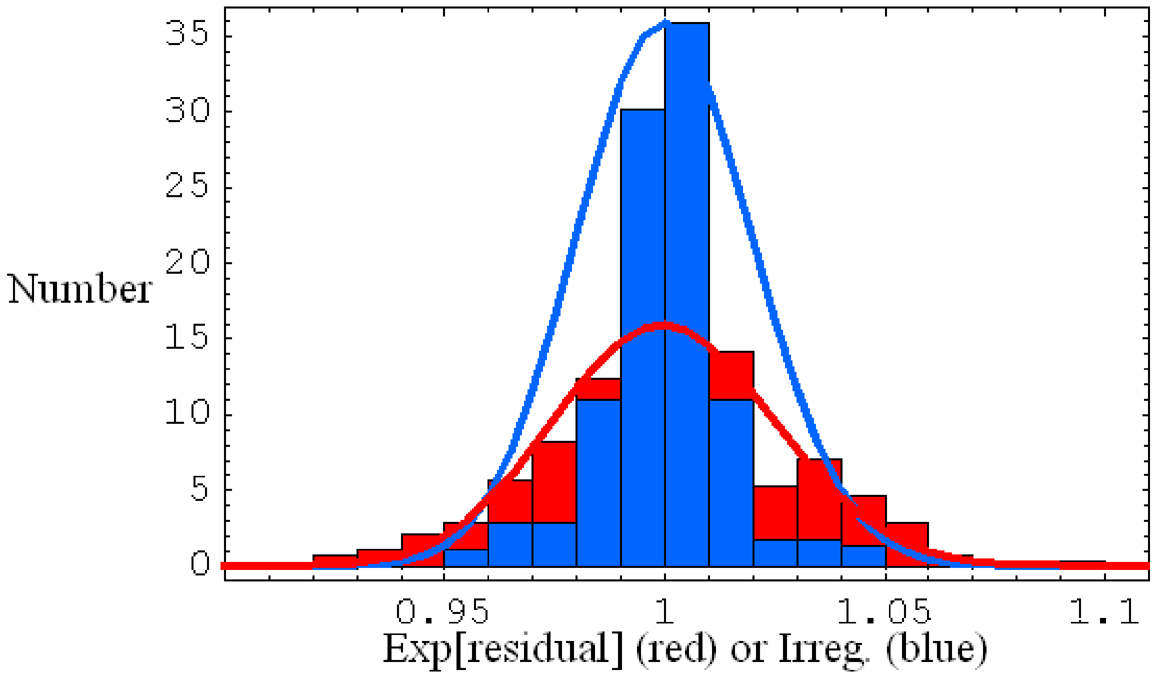


Figure 8: Same as Figure 6 overlayed with best fitting Lognormal distributions according to observed variance of the data (eqation 3). Normal distributions fit equally as well.

The main conclusions of this section are:

- Usage of D13 irregulars to diagnose systematic calendar-related effects (as residuals from underlying X11 trend cycles) is not robust. This is a general problem associated with the Henderson filters. The X11 derived trends are likely to masquerade the real underlying noise structure and systematic patterns sought for. ARIMA model residuals appear to be more sensitive at detecting systematic, non-normal effects in data.

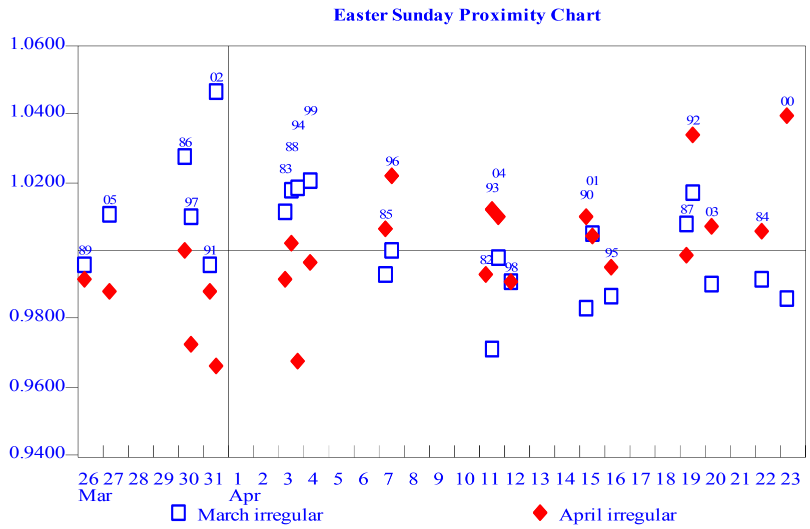
- The standardised ARIMA residuals (on the unit-neutralised scale) can be explained by Gaussian statistics on average, and that parametric tests based on this assumption can be safely applied to explore the significance of EP (and other systematic) effects. This however assumes that the ARIMA model fit has a high level of significance in representing the underlying noise structure and systematics. This is guranteed if it meets the criteria assessed by the X12-ARIMA program.

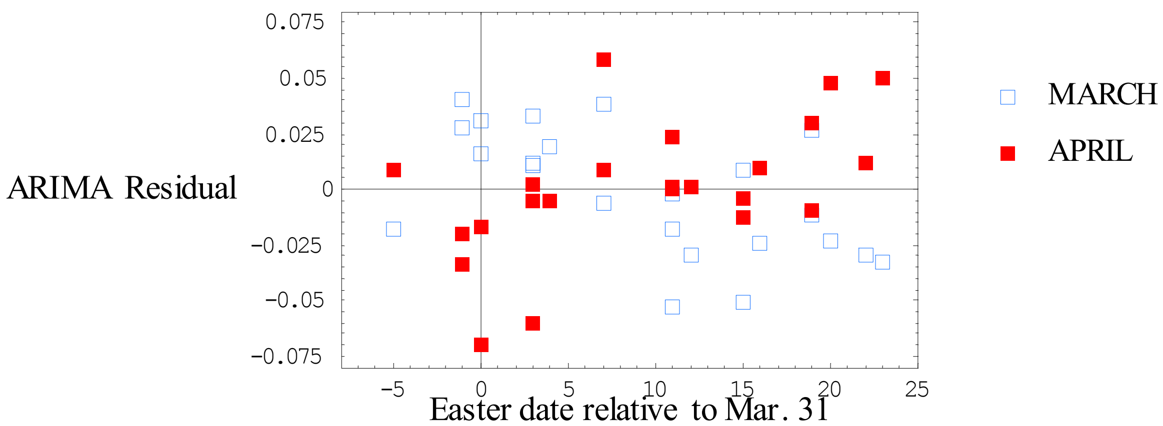
3. Easter Proximity Charts (ARIMA Residuals versus D13 Irregulars)

<u>EP Charts without correction</u>

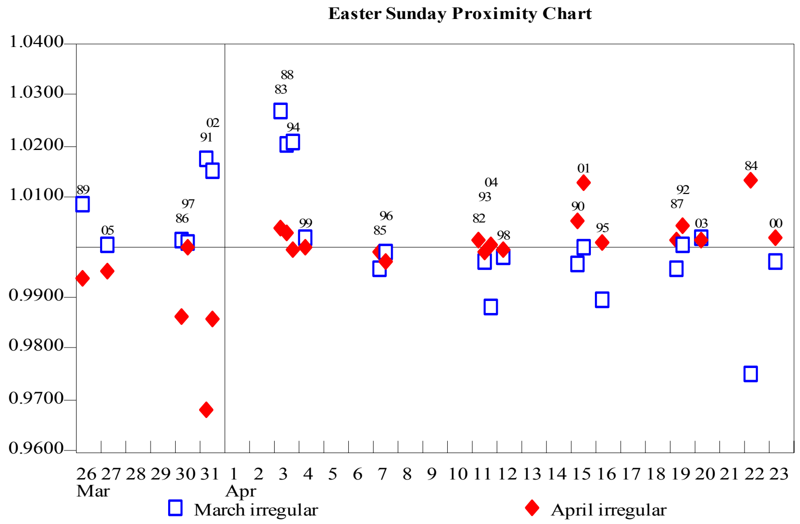
Figure 9: EP charts without the EP correction in terms of D13 irregulars (top plots) and raw-ARIMA residuals (bottom plots) with no EP regressor included. The latter are based on the "best" ARIMA model applied to prior-adjusted (B1) series.

