Quantifying and Improving Consistency between Series

Frank Masci

Version 4: 15/11/2006

A history of related work and other doclinks can be found under the original proposal document: (Subject: Coherence - project proposal; Database: Time Series Analysis WDB; Author: Tom Outteridge; Created: 05/09/2005; Doc Ref: TOUE-6FX82V).

Below we present six "consistency" diagnostic tests that have already been implemented into a prototype software tool [Subject: BPG Consistent Seasonal Adjustment (draft in progress); Database: Time Series Analysis WDB; Author: Lisa Apted; Created: 26/07/2005; Doc Ref: LAPD-6EN5PZ]. This tool is designed to give people a feel on how the consistency checking process will work in a future version of SEASABS. The discussion below is based on the output from this tool. Here's an example:

results_sa.html

1. Diagnostic Tests

Note: *the discussion in this section is independent of whether you're dealing with original, seasonal adjusted (or other) components for a series pair. See section 2 for implications on different series components.*

The crucial summary measures reported by the consistency tool are indicated below in ***red****.*

TEST 1 - "Consistency (SANITY) Check between Main Parameters in Summary Table":

This test does a quick scan of the consistency summary table for a series pair and if their periodicities are the same (e.g. monthly vs. monthly or quarterly vs. quarterly), a report of all inconsistent "main" processing parameters is given. By "main", we mean the following parameters and settings: "Method", "Model", "TMA", "PUB_TMA", "PUB_EWP", "SMA", "Trading day", "Easter" and "Father's Day".

The "Method" field can be either "direct", "indirect", or "not adjusted". This field is only compared if either series is "not adjusted" since we don't necessarily want to declare an expected "direct" vs. "indirect" (ie. aggregative) series pair as inconsistent. However, the prototype tool currently ignores the "Method" comparison since TSA-download does not yet support "no adjusted" cases. This functionality will be present in the SEASABS version.

Note that trend breaks, seasonal breaks, large extremes and additive outliers are excluded from this test since these are covered under a more general framework in the "movement" tests below. If the periodicities of a pair are different (e.g. monthly vs. quarterly), then differences in the above "main" parameters are allowed (or expected). But please check that they do indeed make sense for each series.

A summary of this test is prefixed by lines with "***" in the consistency tool:

E.g. you should review the parameters and/or settings if you see something like:

*** The "X11 TMA, PUB_TMA, <..and/or other>" parameter settings are not consistent for this series pair. Please review.

Otherwise, you'll be satisfied (with regards to these "main" parameters/settings only!) if you see:

*** All main processing parameters for this series pair are consistent (i.e. "Model", "TMA", "PUB_TMA", "PUB_EWP", "SMA", "Trading day", "Easter" and "Father's Day").

If the periodicities of each series in the pair are different, then you'll see:

*** Periodicities are not the same: inconsistencies are expected in processing parameters and settings. Please check all are appropriate.

TEST 2 - "Cross-Correlation for Relatedness":

This presents a simple sanity check to ensure that the series pair are related. This can be classified as a "weak" consistency measure since if an input series pair are declared to be "conceptually" related on apriori grounds (which is the working hypothesis), then it is guaranteed that their movements are correlated to some degree. But it doesn't hurt to check.

- The basis of this test is to ensure that the cross-correlation coefficient between movements is significantly different from zero at some lag h. i.e. we test:

  $$H_0 : \rho(M_1, M_2)_h = 0 \quad \text{versus} \quad H_1 : \rho(M_1, M_2)_h \neq 0$$

- Rejection of this null hypothesis (at some significance level) at any of the interesting lags: h=-3 -> 3 implies that the movements are "related" in some way (note that correlated does not imply equal). In most "conceptually related" series cases, we expect to see a positive significant peak in ρ at lag zero (=CCF(0); i.e. no shift between the series).

- In the consistency tool output, we quote these measures with and without outlying movements included. These outliers are identified using robust linear regression as described in TEST 4 below. The presence of outliers in the joint movement distribution may skew estimates of ρ (i.e. reduce their significance), leading one to reject an otherwise significant underlying correlation (or relatedness). We therefore present estimates with and without inclusion of outliers. The latter gives us a second chance at assessing relatedness. For your convenience, we provide plots of CCF versus h for each of these cases.

- Along with the CCF(0) measures, we also quote 95% confidence intervals expected for the population ρ (from Fisher's z-transformation method). If these intervals include the value zero, then H_0 is accepted at some P-value >5% and no amount of processing can make it better. A concise summary of this test is prefixed by "***" in the consistency tool output summary:
  E.g. you <u>should be concerned</u> if you see something like:

"*** CROSS-CORRELATION COEFFICIENT (WITHOUT OUTLIERS) CONSISTENT WITH ZERO AT >5% LEVEL => SERIES MAY NOT BE RELATED"

Otherwise, all is well (at the <5% chance level!) if you see:

"*** CROSS-CORRELATION COEFFICIENT (WITHOUT OUTLIERS) SIGNIFICANTLY DIFFERENT FROM ZERO AT <5% LEVEL => SERIES ARE INDEED RELATED"

TEST 3 - "Equal-Movement Magnitude Tests":

This test is performed on outlier corrected movements to check whether the underlying movement magnitudes and directions are mutually consistent overall. In other words, let us pose the question: are the movement magnitudes of series 1 generally larger (or smaller) than those in series 2? If they are significantly different (inconsistent), then it raises a warning (query) for the client and an interpretation is needed. Note that the movements can still be significantly correlated even if their mean overall magnitudes differ. No amount of tweaking or processing by TSA can force the movement magnitudes to be globally consistent.

- The first measure we compute is the mean difference in movements (i.e. <M1 - M2>). If this is significantly different from zero (i.e. H_0: <M1-M2> = 0 is rejected according to a two-tailed t-test), then the movement magnitudes are globally <u>inconsistent</u>.

