Quantifying Consistency between Series

Frank Masci

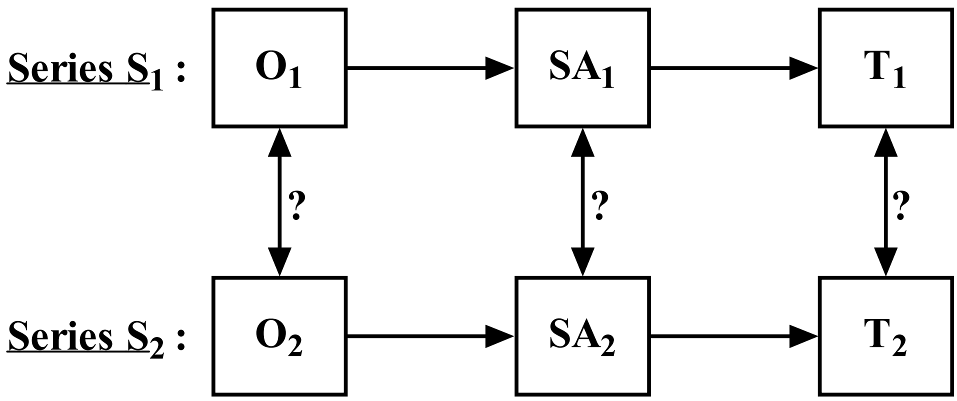
Version 1: 20/12/2005

1. Overview

The goal here is to devise a generic measure to quantify the degree of consistency between any two series, either in original terms, their seasonally adjusted components, or both. This document has been written to encourage discussion and feedback.

The subject of exploring and maintaining consistency between related ABS series is an ongoing project and has been covered extensively. See the doclinks under the original proposal document: (Subject: Coherence - project proposal; Database: Time Series Analysis WDB; Author: Tom Outteridge; Created: 05/09/2005).

Suppose we have two series $\mathbf{S}_1$ and $\mathbf{S}_2$. In the big scheme of things, we would like to check for consistency between the (raw-unprocessed) originals, $\mathbf{O}_1$ and $\mathbf{O}_2$; the seasonally adjusted estimates, $\mathbf{SA}_1$ and $\mathbf{SA}_2$; and the final trend estimates, $\mathbf{T}_1$ and $\mathbf{T}_2$. This is illustrated as follows.



Consistency across all three steps is what one would hope to acheive. If the originals are <u>inconsistent</u> (in the sense that the consistency is "statistically insignificant"; see below), then there's no reason to expect that the seasonally adjusted and trend components are consistent. However, if the originals are consistent but the seasonally adjusted components are not, then this is a problem and one will need to explore the source of inconsistency.

A consistency check in different steps of the above chain can be used to quantify the accuracy of assumptions and methodologies regarding:

(i) consistency of inputs (originals);

e.g: sources of inconsistency can be due to adjustments, tweaks and revisions to raw estimates made by the SMA, possibly due to constraints from benchmarks and accounting/economic standards; changes in periodicity - monthly to quarterly; collection and sampling methods.

(ii) consistency in the seasonal adjustment process;
e.g: sources of inconsistency can be due to different parameter settings - outlier thresholds, SMA and TMA filters; seasonal and trend breaks, moving holidays, trading day, different time spans, seasonal adjustment models (multiplicative, additive) and methods (concurrent, forward-factor, forecasting, direct or indirect).

(iii) consistency in final trend estimation;
e.g: sources of inconsistency could be a consequence of those in (ii) above, but more specific to trend estimation we have Henderson Moving Average properties and all prior factors - trend breaks and large extremes in particular.

2. Consistency Measures

A measure of the consistency between two time series along the above processing chain will need to take into account the stochastic nature of the observed series. In other words, every series is a realisation selected from a multitude of possible values at every time point. Given any two series, it therefore makes sense to consider the statistical significance of the proposed consistency measure. We propose two classes of stochastic measures: (i) cross-correlation methods; and (ii) methods based on point-to-point (%) movements.

(i) cross-correlation methods

Suppose we have two series $\mathbf{S_1}$ and $\mathbf{S_2}$ with N datapoints from each paired off at the same timepoints $t = 0 \ldots N - 1$: $\{\mathbf{S_{1t}}, \mathbf{S_{2t}}\}$. Note that these series can represent any of the components in the above processing chain: O, SA or T. We first make each series stationary by taking first differences:

$$\Delta_1(t) = S_1(t) - S_1(t-1)$$
$$\Delta_2(t) = S_2(t) - S_2(t-1).$$

The reason for this is to remove (or at least minimise) artificial large scale correlations that could arise from "similar" time-dependent means in the two series. We are primarily interested in correlating local structure and movements between the series.