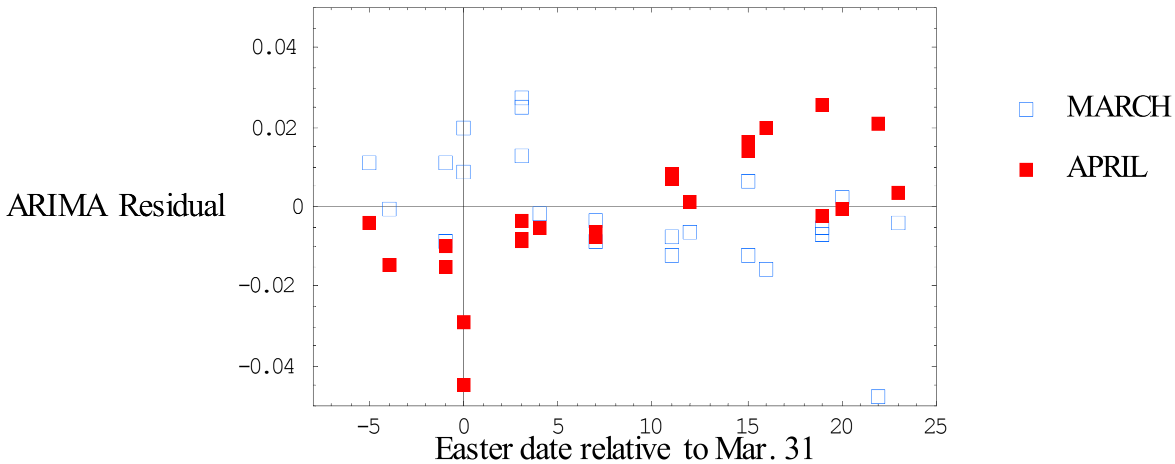
<u>Liquor retailing:</u>



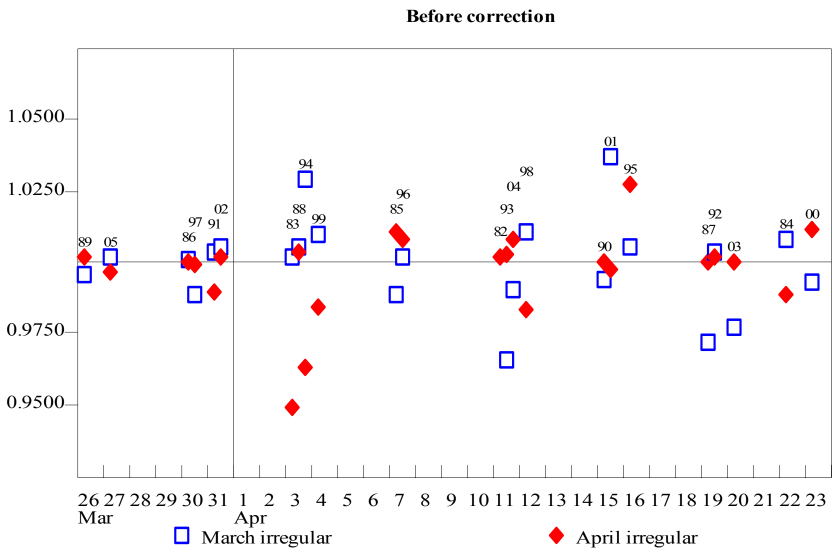


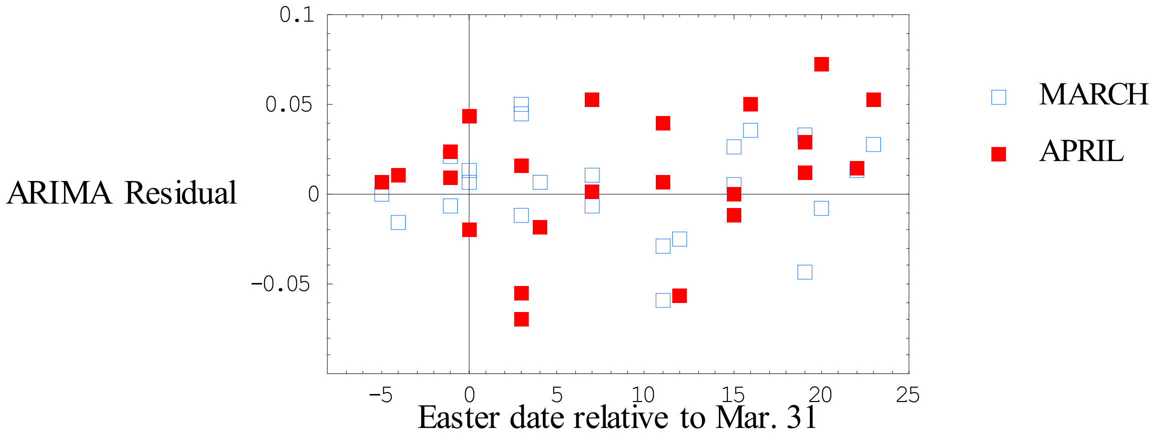
<u>SuperMarket/Grocery retailing:</u>





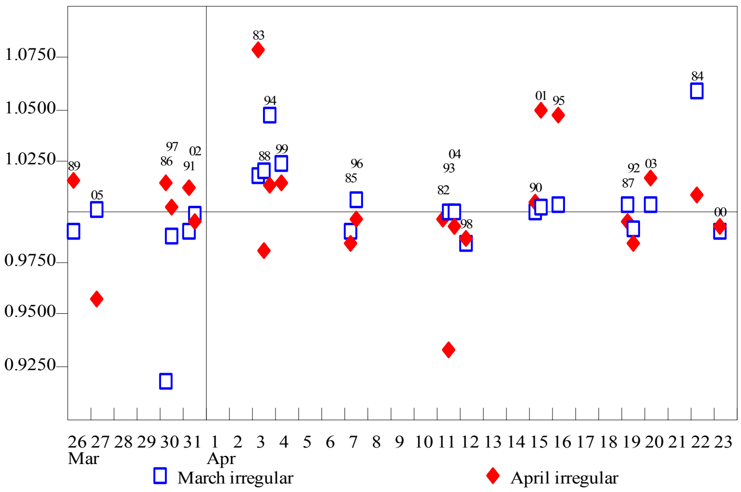
Takeaway-food retailing:

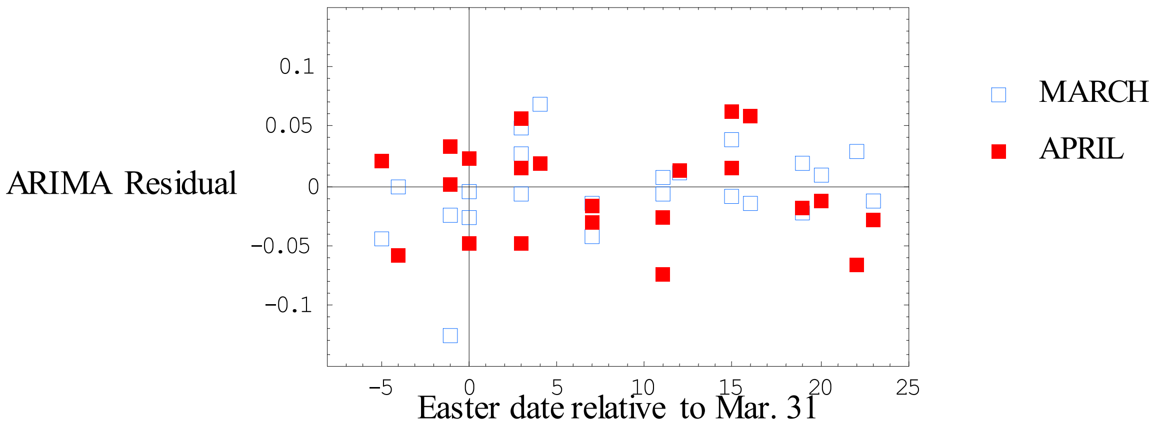




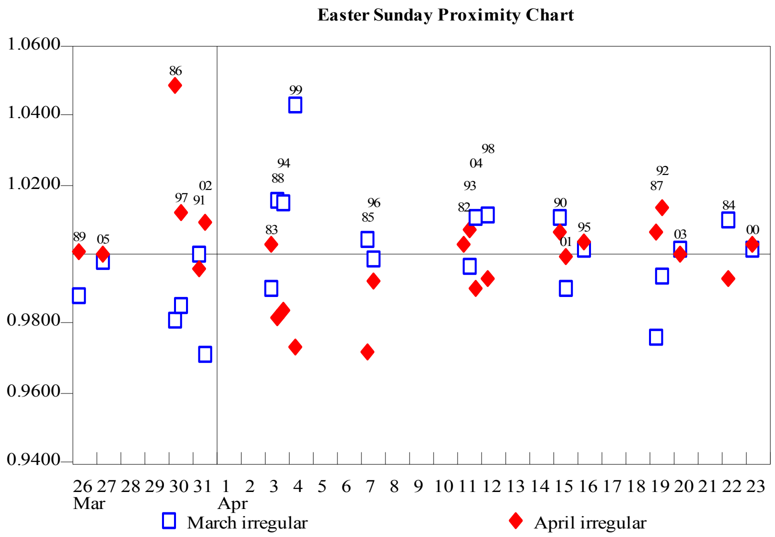
Clothing retail:

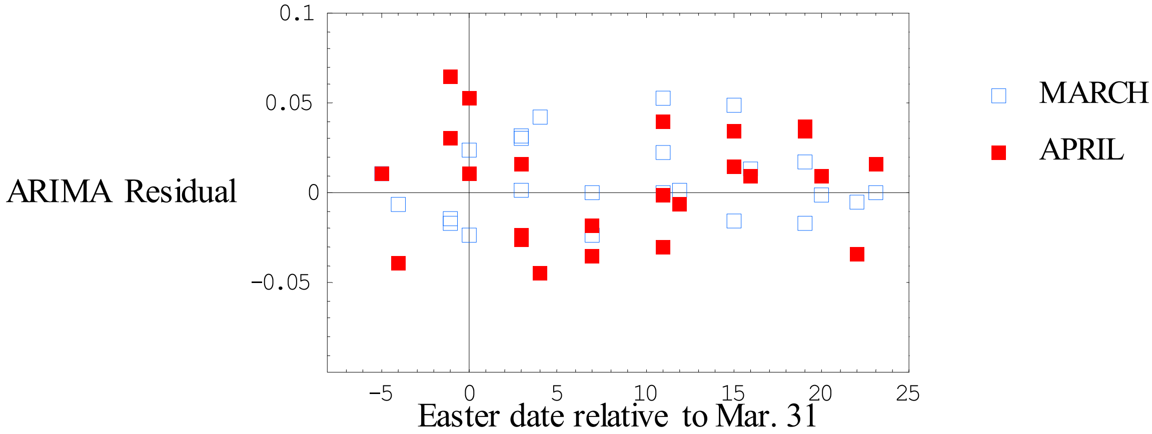
Easter Sunday Proximity Chart



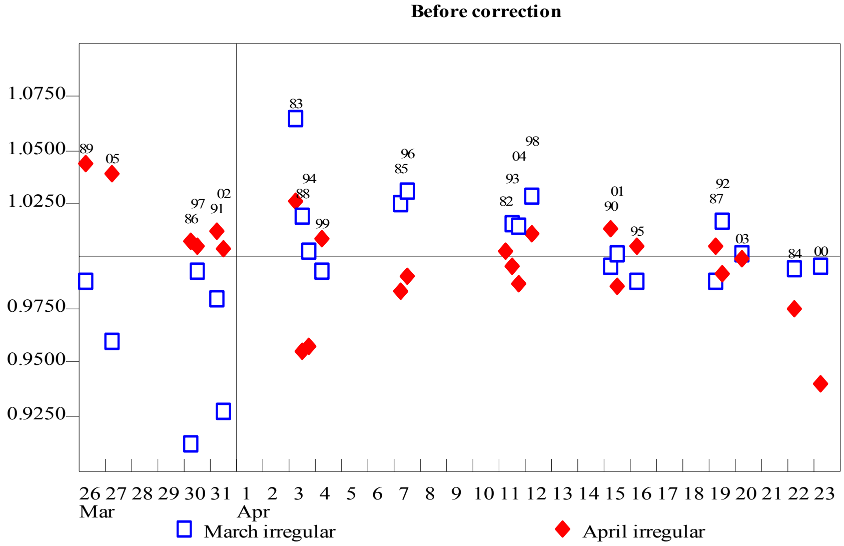


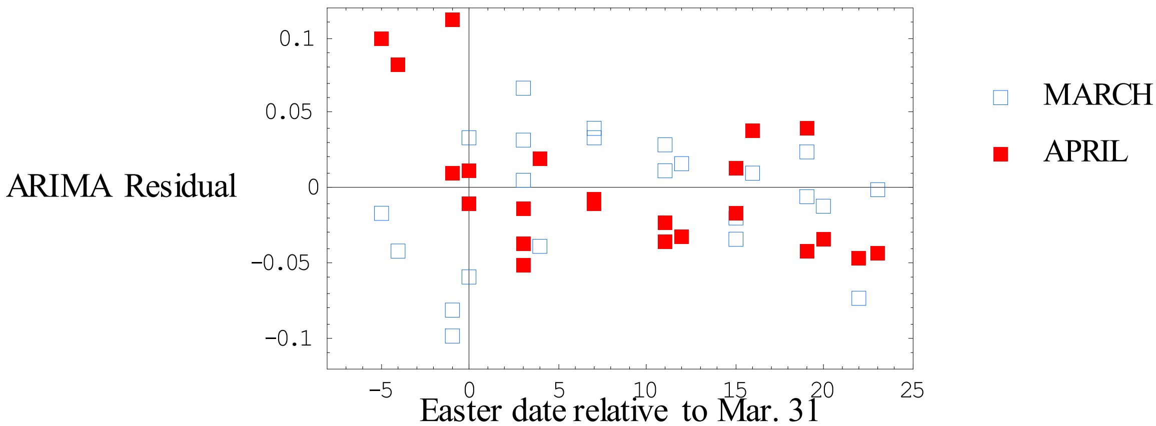
Pharmaceutical, cosmetics and toiletry retail:



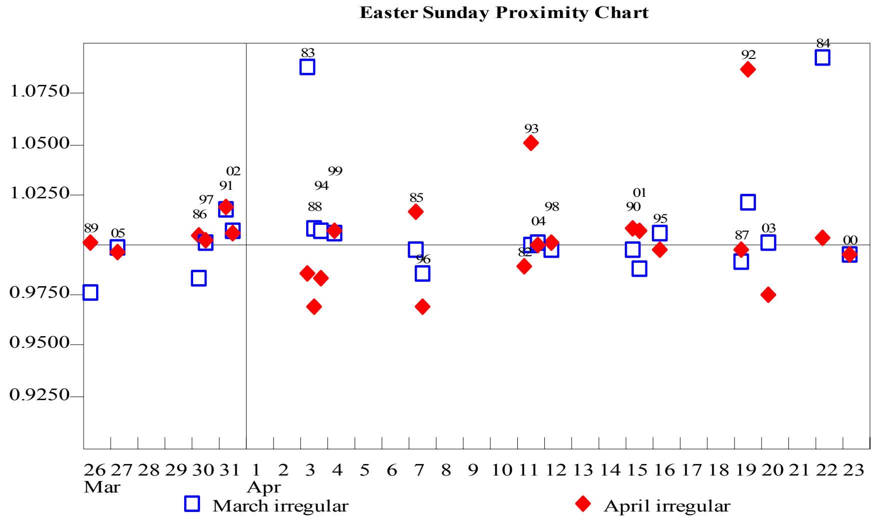


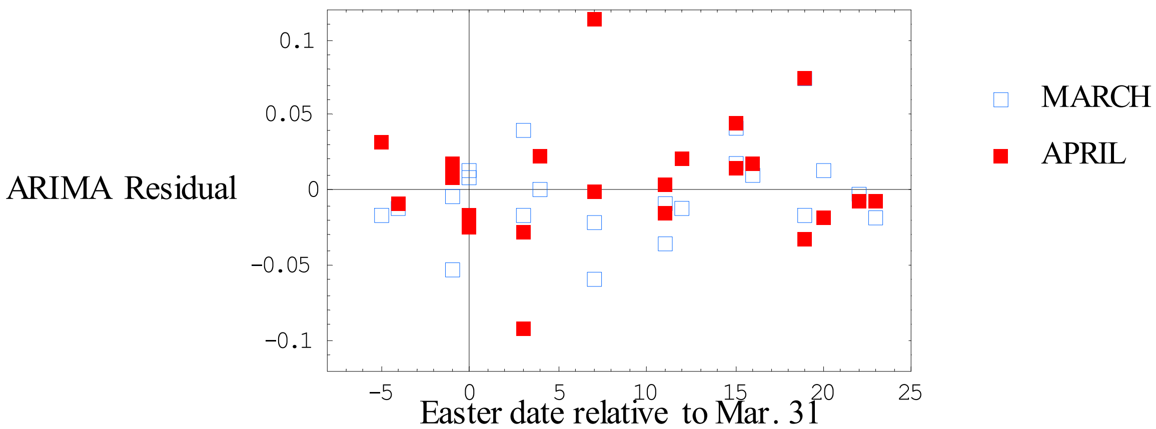
Domestic appliances and recorded music retailing:



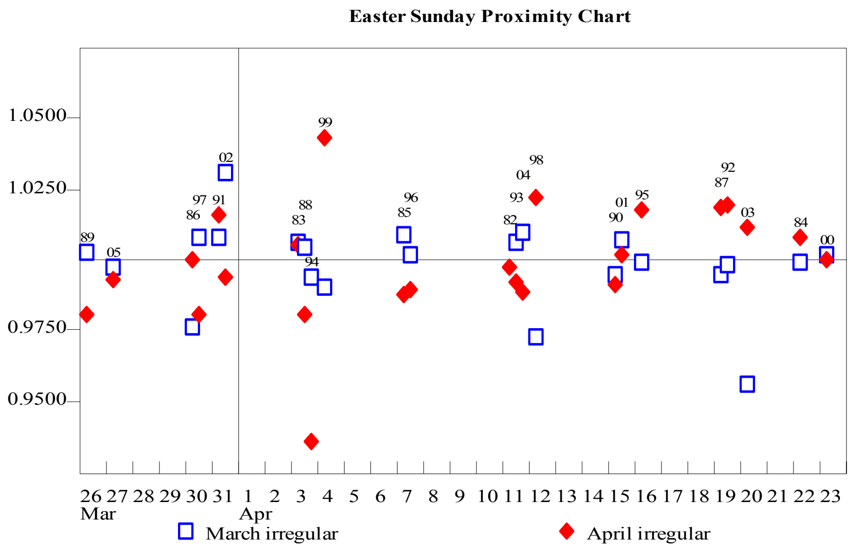


Hardware/Houseware retailing:





Other retailing:



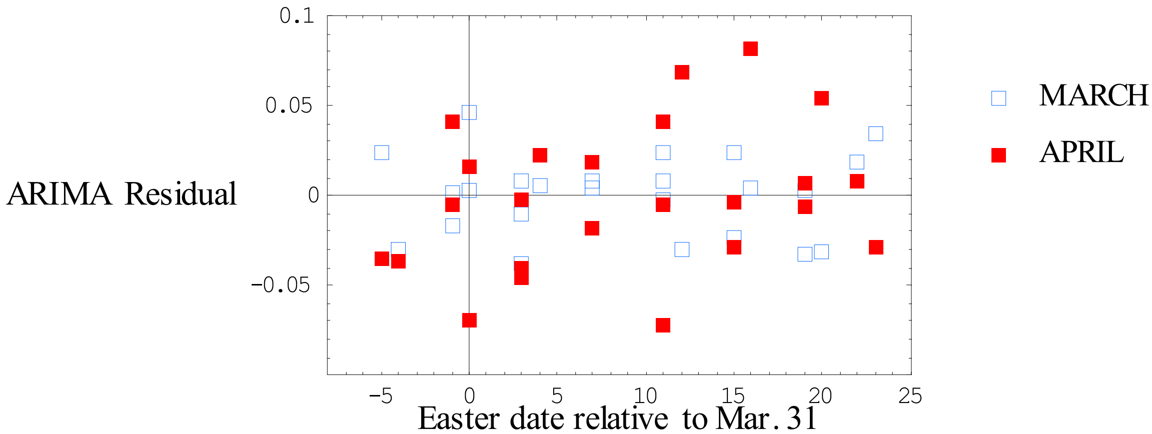


Figure 9: EP charts without the EP correction in terms of D13 irregulars (top plots) and raw-ARIMA residuals (bottom plots) with no EP regressor included. The latter are based on the "best" ARIMA model applied to prior-adjusted (B1) series.

Some points to note from Figure 9:

- The pattern in March and April ARIMA residuals are broadly consistent with corresponding D13 irregulars, as expected.
- This implies that ARIMA residual charts are just as robust at detecting EP effects.
- There are no cases where there is evidence for an Easter effect in the ARIMA residual charts but not the SEASABS D13 charts.

4. Easter Proximity Detection in the Presence of Outliers (version 1)

Before choosing an appropriate EP regressor for regARIMA, we need to ensure that an EP effect actually exists.

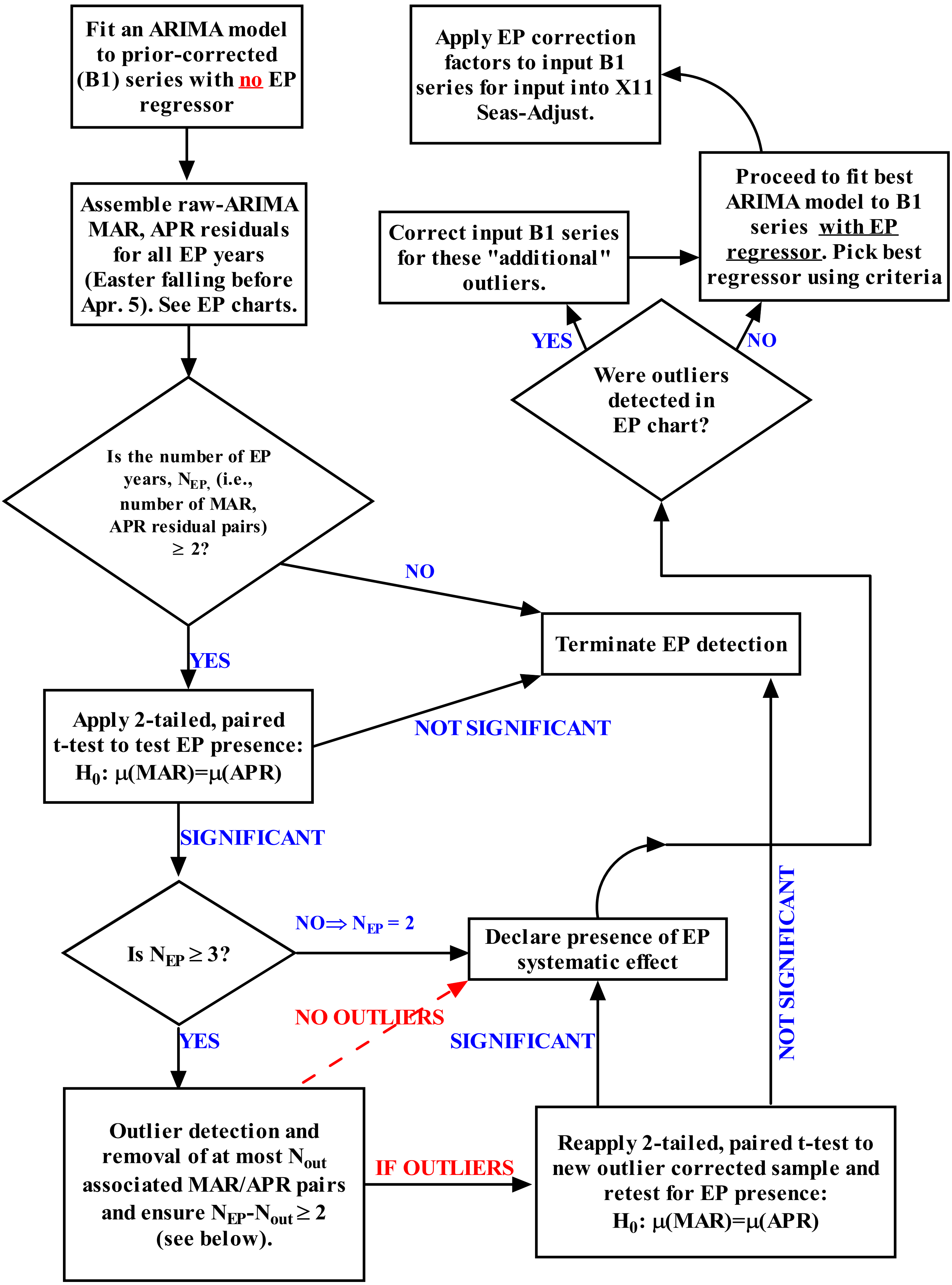
Also, we need to account for the effect of outliers on the significance of an EP effect. Their presence may lead one to erraneously accept an EP effect.