- If the movement magnitudes are found to be significantly different (i.e. <M1 - M2> .ne. 0) overall, then it would be of interest to determine the nature of this difference using a robust linear regression fit to the movements. By robust, we mean "resistant" to outliers as much as possible. E.g. a linear regression fit can be parameterised as:

$$M_2 = \alpha M_1 + \beta,$$

Our goal here is to determine whether a significant global difference in movements is either wholly due to a global relative offset (e.g. non-zero intercept β), a scale factor (e.g. non-unit slope α) or both. In other words,

$$\text{if } \langle M_2 - M_1 \rangle \neq 0 \text{ then either:}$$
$$\langle M_2 - M_1 \rangle \approx \beta \quad \text{OR}$$
$$\langle M_2 - M_1 \rangle \approx (\alpha - 1)\langle M_1 \rangle \Leftrightarrow \frac{\langle M_2 \rangle}{\langle M_1 \rangle} \approx \alpha \quad \text{OR}$$
$$\langle M_2 - M_1 \rangle \approx (\alpha - 1)\langle M_1 \rangle + \beta.$$

If you're feeling ambitious, you may want to show that the variance in the difference of the movements can be written:

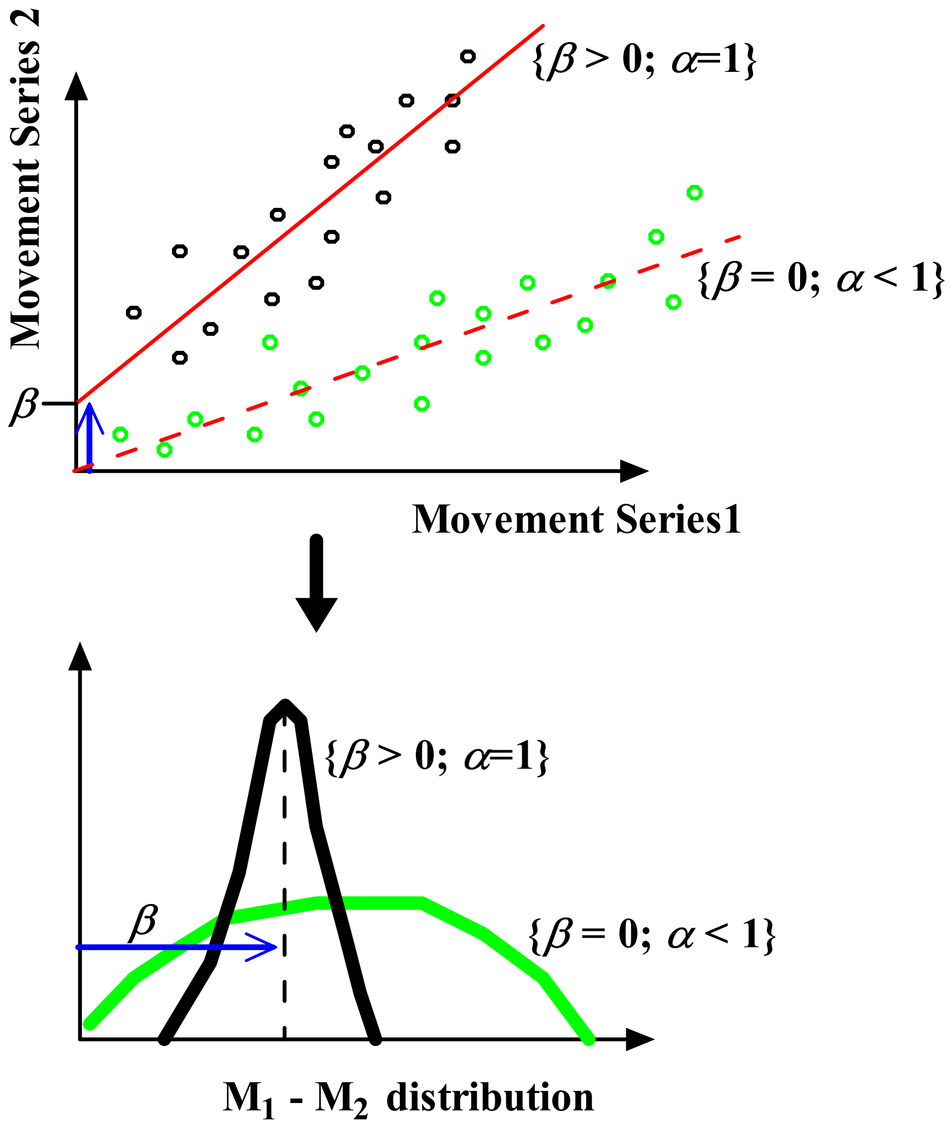
$$\sigma^2_{(M1-M2)} = \begin{cases} \sigma^2_{M1} + \sigma^2_{M2} - 2\rho\sigma_{M1}\sigma_{M2} & |\rho| \leq 1, \\ \sigma^2_{M1}(1-2\alpha) + \sigma^2_{M2}; & |\alpha| < \infty, \\ \sigma^2_{M1} + \sigma^2_{M2}; & \alpha = 0, \\ \sigma^2_{M2} - \sigma^2_{M1}; & \alpha = 1 \end{cases}$$
$$\rho \equiv \alpha\left(\frac{\sigma_{M1}}{\sigma_{M2}}\right),$$

where $\sigma_{M2} \neq 0$ and σ^2_{M1}, σ^2_{M2} are the sample variances of the M_1 and M_2 (conditional) distributions projected onto their axes respectively.

Figure 1 below illustrates these concepts.

- For simplicity, we test for a non-zero intercept and a non-unit slope independently (not jointly) using t-tests on results of the linear regression:

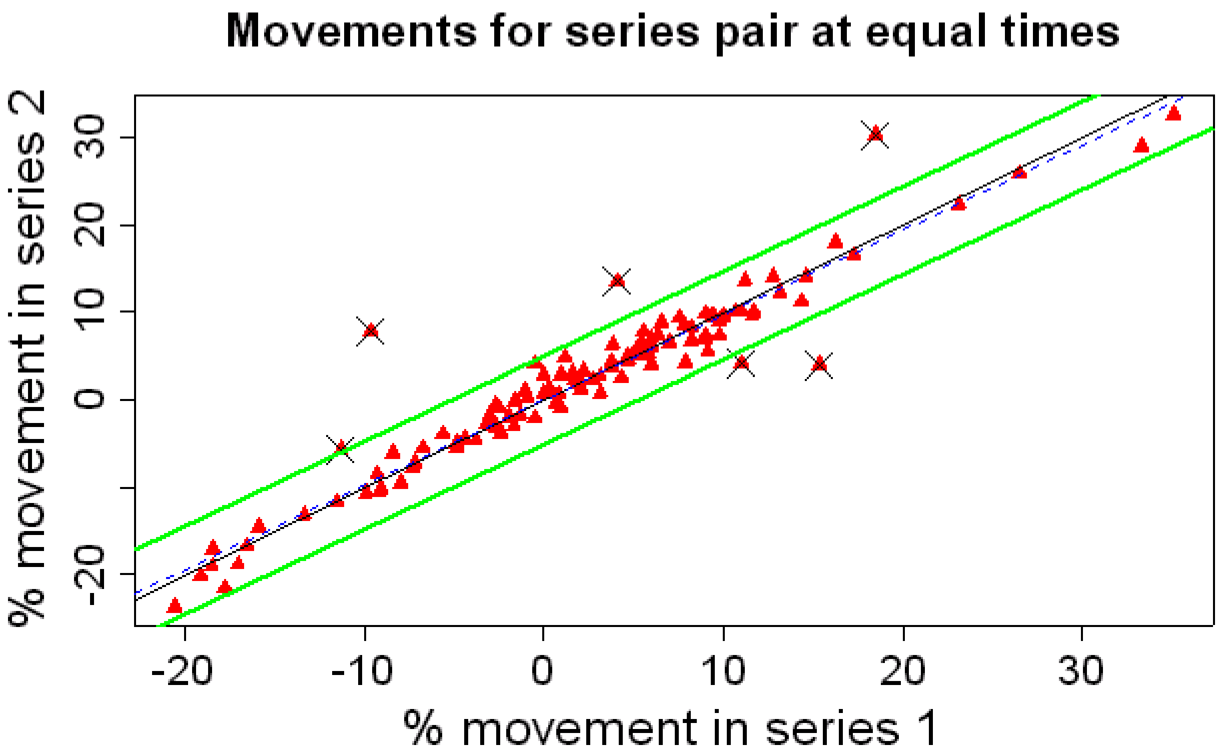
$$H_0 : \alpha = 1 \quad \text{versus} \quad H_1 : \alpha \neq 1 \qquad \text{and}$$
$$H_0 : \beta = 0 \quad \text{versus} \quad H_1 : \beta \neq 0$$



<u>Figure 1:</u> interpretation of non-zero intercept / non-unit slope.

- A significantly non-zero intercept will have a different interpretation from a significantly non-unit slope. Either or both of these imply an inconsistency in movement magnitudes in the broad sense and the above test is an attempt to probe the relationship of the movement magnitudes further. Whether the movement magnitudes from two series are related in a multiplicative or additive way will be of great interest to the client.

- The consistency tool output includes a plot of the movements in series 1 versus that in series 2 at equal times. E.g:



<u>Figure 2:</u> example movement versus movement plot from consistency tool.

The <u>black solid line</u> defines the "line of equal movements", the <u>blue dashed line</u> is the robust linear regression fit, the <u>green lines</u> encompass a ~99.73% (~3σ) confidence interval for the residuals about this fit and the crosses indicate outliers falling outside this interval. See TEST 4 for their detection method. Obviously the more the blue dashed line coincides with the black line, the more "consistent" are the movement magnitudes.

- The consistency tool outputs diagnostics of the linear regression and a concise summary of this test is given by lines prefixed by "***":

E.g. you should make a note and let the client know if you see something like:

*** MEAN DIFFERENCE IN MOVEMENTS SIGNIFICANTLY DIFFERENT FROM ZERO AT 5% LEVEL

which could be due to either of following:

*** FITTED SLOPE SIGNIFICANTLY DIFFERENT FROM ONE AT 5% LEVEL
*** FITTED INTERCEPT SIGNIFICANTLY DIFFERENT FROM ZERO AT 5% LEVEL

Otherwise, all is well if you see (for example with p-values >5%):

*** MEAN DIFFERENCE IN MOVEMENTS CONSISTENT WITH ZERO AT 36.97 % LEVEL
*** FITTED SLOPE CONSISTENT WITH ONE AT 16.45 % LEVEL
*** FITTED INTERCEPT CONSISTENT WITH ZERO AT 65.23 % LEVEL

TEST 4 - "Movement versus Movement Outlier Test":

This test is probably the most useful of all. The basis of this test is the detection of outliers in the movement versus movement plot (e.g. see Figure 2 above). Depending on the series component under investigation, the presence of discrepant movements reveal inconsistencies at the same time points between two series. For example, if "prior-corrected" (D11) seasonally adjusted series are being compared, then the presence of a large movement in one series but not the other may imply that there has been an inconsistent prior-correction applied, or, that the actual originals are conceptually different at those timepoints.

- Outlier movements are detected by identifying those residuals r_i about a "robust" linear regression fit that exceed some thresold "t " of the Median Absolute Deviation (*MAD* measure) from the median. i.e. an outlier is declared if:

$$\frac{\left|r_i - median\left\{r_i\right\}\right|}{MAD} > t, \qquad \text{where}$$

$$MAD = 1.4826 * median\left\{\left|r_i - med\left\{r_i\right\}\right|\right\}$$

The numerical prefactor is a correction for asymptotic normality so that *MAD* ~ σ for large N . In this limit, the threshold "t " can therefore be expressed in terms of the number of standard deviations (e.g. 2σ , 3σ etc..; the consistency tool has this hardcoded at 3σ). The green bands in Figure 2 above represent: α* M1 + β +/- 3σ (with α and β from TEST 3). We use the "MAD" measure because of its robustness.

- A list of all outliers detected are reported in the diagnostics output. It is recommended that the user trace their cause. An example output is as follows:

*** NUMBER OF > 3 SIGMA OUTLIERS DETECTED IN MOVEMENT VS MOVEMENT PLOT = 6 :

```
* ALL OUTLIERS:
* MNTH or QTR/YEAR | MVT1(%) | MVT2(%) | approx.#SIGMA | P-VALUE:
* ------------------------------------------------------------------------
* 3/1977      | -9.61    | 7.78     | 10.12     | 4.307e-024
* 4/1977      | 15.35    | 3.97     | -6.32     | 2.539e-010
* 3/1978      | 4.09     | 13.46    | 5.65      | 1.609e-008
* 3/1979      | 18.47    | 30.30    | 7.35      | 2.039e-013
* 1/1981      | 11.02    | 4.08     | -3.79     | 1.486e-004
* 3/1981      |-11.28    | -5.64    | 3.20      | 1.357e-003
* ------------------------------------------------------------------------
```

TEST 5 - "Test for Significant Temporal Pattern in Movement Differences":

We attempt to search for a regular pattern in "excess" (or outlying) movement difference (M1 - M2) as a function of time. By excess, we mean values of the movement differences that exceed some threshold (see below). For instance, if a significant pattern is found between seasonally adjusted components on timescales of 12 months, then this may imply that residual seasonality is present in one series but not the other. Note that this test is not sensitive at detecting residual seasonal movements of the *same magnitude* in both series. If we are examining original (unmodified) series, then a significant pattern in thresholded "M1 - M2" values may imply different seasonal patterns.

- The first step involves finding all movement differences D_i = $M1_i$ - $M2_i$ that exceed some thresold "*t* " of the Median Absolute Deviation (*MAD* measure) from the median. In other words, the detection of excess movement differences uses the same algorithm as used for detecting *residual* outliers in TEST 4 with r_i replaced by D_i .