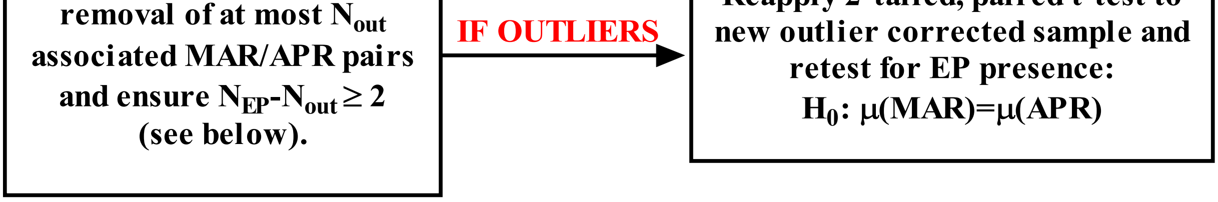
We propose the following EP detection and downstream processing strategy. Some of the steps are further discussed after the flowchart. We present two methods depending on the nature of the input series:

Method 1: Input series is "prior-corrected" B1 series from SEASABS:

This method uses a sequential approach to detect and correct for the EP and possibly other systematic effects. It assumes an input series that has been "prior-corrected" using the SEASABS empirical algorithms: e.g: large global extremes or outliers (as opposed to outliers detected in the second pass that may specifically skew the EP effect), trend breaks, seasonal breaks, and optionally trading day and other moving holidays.

Note that this method is not restricted to only correcting for the EP effect in the regARIMA step since one has the flexibility of omitting any prior-corrections in SEASABS and including them together with the EP regressor (i.e., using a "multiregressor") later on. For instance, one can omit trading day corrections in the input B1 series, but then include them using a generalized regressor in the regARIMA step. This allows one to "mix" SEASABS prior-corrections with regARIMA corrections and is particularly useful for cases when one does not know how to formulate everything in terms of regressors.
A more generalised version of this is encompassed by method 2 below.


Fit an ARIMA model to prior-corrected (B1) series with no EP regressor
Assemble raw-ARIMA MAR, APR residuals for all EP years (Easter falling before Apr. 5). See EP charts.
Is the number of EP years, N_{EP}, (i.e., number of MAR, APR residual pairs) ≥ 2?
NO
YES
Terminate EP detection
Apply 2-tailed, paired t-test to test EP presence: H_0: μ(MAR)=μ(APR)
NOT SIGNIFICANT
SIGNIFICANT
Is $N_{EP} \geq 3$?
NO⇒ N_{EP} = 2
Declare presence of EP systematic effect
YES
NO OUTLIERS
SIGNIFICANT
Outlier detection and removal of at most N_{out} associated MAR/APR pairs and ensure N_{EP}-$N_{out} \geq 2$ (see below).
IF OUTLIERS
Reapply 2-tailed, paired t-test to new outlier corrected sample and retest for EP presence: H_0: μ(MAR)=μ(APR)
NOT SIGNIFICANT
Were outliers detected in EP chart?
YES
NO
Correct input B1 series for these "additional" outliers.
Proceed to fit best ARIMA model to B1 series with EP regressor. Pick best regressor using criteria
Apply EP correction factors to input B1 series for input into X11 Seas-Adjust.



Notes:

Minimum number of Easter Proximity March/April Pairs?

This is certainly an open question. This number can be a variable and input by the user. The above flow chart assumes a starting value of 2 EP years, that is 2 Mar/Apr pairs or 4 months in total before attempting to diagnose an EP effect. As a good working measure (as seen from the EP charts for retail series), we assume that an EP year represents a year where Easter Sunday falls on or before April 5. This date is very rubbery and a later date will make little difference for testing the existence of an EP effect. In other words, the EP effect is seen to be strongest when Easter falls close to the Mar/Apr boundary. Given a nominal pre-Easter window of w ~ 7-10 days, the assumption of a later Easter date (well into April) will wash out separations in the Mar/April residuals due to the EP effect.

Easter Proximity Detection and Significance

We apply a two-tailed, paired-sample t-test to test the null hypothesis of no significant difference in the mean values of the Mar and April residuals over all EP years. It is two tailed since the mean March residual can be greater than the mean April residual and vice versa. We use the paired version of this test since the March and April residuals are often (or expected to be) correlated. Thus for any EP year, there is an associated pair of residuals: ($\mathbf{R}_{MAR}$, $\mathbf{R}_{APR}$) which depend on each other according to the date of Easter. If $\mathbf{N}_{EP}$ is the number of associated Mar, April pairs (or EP years), then the paired sample t-test formulas are:

$$\mathrm{Cov}(R_{MAR}, R_{APR}) \equiv \frac{1}{N_{EP}-1}\sum_{i=1}^{N_{EP}}(R_{MAR\_i}-\mu_{MAR})(R_{APR\_i}-\mu_{APR})$$

$$\sigma_R = \left[\frac{\sigma^2_{R\_MAR}+\sigma^2_{R\_APR}-2\mathrm{Cov}(R_{MAR}, R_{APR})}{N_{EP}}\right]^{1/2}$$

$$t = \frac{\mu_{MAR}-\mu_{APR}}{\sigma_R}$$

The significance of "t" at some α level is evaluated for $N_{EP}-1$ degrees of freedom

Outlier Detection?

There are a plethora of outlier detection algorithms out there and we need to select a robust approach that is applicable to moderately small samples. For example, we expect to have N <~ 36 EP Mar+Apr data points, that is, there are no more than 18 EP years between 1955-2005 which is the maximum span covered by ABS series. In particular, the majority of our retail series go back to ~1980, which only encompasses about 10 EP years.

We present two outlier detection/correction methods. These have been proven to be robust for small to moderately small samples. The second method below may be useful for small samples. Either way, I believe the first method is most robust .

(i) The Median Z-score test

This method is very generic, simple and uses non-parametric estimators which are relatively insensitive to the presence of outliers. It is based on the following steps:

- Compute the "Median Absolute Deviation" from the median (MAD) for the entire data set of residuals $\{R_i\}$:

  $$MAD = med\{|R_i - R_{med}|\},$$

  where $R_{med} = med\{R_i\}$

- Next, we compute the "z-score" for for each residual:

  $$Z_i = [R_i - R_{med}] / (1.4826*MAD)$$

  Why the factor 1.4826? This is a correction factor included for completeness to ensure that the denominator:
  1.4826*MAD = σ in the asymptotic limit of a normal distribution, i.e., in the large sample limit. For your interest, the
  MAD estimator represents the 75th percentile in a normal distribution.