- If an excess is detected, we assign an indicator variable *i(t)* with the value 1, or 0 if the movement difference falls below the threshold. In the consistency tool, this threshold is hardcoded at 2σ. In algorithmic terms,

$$
\begin{aligned}
&if\left(\frac{|D_i - median\{D_i\}|}{MAD} > t\right), \quad \text{where } MAD = 1.4826 * median\{|D_i - med\{D_i\}|\}, \\
&\{\ \ i(t) = 1\ \ \} \\
&else \\
&\{\ \ i(t) = 0\ \ \}
\end{aligned}
$$

- Given the time sequence of indicator variables, *i(t)* , we then explore whether it exhibits a significant pattern by computing its autocorrelation and partial-autocorrelation functions (ACF[i(t)] and PACF[i(t)] respectively) at a number of interesting time lags h = 1...12. Our null and alternative hypotheses are:

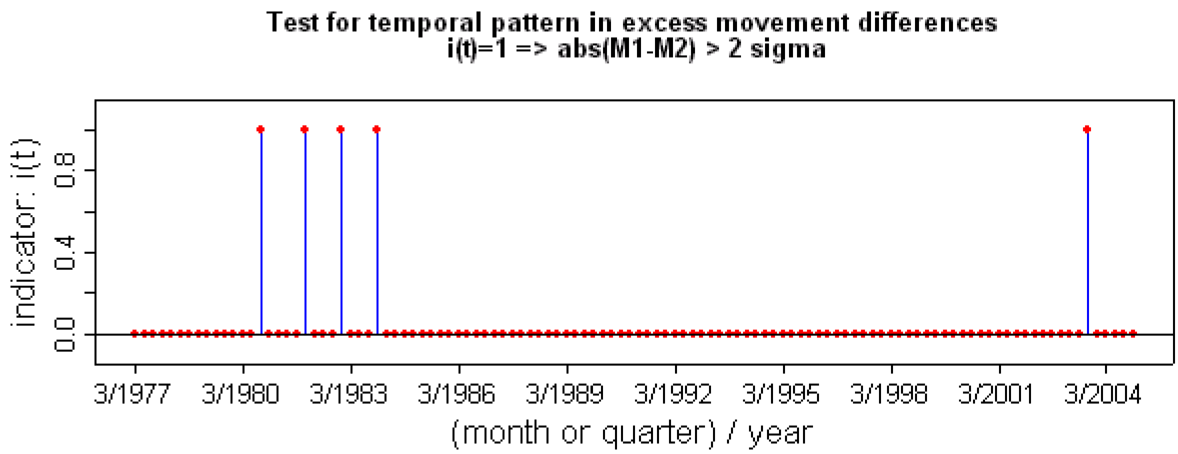
$$
\begin{aligned}
&H_0 : ACF[i(t)]_{lag\ h} = 0 \quad \text{versus} \quad H_1 : ACF[i(t)]_{lag\ h} \neq 0 \quad \text{and} \\
&H_0 : PACF[i(t)]_{lag\ h} = 0 \quad \text{versus} \quad H_1 : PACF[i(t)]_{lag\ h} \neq 0
\end{aligned}
$$

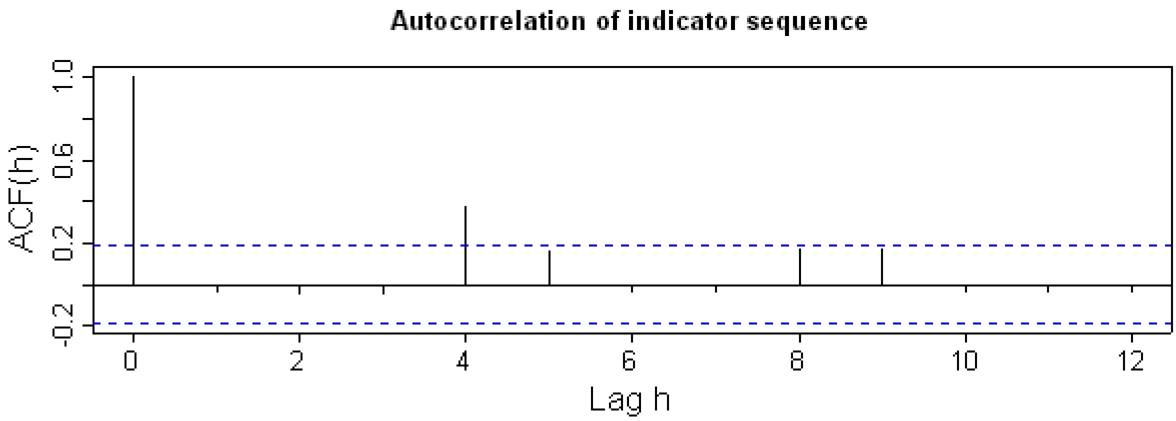
- Rejection of H_0 for both the ACF <u>and</u> PACF at any lag implies a significant temporal pattern in thresholded movement differences (M1 - M2) at that time lag. On the other hand, acceptance of H_0 at all lags implies no temporal patterns are present. The reason why we consider both the ACF and PACF is because the ACF values on their own do not measure the "true" autocorrelation at a specific lag. A significant ACF value at say lag 12 could be due to the indirect propagation of autocorrelation signal at smaller lags. The PACF values have the effects from intervening lags removed. A joint comparison of the ACF and PACF values makes this test more robust.

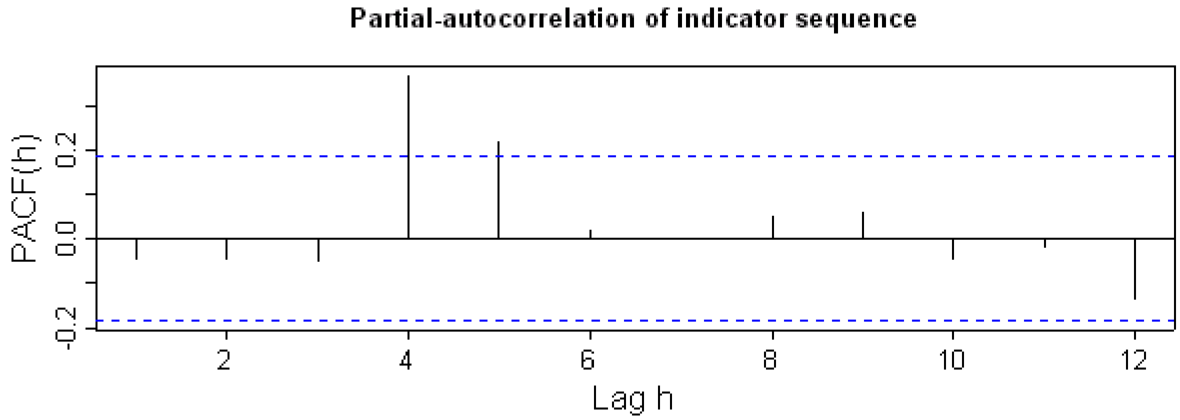
- The consistency tool produces as output, plots of the indicator variable *i(t)* and the ACF and PACF of i(t) at lags h=1 to 12 lags. The ACF and PACF plots show the 95% confidence intervals (~ +/-1.96/sqrt[*N*]) outside which values can be considered significant. There is also a summary of this test. If a significant temporal pattern is found, you'll see, for example:

* MOVEMENT DIFFERENCE THRESHOLD = 2 SIGMA
* NUMBER OF |M1 - M2| VALUES ABOVE THRESHOLD = 5
* LAGS WHERE ACF & PACF[i(t)] ARE GREATER THAN 95% CL VALUE: LAG 4 ;
* => REJECT "H0: ACF & PACF[i(t)]=0" OF NO PATTERN IN MOVEMENT DIFFERENCES
* => SYSTEMATIC INCONSISTENCY EXISTS AT ABOVE (LAG) PERIODICITIES (CHECK ACF PLOTS)

It's important to never believe these conclusions outright. As the last line implies: "CHECK ACF PLOTS". You want to ensure that there is indeed a systematic pattern in the indicator { i(t) } sequence. The plots corresponding to this example are below. As you see, the reported significant autocorrelation at lag 4 only occurs over a three year span. Since this is a quarterly series, it implies an excess seasonal movement difference at these consecutive three years between the two series. This warrants a look at the actual S*I charts for these series.







TEST 6 - "Unit Root & Cointegration Tests":

Goals and Motivation:

The above diagnostic tests exclusively focused on the movements (first-differences) in each series of a conceptually related pair. The reason for this is that first differences are likely to be 'stationary' for most economic/business-like series. If one attempts to correlate levels directly, then correlation measures may be spurious due to the presence of non-stationarity (a feature of most economic series since they usually show an upwards trend). Correlation measures will be biased towards explaining global long-term relationships in the trend levels since both depend (usually linearly) on a third variable - "time". The correlation measures derived therefrom will not be a true representation of relationships on the short timescales sought for (i.e. on which the series are sampled).

To see this via an example, consider two non-stationary series X_t and Y_t , say with deterministic trends of unequal slopes c_1 , c_2:
$X_t = c_1.t + e_t$
$Y_t = c_2.t + e'_t$,
where e_t and e'_t are zero mean IID noise sequences.
The covariance between these series can be written:
$\mathrm{cov}[X_t, Y_t] = \langle(X_t - \langle X_t\rangle)(Y_t - \langle Y_t\rangle)\rangle$
$= c_1.c_2.\mathrm{var}(t)$
which is clearly non-zero and depends on the variance of the timepoints in the span. This means that the correlation is due to the relationship in the overall trends, not between underlying point-to-point movements across series where correlations are expected to be zero (i.e. $\mathrm{cov}[e_t.e'_t] = 0$).
Furthermore, consider a linear regression between these two series: X_t and Y_t :
$Y_t = B.X_t + A$,
where A and B are parameters of the regression. The residuals from this regression:
$u_t = Y_t - B.X_t - A$
$= (c_2 - B.c_1).t + e'_t - B.e_t - A$
will be non-stationary and highly autocorrelated at all lags. Consequently correlation measures and t-statistics on parameter estimates will be biased (former overestimated and latter underestimated) since there are 'unexplained'

non-white noise residuals in the regression. Cointegration tests can help alleviate and/or assess cases that may lead to spurious correlations.

Cointegration is provided here for informational purposes and explores whether there exists a stable 'long-run' (ie. equilibrium) relationship between two series, or put another way, whether they have common long-term stochastic trends. If so, the series are 'cointegrated'. If a pair of series are cointegrated, then one of them may 'cause' the other. Since the series pairs fed into the consistency tool are expected (or known a-priori) to be conceptually related, cointegration should follow naturally. Below we provide a brief overview of the tests performed. For a more in depth discussion with links to lecture notes, see (Subject: Cointegration notes/articles..; Database: Time Series Analysis WDB; Author: Frank Masci; Created: 14/06/2006; Doc Ref: FMAI-6QR3PB).

Unit Root Tests:

- Before testing for cointegration, we must first determine if both series are consistent with "integrated" processes of the same order. By "integrated", we mean they have underlying 'non-stationary' stochastic trends (just like a classic random walk), as opposed to deterministic trends which can also be non-stationary. An 'integrated' process results from a cumulative sum of repeated stochastic disturbances over time. Consider the following process and it's first difference:

$$y_t = \rho y_{t-1} + e_t$$
$$\Delta y_t = \delta y_{t-1} + e_t,$$
where $\delta = \rho - 1$ and $\Delta y_t = y_t - y_{t-1}$.

The classic random walk is defined by $\rho = 1$ and its first difference has $\delta = 0$. This model is said to have a "unit root" and is non-stationary. If a series with an underlying stochastic trend has to be differenced once in order to make it stationary, then it is called "integrated of order one", or I(1). A series which needs to be differenced d times before it is stationary is called I(d). We limit ourselves to testing for I(1) cases only since these underly most economic-like series. We perform two separate unit-root tests on each series to determine if each is I(1):

- The first unit root test is the **Augmented Dickey-Fuller (ADF)** test (e.g. "ur.df" function in the "urca" R package). This is based on performing the following general regression:

$$\Delta y_t = \beta_1 + \beta_2 t + \delta y_{t-1} + \sum_{i=1}^{m} \alpha_i \Delta y_{t-i} + e_t, \qquad (1)$$

i.e. 'nested' constant and trend terms are included and if the error term e_t is autocorrelated, lagged difference terms are added to soak up any residual autocorrelations so to ensure whiteness in e_t . m is the number of lagged difference terms to include (m = 0 corresponds to a straight 'Dickey-Fuller' test). The ADF test is based on testing the null and alternative hypotheses:

$H_0 : \delta = 0 \Rightarrow y_t$ contains a unit root and is integrated with order one = "I(1)"

versus

$H_1 : -2 < \delta < 0 \Rightarrow y_t$ stationary "stochastic" process (not integreated)

- The second unit root test performed is called the **Phillips-Perron Test** (e.g. "ur.pp" function in the "urca" R package). This is a variant of the ADF test in that different estimators and statistics for the null distribution in δ and coefficients for lagged differences (α_i) are used, mostly based on semi-parametric methods. The method includes a more careful consideration of possible correlations between all parameter estimates (slope and intercept terms included). This test is known to be more powerful than the ADF test when structural breaks (or sudden impacts) are present in a series. Breaks can render a series non-stationary (integrated) when it is really stationary (i.e. type II error). Like the ADF test, this test also has 'poor' finite sample behaviour (say for spans N <~ 100 observations) since critical values are derived in the large N asymptotic limit. This test become equivalent to the ADF test when all lags are explicitly omitted (all α_i = 0).

- A difficulty with the above tests is deciding how many lagged differences to include in the regression (i.e. m). In general, m should be large enough to remove any evidence of autocorrelation in the regression residuals e_t . For the time being, we have picked "m = 2" for both tests since we have prior knowledge from ARMA modelling that most series in our collections exhibit AR(<3) behaviour. Incorrect specification of m could pose a problem for series exhibiting MA(q) behaviour since equivalent AR(p) representations require p and hence m to be large (in principle infinite!). Further research into optimising the number of lags to include in these tests is needed, possibly via the AIC or other robust methods.

- Another difficulty is whether to include a deterministic trend (non-zero β_2) when fitting eqn (1). The assumption of a deterministic (linear) trend may masquerade a stochastic trend being sought for. In other words, if one allows a deterministic term in the fit (ie. β_2 .ne. 0), this may 'explain away' the real underlying stochastic component, therefore giving a misleading I(0) (stationary) process (ie. type I error). To avoid this, we take a conservative approach and assume that any non-stationary trend is potentially an I(1) process. We therefore omit fitting a deterministic trend term in all cases by explicitly setting β_2 = 0.

- A prelude to performing cointegration tests is to ensure that both series are integrated with the same non-zero order d: "I(d)" [d > 0]. This will become clearer below. Aside from this requirement, if a pair of series are not integrated of the same order, then they may not be related at all. The prior assumption of a 'conceptually related' series pair is invalidated and this must be brought to the client's attention. We apply two cointegration tests, each based on different estimation methods but which in the long run are expected to converge:

Cointegration Test 1:

- The first is called the **<u>"Augmented Engle-Granger (AEG) test"</u>**. This is based on first performing a linear regression between two series: X_t and Y_t :

  $Y_t = b.X_t + a$,

  where a and b are parameters. This is rearranged to obtain the actual regression residuals:

  $u_t = Y_t - b.X_t - a.$

  Now the series u_t can be seen as a linear combination of series X_t and Y_t and <u>if it is stationary</u>, then effectively the trends in X_t and Y_t (stochastic or otherwise) 'cancel out'. In this case, X_t and Y_t are cointegrated with cointegrating parameter " b ". Note that if the series are described by multiplicative decomposition models, then logs of the series are regressed instead. This is to ensure stationarity in the variance of regression residuals. However, this is <u>not</u> a strict requirement for the AEG test since it is based entirely on testing for stationarity in the <u>mean</u> of the residuals.

- The basis of this test is to first ensure that two series are integrated with the same, <u>non-zero</u> order, i.e. both have underlying <u>stochastic</u> (non-constant) trends. If not, then there's no point in going on since u_t will be expected to be either non-stationary (the series don't track each other), or, if both input series are I(0) (already stationary), then there will be <u>no 'common' long-term trends</u> to correlate. In other words we cannot infer any long-run equilibrium relationship between two series if they are made (or are already) stationary. This is the basis of cointegration.

- Note that a linear combination between two I(0) processes can still be stationary if they exhibit similar underlying autocorrelation structures. We can therefore still attempt to perform a regression between two series if both are I(0). Correlation measures derived therefrom are <u>not</u> expected to be spurious. However, as discussed, we cannot proceed with cointegration tests since the relationship (if any) is only local (i.e. in the short run). Cointegration aims to explore long-run steady-state causal relationships between integrated (non-stationary) processes.

- Therefore, given the requirement of <u>equal non-zero</u> integration order, two time series are cointegrated if the residual term from a regression of one on the other is <u>stationary</u> in the long run. Note that two independent random walks are both I(1) but they are very unlikely to be cointegrated since the chance that they will track each other in the long run (hence giving stationary regression residuals) is miniscule.

- The AEG test is based on applying ADF test described above to the residuals series u_t with maximum lag order m = 2. The null and alternative hypothesis are:

  $H_0 : \delta = 0 \;\Rightarrow u_t$ non-stationary $\Rightarrow$ series pair non-cointegrating

  versus

  $H_1 : -2 < \delta < 0 \;\Rightarrow u_t$ stationary $\Rightarrow$ series pair is cointegrating

  The usual critical values for the ADF statistic as used for the unit-root test above are not appropriate here due to the combination and interplay of two integrated processes. Instead we use the critical τ statistics for Engle-Granger cointegration as derived by MacKinnon, J. G., 1996. These critical values also contain finite sample corrections (ie. depend on the number of observations) so that large-sample requirements (as assumed in the univariate ADF tests) can be relaxed.

Cointegration Test 2:

The second test is considered more general in the econometrics world. This is "Johansen's Vector Auto-Regressive, Error Correction Model" (VAR-ECM) procedure. We use this method only to determine if a series pair is cointegrated and not to assess or correct deviations from long-run equilibrium.

- Given two series X_t and Y_t integrated with the same non-zero order, this test is based on expressing the first differences (movements) in a VAR-ECM framework. The existence of a 'unique' error-correction model (ECM) representation will ensure that the series are cointegrated. In a nutshell, the VAR-ECM for two series can be written:

$$\begin{pmatrix} \Delta X_t \\ \Delta Y_t \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} + \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} t + \begin{pmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \end{pmatrix} \begin{pmatrix} X_{t-1} \\ Y_{t-1} \end{pmatrix} + \sum_{i=1}^{m} M_i \begin{pmatrix} \Delta X_{t-i} \\ \Delta Y_{t-i} \end{pmatrix} + \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix}, \qquad (2)$$

where from left to right we have the vectors and matrices: first differences in the series, column vectors of constant intercept and linear (deterministic) trend coefficients, a 2x2 coefficient (or cointegrating) matrix multiplying the lagged series, a sum of lagged differences with coefficient matrix M_i and maximum lag order m , and a vector of (possibly correlated) white noise components e_t .