- We then declare a particular residual value R_i to be an outlier if the absolute value of its Z-score satisfies:

  $$|Z_i| > k \;,$$

  where **k** is some predetermined threshold (say typically 2.5 or 3.0). Note that **k** here can be taken to represent the number of standard deviations, σ, for a sample which is distributed as $\sim N(R_{med}, \sigma^2)$. The associated Mar/Apr pair of residuals containing the suspect outlier R_i are removed from the sample before proceeding.

(ii) A Jacknife "survival" approach

This method may be preferred for the smallest samples, e.g., for N <~ 8 (4 EP years). It implicitly assumes that a single suspect outlier is present at either end of the residual distribution. It is based on the following steps:

- apply a t-test to test for the significance of the EP effect: H_0: μ**(MAR) =** μ**(APR)**. If this test results in " μ**(MAR)** not equal to μ**(APR)** ", we move on to the next step, otherwise we stop here. This is illustrated in the above flowchart.

- find the maximum of the absolute values of the MAR and APR residuals: $\mathbf{R_{max} = max\{|R_i|\}}$.

- assume this is a potential, suspect outlier and remove the associated MAR/APR residual pair containing this point.

- With this suspect outlier removed, reapply the paired-sample t-test to test for the significance of the EP effect:
  $\mathbf{H_0: \mu(MAR) = \mu(APR)}$.

- If this second t-test "survives" in maintaining a significant " μ**(MAR)** not equal to. μ**(APR)** ", then we declare that an EP effect genuinely exists.

Therefore, this approach really does not directly detect outliers. An outlier may, or may not be present. Regardless of whether the maximum value pair was an outlier or not, its removal from the sample just makes the test for the presence of an EP effect more stringent, and hence more significant if it "survives" the second-pass t-test.

Selecting the "Best" EP regressor under regARIMA.

I envisage that we'll first want to isolate a robust set of EP regressors that cover most real series cases (e.g., from retail to tourism). Given this list, we would then cycle through each regressor sequentially in separate regARIMA model runs and select the "best" regressor according to a set of significance criteria. The X12-ARIMA program generates regression test statistics such as the t and χ^2 values for each estimated regressor parameter. These can be used to eliminate nonsignificant regressors from our list. Furthermore, the global AICC statistic can also aid in isolating the best regARIMA model. My belief is that a combination of all these criteria will need to be used.

Method 2: "Super-Regressor" regARIMA approach with Original Series as Input:

This method represents a pure regARIMA modelling approach where all possible systematic calendar effects are fit simultaneously in the estimation process. This includes additive outliers, trend and seasonal breaks and trading day. It is currently being explored (and to some extent used) by the U.S. Census Bureau. Version 0.3 of X12-ARIMA incorporates an improved automatic regARIMA model selection procedure (see http://www.census.gov/ts/papers/asa2002kmm.pdf).

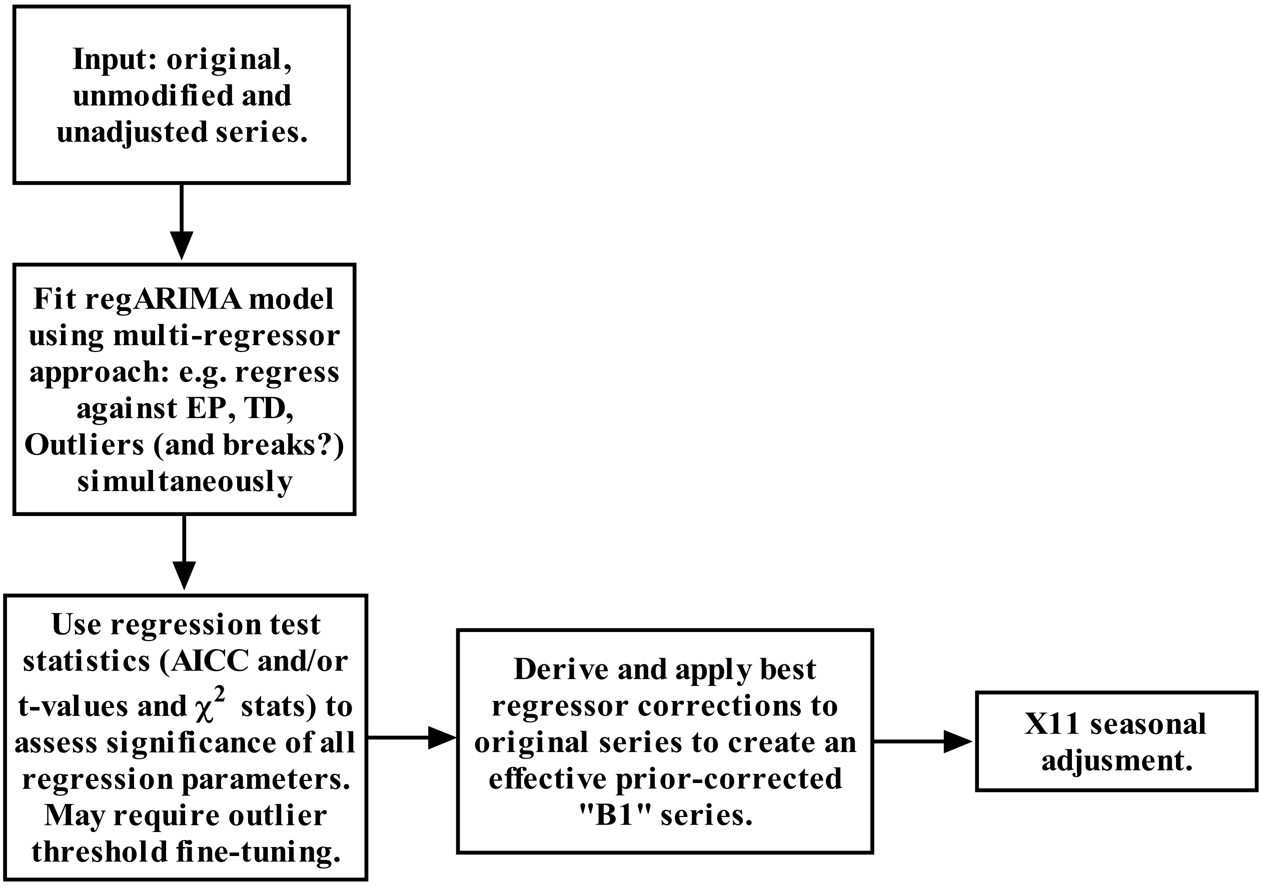
The significance and/or fine tuning of all regression parameters assumed in the "super-regressor" (i.e., third step in flowchart below), can be tested using either or both of the following criteria:

1. The minimum AICC statistic. The X12-ARIMA program has an input specification option "*aictest* " which calculates AICC values for a regARIMA model with and without a specific regressor included and chooses the model with the minimum AICC. This treats each regressor separetely and sequentially goes through each one to determine whether it improves the overall AICC. If so, the regressor is retained in the final simultaneous estimation.
2. Perform significance tests for each regressor parameter. Regression test statisitcs such as the t and χ^2 values are used to eliminate nonsignificant parameter estimates for all regressors. The best regressors can be selected according to the largest (most significant) t-values and/or smallest χ^2 values depending on the nature of the regressor. Note that if X12-ARIMA program identifies any outliers, they can be included in the final regARIMA model.

Some potential pitfalls in using this approach (which have not yet been explored or considered) are:

- The different types of regressors assumed in the "super-regressor" model may be correlated with one another. Does the regARIMA estimation process take this into account? Ignoring this will bias parameter estimates, their standard errors and hence their significance for inclusion in the "super-regressor.
- Related to previous bullet point, are all assumed regressors unique enough? In other words, we could be estimating redundant parameters across the different regressors, some of which may cancel in the end.
- We may need significantly long series spans to adequately model the full set of regressors and fit their parameters. The optimal spans are yet to be explored.
- If outlier detection is to be included in the "super-regressor", how sensitive are all the parameter estimates to the assumed outlier critical value?