- Let's consider the 'cointegrating matrix' in the above representation, i.e:

$$\Pi = \begin{pmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \end{pmatrix}$$

The test for cointegration between X_t and Y_t is based on looking at the rank of the Π matrix. The rank of a matrix, r , is equal to the number of its non-zero eigenvalues, or equivalently, the number of linearly independent rows or columns. In this context, r is therefore equal to the number of independent cointegrating relations that exist amongst X_t and Y_t . For example, if Π had precisely the form:

$$\Pi = \begin{pmatrix} \alpha_1 & \alpha_1\beta \\ \alpha_2 & \alpha_2\beta \end{pmatrix},$$

then rank(Π) = 1 since the second row is "α_2 / α_1" times the first row and therefore there is only one linearly independent vector involved. In this case, eqn (2) can be written (omitting constant and lag terms):

$$\Delta X_t = \alpha_1 X_{t-1} + \alpha_1 \beta Y_{t-1} + e_{1t} \;\; \equiv \;\; \alpha_1 \left(X_{t-1} + \beta Y_{t-1} \right) + e_{1t}$$

$$\Delta Y_t = \alpha_2 X_{t-1} + \alpha_2 \beta Y_{t-1} + e_{2t} \;\; \equiv \;\; \alpha_2 \left(X_{t-1} + \beta Y_{t-1} \right) + e_{1t} \,,$$

so that the 'unique' cointegrating relation is " $X_{t-1} + \beta Y_{t-1}$ " with cointegrating vector: $(1 \quad \beta)$. Hence the linear combination:

$$X_{t-1} + \beta Y_{t-1} = \begin{cases} \dfrac{1}{\alpha_1}\left(\Delta X_t - e_{1t}\right) \\ \dfrac{1}{\alpha_2}\left(\Delta Y_t - e_{2t}\right) \end{cases}$$

is <u>stationary</u> since the ΔX_t and ΔY_t and e_t are stationary themselves. Each row of Π therefore represents the same cointegrating vector. We can now make the analogy with test 1 above where stationarity in the regression residuals was the criterion for cointegration.

- Note that if the rank of Π were equal to its dimension, 2, then it's determinant will be non-zero and Π^{-1} exists. Pre-multiplying the VAR-ECM (eqn 2) with Π^{-1} gives (omitting constant and lag terms):

$$\Pi^{-1}\begin{pmatrix} \Delta X_t \\ \Delta Y_t \end{pmatrix} = \begin{pmatrix} X_{t-1} \\ Y_{t-1} \end{pmatrix} + \begin{pmatrix} e'_{1t} \\ e'_{2t} \end{pmatrix}.$$

Since the left hand side is stationary, all terms on the right must also be stationary. This implies there cannot be a cointegrating relationship between X_t and Y_t since both must already be stationary. As discussed above, two stationary series can never be cointegrated (i.e. exhibit a long run dependence).

- In practice, elements of the cointegrating matrix Π are estimated by fitting eqn (2) to 'stochastic' time series data and therefore, the rank of Π needs to be estimated by testing for its number of <u>*significantly*</u> <u>non-zero</u> eigenvalues. This is known as Johansen's (1988) procedure and essentially involves testing for a null hypothesis of r =1 cointegrating vectors [where r = rank(Π)] against the alternative of r = 2. For more information on the specific tests, please see articles in (Subject: Cointegration notes/articles..; Database: Time Series Analysis WDB; Author: Frank Masci; Created: 14/06/2006; Doc Ref: FMAI-6QR3PB). To carry out this test, we use the "ca.jo" function in the "urca" R package.

- The two cointegration tests above are most powerful when series spans are reasonably long (say N >~ 100 observations) and if they have not been modified too much (ie. by the process of seasonal adjustment). Also, the same limitations that plague the unit-root tests above apply: presence of structural breaks/impacts, the role of the deterministic trend and specification of the maximum lag order.

<u>Consistency Tool Outputs:</u>

The following diagnostics are generated from the sequence of tests:

(i). The ADF and Phillips-Perron unit root tests are performed on each series in succession. Following the output test statistics, a summary of these tests is given by red lines prefixed by "***":

e.g. if either the ADF or Phillips-Perron tests indicate that both series are integrated (non-stationary), you will see:

*** According to either test, series 1 = integrated I(1) process [at >p% level]
*** According to either test, series 2 = integrated I(1) process [at >p% level]
(where p=5 or 10)
or if both series are stationary you will see:
*** According to either test, series 1 ~ I(0) process [at <p% level]
*** According to either test, series 2 ~ I(0) process [at <p% level] (where p=1 or 5)
or you may see a combination of these if one series is integrated and the other stationary.

(ii). Then there is a statement reporting whether both series are integrated with the same order (zero included). If so, you will see for example:
*** Series have same integration order: [1 1]
otherwise, if the series are not integrated with the same order, you will see, for example:
*** WARNING: series pair not integrated with same order: [0 1]

This is serious since it implies that the input series may not be related at all. The prior assumption of a 'conceptually related' series pair is invalidated and must be brought to the client's attention. Please review the diagnostics and plots for marginal cases, e.g. if outlier/break effects are skewing the results so that series can still be declared to be integrated with the same order.

(iii). If both series are integrated with the same order (zero included!), then one series is regressed on the other using robust ('outlier resistant') regression. Actually for series with multiplicative decomposition models, the logs of the series are regressed on each other to ensure stationarity in the variance of the residuals (however this requirement is not important; see discussion above). Slope, intercept and correlation coefficient estimates are reported. Also, it is worth mentioning that if both series have the same integration order, then plots of regression residuals versus time and its ACF are also generated. These can be viewed by clicking on the "* TEST 6: UNIT ROOT & COINTEGRATION TESTS: PLOTS..." title link.

(iv). If both series are integrated with the same non-zero order, i.e. if both are I(1), then we perform the two cointegration tests: the "Augmented Engle-Granger" test (TEST1), and Johansen's VAR-ECM procedure (TEST 2). Following some output test statistics, the outcome of both tests is:

*** According to either AEG or Johansen tests, series pair is cointegrating
or
*** According to both AEG and Johansen tests, series pair is not cointegrating

If the series are not cointegrating (but both are indeed I(1)), then it also implies that the input series may not be conceptually related in the long-run. In other words, there are no common long-term trends. This should be brought to the the client's attention. The assumption of relatedness should be questioned. Please review the diagnostics and

plots for marginal cases.

If both series were not integrated with the same non-zero order (from results of unit root tests above), then you would see:
*** Series pair not cointegrating since components not integrated with same non-zero order

2. Which Series Component and Why?

After much testing, discussion and debate, it has been decided that three types of series components will satisfy all our needs on assessing consistency: originals ("O"), standard ABS seasonally adjusted ("SA"; i.e. with priors included), and, X11 "prior-corrected" seasonally adjusted (= D11 table output labelled "pSA"). These will allow us to quantify the accuracy of different assumptions and methodologies. The specific questions that can be addressed using the above tests on each of these components are as follows:

Originals: {O_1, O_2}

1. Are the series conceptually related? If not, the client should be notified that the series are not consistent or related on conceptual grounds. Is it worth the effort to proceed with the consistency check?
2. Have any revisions in the data destroyed any prior association (relatedness) for a series pair?
3. Do the raw series have different seasonal patterns, strengths? If so, this does not imply inconsistency!
4. Do the raw series differ in any other systematic calendar-related effects?
5. Do the raw series have different outlier, break strengths or numbers thereof?