Some of these issues can be addressed using simulated series with known input systematic effects.



Next steps:

- Which method above should we aim towards? I presume the "big picture" encompasses both since we want to retain our conservative SEASABS approach. Is there a priority?

- In the method 1 approach above, we want to isolate a robust set of EP regressors that cover most series cases. After this, we would like to explore how the best regressors can be chosen under a regARIMA model using the AICC and/or regression test critieria output by the X12-ARIMA program.

- Tie it all together. Plan possible modifications to the X12-ARIMA interface e.g., allow user to select from a list of "popular" EP regressors, or, allow an automodelling feature that selects the best EP regressor according to some criteria.

5. Easter Proximity Detection in the Presence of Outliers (version 2)

The following steps were used to generate the sequence of charts below:

1. Outliers were initially simulated (added to the original series at random points, including some Easter-proximity timepoints). These were then identified in X12-ARIMA using original input series + automatic outlier detection option (ao's only). Best ARIMA model was also identified in this run: (2 1 2)(0 1 1).

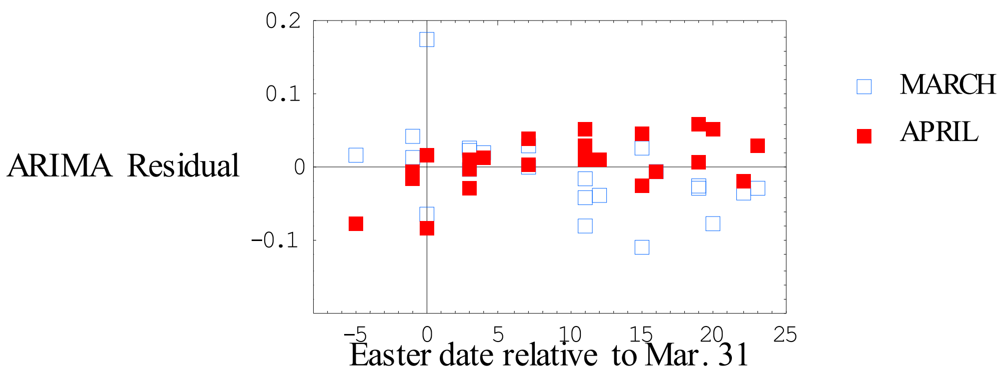
2. Regression variables for the locations of these outliers were added to 18 separate Easter Proximity (EP) regression matrices which correspond to models believed to representand real-world scenarios.

3. Each EP+AO regression matrix was used in separate regARIMA runs to select the best regression matrix according to smallest AICC statistic and most significant regression t-stats for all simultaneously estimated EP+AO model parameters. Note that the automatic outlier option was turned off in these runs to avoid distorting the best EP regressor selection.

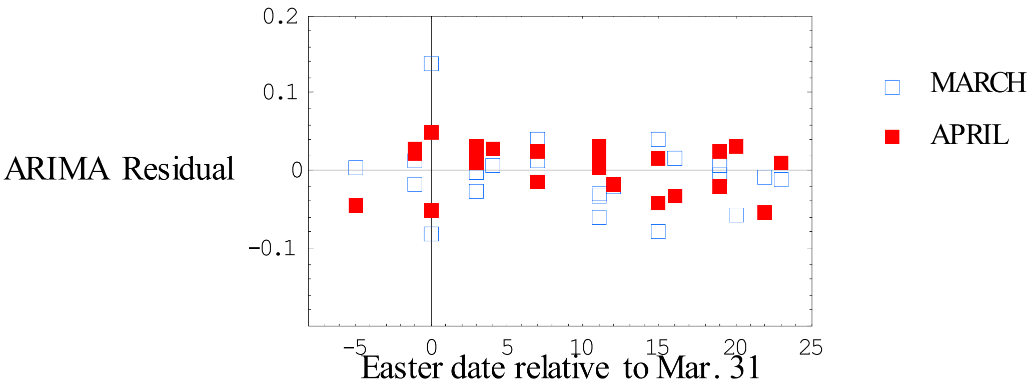
4. Final tweak/check of ARIMA model structure with the best regression (EP+AO) model included. This resulted in a slightly new "better" model [(2 1 1)(0 1 1)] according to the AICC.

Note that if we would have started with the prior-corrected (B1) series from Seasabs, the initial automatic outlier detection step in X12-ARIMA would not have been necessary. In this case, we would have just proceeded with the best EP selection (the AOs being implicitly assumed in the input B1 series).

Original series with ARIMA model fit only and no regression:
(ARIMA only)



Original series with regARIMA model fit with best EP regressor included:
(ARIMA + EP)



Original series with regARIMA model fit with best EP regressor and AO regressor included:
(ARIMA + EP + AO regressors)

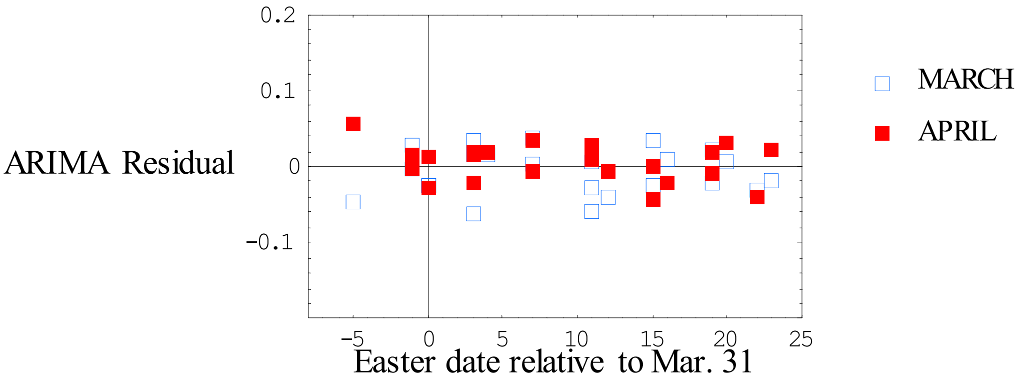


Figure 10: EP charts for "liquor retail series" in terms of ARIMA model residuals. TOP: using original series data (uncorrected for priors) with no regressors included; MIDDLE: with just EP regressor model included; BOTTOM: with both EP and AO (additive outlier) regressors included. The AO regressor was defined by locating potential outliers from a prior ARIMA run with automatic outlier detection. The same "best-fitting" ARIMA model (defined from a prior ARIMA run with no regressors) was used in all cases.

(reg)ARIMA output statistics for above cases:

ARIMA fit only:
Likelihood Statistics: Model (2 1 2)(0 1 1)

Effective number of observations (nefobs) 268

Number of parameters estimated (np)	6
Log likelihood	434.7223
Transformation Adjustment	-1396.5373
Adjusted Log likelihood (L)	-961.8150
AIC	1935.6300
AICC (F-corrected-AIC)	1935.9519
Hannan Quinn	1944.2839
BIC	1957.1759

ARIMA + best EP regressor:

Regression Model

Variable	Parameter Estimate	Standard Error	t-value
User-defined			
xb_whp_703	0.0416	0.00980	4.24

Likelihood Statistics: Model (2 1 2)(0 1 1)

Effective number of observations (nefobs)	268
Number of parameters estimated (np)	7
Log likelihood	442.9403
Transformation Adjustment	-1396.5373
Adjusted Log likelihood (L)	-953.5970
AIC	1921.1939
AICC (F-corrected-AIC)	1921.6247
Hannan Quinn	1931.2901
BIC	1946.3308

ARIMA + best EP + AO:

Regression Model

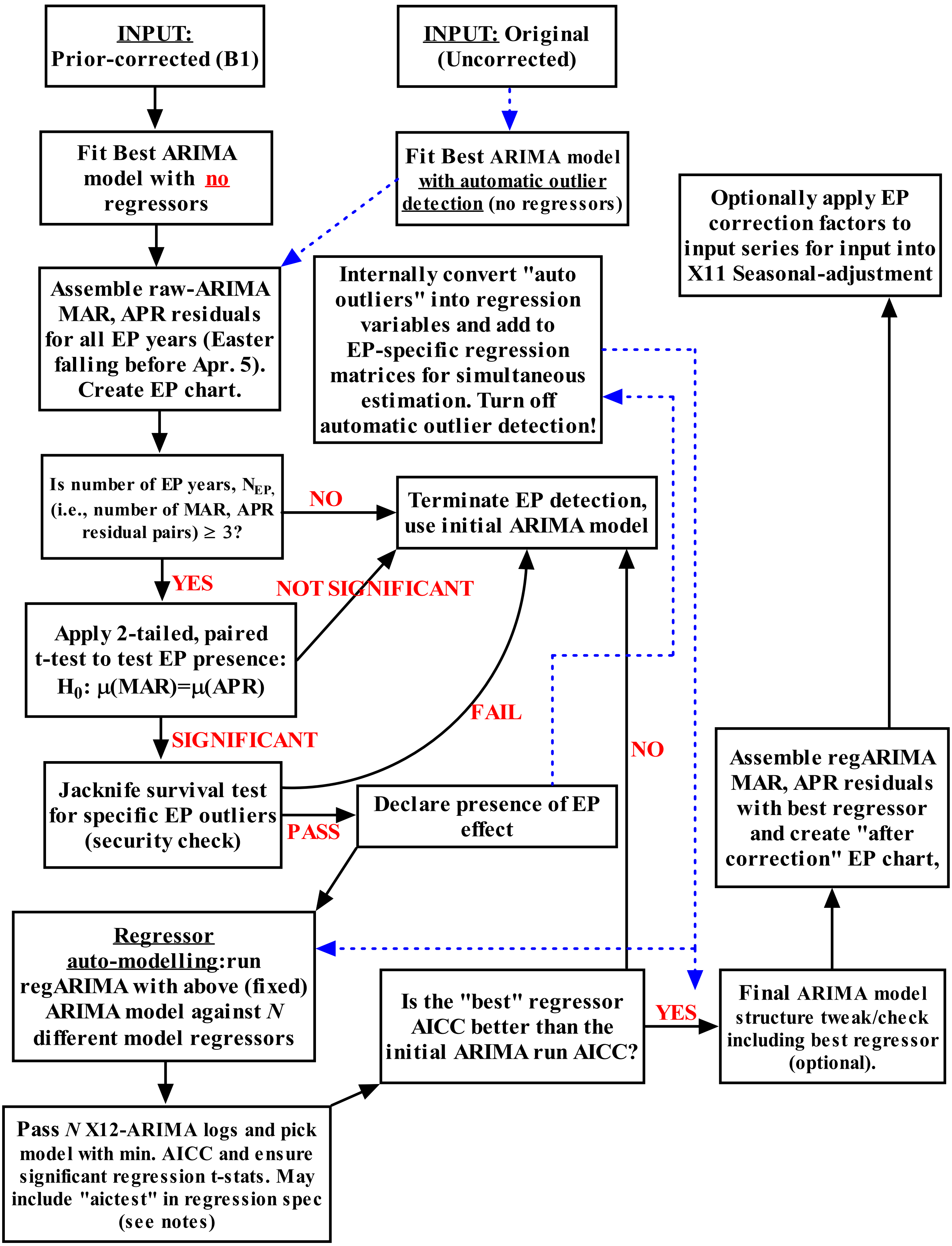
Variable	Parameter Estimate	Standard Error	t-value
User-defined			
xb_whp_703	0.0210	0.00750	2.80
ao1	-0.1103	0.02690	-4.10
ao2	-0.2156	0.02860	-7.54
ao3	-0.1203	0.02731	-4.40
ao4	-0.2286	0.02920	-7.83
ao5	0.1741	0.02774	6.28

Likelihood Statistics: Model (2 1 2)(0 1 1)

```
Effective number of observations (nefobs)              268
Number of parameters estimated (np)                     12
Log likelihood                              499.3473
Transformation Adjustment                        -1396.5373
Adjusted Log likelihood (L)                    -897.1900
AIC                                     1818.3799
AICC (F-corrected-AIC)                        1819.6034
Hannan Quinn                               1835.6877
BIC                                     1861.4718
----------------------------------------------------------------
```

(New) Method 3: Input series can be either "prior-corrected" B1 series from SEASABS or "Original":

This method unifies the two possible types of input series for X12-ARIMA into a generalised framework in correcting for Easter proximity (and possibly other calendar) effects. It allows for straightforward addition of other regressors (as well as multi-regressors) in future. The specific processing steps for the "Original Series" input case are followed by the blue arrows. The two input cases have many steps in common (flow represented by black arrows). See "NOTES" section below for more details on specific steps.



INPUT: Prior-corrected (B1)
INPUT: Original (Uncorrected)
Fit Best ARIMA model with no regressors
Fit Best ARIMA model with automatic outlier detection (no regressors)
Optionally apply EP correction factors to input series for input into X11 Seasonal-adjustment
Assemble raw-ARIMA MAR, APR residuals for all EP years (Easter falling before Apr. 5). Create EP chart.
Internally convert "auto outliers" into regression variables and add to EP-specific regression matrices for simultaneous estimation. Turn off automatic outlier detection!
Is number of EP years, N_{EP}, (i.e., number of MAR, APR residual pairs) ≥ 3?
NO
Terminate EP detection, use initial ARIMA model
YES
NOT SIGNIFICANT
Apply 2-tailed, paired t-test to test EP presence: H_0: μ(MAR)=μ(APR)
FAIL
SIGNIFICANT
NO
Assemble regARIMA MAR, APR residuals with best regressor and create "after correction" EP chart,
Jacknife survival test for specific EP outliers (security check)
PASS
Declare presence of EP effect
Regressor auto-modelling:run regARIMA with above (fixed) ARIMA model against N different model regressors
Is the "best" regressor AICC better than the initial ARIMA run AICC?
YES
Final ARIMA model structure tweak/check including best regressor (optional).
Pass N X12-ARIMA logs and pick model with min. AICC and ensure significant regression t-stats. May include "aictest" in regression spec (see notes)

Notes:

- ARIMA model residuals are generated using the save=(residuals); print=(residuals) options in the estimate{...} X12-ARIMA spec.
- The Easter Proximity detection test was desribed in section 4 above.
- Jacknife EP-specific outlier test was described in section 4 above.
- A regressor automodelling ("autoregressor") feature needs to be implemented where once cycles through all regression matrices in the list (already implemented) for the specified ARIMA model.
- Note that X12-ARIMA automodelling option which cycles through just ARIMA models should remain. When a user also specifies "autoregressor" modelling, it should be used in conjuction with the ARIMA automodelling. In other words, for each regressor model in the list (say N regression matrices), run the X12-ARIMA automodelling feature (all ARIMA models) so that results will be written to N separate log files.
- In the "autoregressor" step, it may be necessary to concatenate the N regressor-specific log files for efficiency in storage, or otherwise, name each output log file according to the unique regressor-matrix name.
- Passing of the N X12-ARIMA output logs involves picking out lines containing the crucial "AICC" statistic and line(s) containing each regression parameter with t-statistic. May also want to check χ^2 values for "groups of regressors" to ensure they are significant.
- The option "*aictest* " in the regression{...} spec calculates AICC values for a regARIMA model with and without the specific user-regressors included and chooses the model with the minimum AICC.
- Following the conversion of auto-detected outliers to regression variables for the "original" input series case, the automatic outlier option needs to be turned off before the subsequent autoregressor-regARIMA runs (next step) to avoid distorting the best EP regressor selection.
- For the "original" input series case, the user can specify multiple regression variables in each regression matrix in the list (each matrix in the list is envisaged to cover a range of models where the best will be picked by the AICC test). This simplies the "big picture" with an outlook to the future. The processing flow above is not just restricted to correcting for the EP effect.

Outputs:

- Easter proximity charts (before and after correction).

- Dialogue summary window with EP regression parameter estimates and t-stats. This can be generalised in that the results for all regressors can be added. EP specific estimates and stats can also be written on the EP chart.

- New series corrected for desired calendar related effects (not only the EP effect) and outliers for input into X11 seasonal adjustment.