ABS Seasonally Adjusted: {SA_1, SA_2}

1. Like above: are the series conceptually related ("minus" seasonality)?
2. What are all the discrepant priors (outliers, breaks, calendar related effects) between the two series?
3. Was seasonality adequately and consistently removed from each series?

(D11) "Prior-Corrected" Seasonally Adjusted: {pSA_1, pSA_2}

1. Like above: are the series conceptually related?
2. Were all prior-corrections consistently and correctly applied to each series? This includes all calendar-related effects.
3. Are there any remaining, discrepant priors not accounted for between the two each series?
4. Was seasonality adequately and consistently removed from each series?

- *It appears that the* "pSA" *components are the most useful and efficient to assess consistency, simply because they effectively represent our "end product" with priors and seasonality removed. Any remaining priors that are discordant between these series (outliers in the movement vs. movement plots - TEST 4) would not have been accounted for in your nominal, initial list of priors, therefore allowing for easier identification.*

- *However, "pSA" (D11) components are not yet available for downloading* *using "TSA-download" from which the output data format is needed for the consistency tool. Nonetheless, you are not at a loss. It is advised that you use "ABS SA" components instead. The only difference is that the movement vs. movement plots (as in TEST 4) will contain all discordant priors between the two series, regardless if they've been corrected or not as specified in your nominal, initial list of priors. Therefore, you will just need to ensure that all the discordant priors are represented in your initial list of priors.*

- *For your information, diagnostic test results will be generated for all of the above three series components in the final SEASABS version.*

3. Recipe: what constitutes a "Pass" or "Fail" and how do I "fix" a "Failure"?

- Disclaimer: our goal is to achieve the best possible consistency between a pair of "prior-corrected" (D11) seasonally adjusted series by accounting for all priors and systematic differences as best we can. This is to be achieved at a level where any residual differences are purely stochastic in nature.

- The main output test diagnostics were outlined in red in section 1 above. For concise reporting purposes (to be implemented in a future SEASABS), we would like to assign a "PASS" (P) or "FAIL" (F) flag to each test. Note that all tests above don't treat the word "FAIL" in the same context. For example, if series of a pair are found to be totally unrelated (i.e. TEST 2 "failure") then it's pretty serious, while if the movements in series 1 are overall different from those in series 2 (i.e. TEST 3 "failure") then it's really just a warning that should be recorded (and forwarded to the client).

- Also, a "Pass" or "Fail" must never be taken as the final truth since there may be borderline cases (e.g. that marginally satisfy critical p-values) that can only be decided upon by examining all diagnostics and plots. Broadly speaking, below we define the "Pass" and "Failure" modes for each test and possible solutions to the "Failed" (inconsistent) cases. These will evolve as this tool is put into practice.

- *The below assumes you have first generated consistency summary tables and diagnostics using either "ABS seasonally adjusted", or, "prior-corrected (D11) seasonally adjusted" series components. At the time of writing, only "ABS seasonally adjusted" components are available for downloading via "TSA-download". Instructions on how to execute the prototype consistency tool are in:* [Subject: Coherence and Consistency (draft in progress) (Frank); Database: Time Series Analysis WDB; Author: Lisa Apted; Created: 26/07/2005; Doc Ref: LAPD-6EN5PZ].

- **TEST 1:**

- Pass: If all the "main" processing parameters and settings are consistent (only checked if periodicities the same).
- Fail: If any of the "main" processing parameters or settings are inconsistent (only checked if periodicities the same).
- Solution to Failure: Go back and fix the inconsistent parameter settings in

SEASABS if you think they should all be the same - i.e. if there are no justifiable reasons why they should be different. An example of a justifiable reason on retaining different seasonal adjustment parameters is the SMA filter length if the series have grossly different spans or levels of volatility.

- **TEST 2:**

- Pass: If the zero-lag cross-correlation coefficient is significantly different from zero.
- Fail: If the zero-lag cross-correlation coefficient is consistent with zero.
- **Solution to Failure:** Series movements and hence the series in original terms are outright unrelated and inconsistent. Not much you can do except notify the client if they really expected consistency. If so, then do your best to address any failures in TESTS 1, 4 and 5.

- **TEST 3:**

- Pass: If the mean difference in movements is consistent with zero.
- Fail: If the mean difference in movements is significantly different from zero.
- **Solution to Failure:** Is this due to a non-unit slope (=> movements related by scale factor), non-zero intercept (=>movements related additively by an offset) or both? There's not much you can do to fix this. You might just want to report it to the client for their information. This may have theoretical economic implications!

- **TEST 4:**

- Pass: If no outliers were found from the movement vs. movement plot.
- Fail: If any outliers were found from the movement vs. movement plot.
- **Solution to Failure:** This usually implies a large extreme, outlier, break or some deviant present in one series but not the other. If they are significant (i.e. relatively large), then make sure they are represented in your nominal list of priors in the summary table. It is recommended that you compare either the seasonally adjusted or original series components at the outlying timepoints to determine which series in the pair needs the correction. If a prior is not represented in your nominal list, then set them explicitly in SEASABS, seasonally re-adjust and re-run the consistency check.

- **TEST 5:**

- Pass: If no significantly non-zero ACF & PACF values were reported implying no evidence for a temporal pattern in movement differences.
- Fail: If significantly non-zero ACF & PACF values were reported implying evidence for a temporal pattern in movement differences at the reported time-lags.
- **Solution to Failure:** If a significant temporal pattern is found at a seasonal annual lag (4 for quarterly; 12 for monthly) and if you are looking at "seasonally adjusted" components, then it implies that residual seasonality is present in one series relative to the other. It is advised that you peruse the seasonal factors and/or S*I charts and ensure that seasonality is eradicated as much as possible in the seasonally adjusted components. A pattern at a non-seasonal (non-annual) lag would be intriguing and suggests the presence of autocorrelated behaviour on some underlying business cycle.

- **TEST 6:**

- Pass: If **(i)** both series are integrated with the same order (zero included), i.e.

both are I(d) where d >= 0 and **(ii)** if the series pair is cointegrating according to some significance level.
- Fail: If either of these conditions are not met.
- **Solution to Failure:** A failure in either (i) or (ii) implies that the prior assumption of a 'conceptually related' series pair is not a good one. The series don't track each other in the 'long-run' (i.e. trends not common over long spans). Please first peruse the diagnostics and plots for marginal cases (e.g. outliers/breaks skewing the results). You might want to bring this to the client's attention and ask whether they really expected 'long-term consistency'. This may have unexpected theoretical economic implications!