We define the "linear cross-correlation coefficient" at lag h between the first-differenced series:

$$\rho(\Delta_1,\Delta_2)_h = \frac{\operatorname{cov}(\Delta_1,\Delta_2)_h}{\sigma_{\Delta 1}\sigma_{\Delta 2}} = \frac{\sum_{t=0+h}^{N-1}\left[\Delta_1(t)-\langle\Delta_1\rangle\right]\left[\Delta_2(t-h)-\langle\Delta_2\rangle\right]}{\sqrt{\sum_{t=0+h}^{N-1}\left[\Delta_1(t)-\langle\Delta_1\rangle\right]^2}\sqrt{\sum_{t=0}^{N-1-h}\left[\Delta_2(t)-\langle\Delta_2\rangle\right]^2}} \tag{1}$$

where

$h = 0, \pm 1, \pm 2, \ldots$

$$\langle\Delta_1\rangle = \frac{1}{N-1-h}\sum_{t=0+h}^{N-1}\Delta_1(t)$$

$$\langle\Delta_2\rangle = \frac{1}{N-1-h}\sum_{t=0}^{N-1-h}\Delta_2(t)$$

The quantity in equation (1) has the range $-1 \le \rho \le 1$, and values $\rho > 0$ indicate "positive" correlation, irrespective of the magnitude of the slope between the **{**Δ_{1t}, Δ_{2t}**}** and their relative magnitude (see "Issues" section below). Positive, statistically significant values of ρ can therefore be taken to indicate some level of consistency between the series.

Parametric Test:

The statistical significance of ρ can be assessed depending on the number of overlapping data points, *N* (corrected for any non-zero relative lag). We define our null and alternative hypotheses to be:

$$H_0 : \rho(\Delta_1,\Delta_2)_h = 0 \quad \text{versus} \quad H_1 : \rho(\Delta_1,\Delta_2)_h > 0.$$

Note that ρ is ignorant of the individual distributions for Δ_1 and Δ_2 and it is difficult to compute or assume a distribution under $\mathbf{H_0}$. As a rough and ready guess, we can assume that the distributions of Δ_1 and Δ_2 jointly form a bivariate normal distribution, a reasonable approximation at least not too far out in their tails. Thus as a starting point, we propose a parametric test. There are two limiting cases:

If N is large, say typically N > 500 and if the "tails" of the distributions for Δ_1 and Δ_2 die off sufficiently rapidly, then under $\mathbf{H_0}$, ρ **~ N(0, 1/N)**. Therefore, the critical (one-sided) value above which the observed value of ρ can be considered significant (at the 5% level) is $\rho_{crit} = 1.96/\sqrt{N}$.

If N is small to moderately large, it turns out that the statistic:

$$t = \rho\sqrt{\frac{N-2}{1-\rho^2}} \tag{2}$$

is distributed under $\mathbf{H_0}$ like Student's *t* - distribution with *N* - 2 degrees of freedom (Press et al. 1992 and references therein). Note that as *N* becomes large, this statistic becomes normally distributed and ρ **~ N(0, 1/N)** as above. Thus, one will never do worse by using equation (2), even if the assumption of a bivariate normal is not a good one.

Non-parametric Tests:

If we are concerned with the assumption of an underlying bivariate-normal distribution for the joint distribution in the differenced-series data: $\{\Delta_{1t}, \Delta_{2t}\}$, i.e. if we think outliers or other artifacts in the series can significantly skew this assumption, then we can use one of the more robust *non-parametric* tests: e.g. *Spearman Rank-Order correlation coefficient,* or *Kendall's Tau* .

These tests are based on ranking the data in Δ_1 and the data Δ_2 separetely. There is some loss of information in replacing the data by ranks, but this method is more robust to unplanned effects in the data (e.g. outliers). If a non-parametric correlation is significant, then a parametric test will certainly always find it significant (with much higher significance as a matter of fact). This makes non-parametric tests more stringent. For the formalism on how these tests are constructed, see any good statistics book.

(ii) point-to-point movement methods

For any two “potentially consistent” series $\mathbf{S_1}$ and $\mathbf{S_2}$, we define the fractional (or relative) point-to-point movement in each series as follows:

$$M_1(t) = \frac{S_1(t) - S_1(t-1)}{S_1(t-1)} \approx \log_e S_1(t) - \log_e S_1(t-1)$$
$$M_2(t) = \frac{S_2(t) - S_2(t-1)}{S_2(t-1)} \approx \log_e S_2(t) - \log_e S_2(t-1), \tag{3}$$

where the approximations are good to < 4% for movements $\mathbf{M_1}$, $\mathbf{M_2}$ < 30%. Like above, the series $\mathbf{S_1}$ and $\mathbf{S_2}$ can be any one of the components O, SA or T from the above processing chain. Also, multiplying (3) by 100 gives the percentage movement from t - 1 to t. We shall stick to “fractional” movements in what follows. As a measure of consistency between $\mathbf{S_1}$ and $\mathbf{S_2}$, we can compare the distribution in $\mathbf{M_1}$ values at all time-points to those of $\mathbf{M_2}$ and apply a statistical test to determine if they were drawn from the same population distribution. In other words, our null hypothesis is that the distribution in $\mathbf{M_1(t)}$, $\mathbf{D(M_1)}$, is consistent with the distribution in $\mathbf{M_2(t)}$, $\mathbf{D(M_2)}$:

$$H_0 : D(M_1) = D(M_2) \quad \text{versus} \quad \mathrm{H}_1 : D(M_1) \neq D(M_2) \tag{4}$$

Failure to disprove the null hypothesis only shows that the series in relative movements: $\mathbf{M_1(t)}$ and $\mathbf{M_2(t)}$ are consistent. One can then only conclude that the input test series $\mathbf{S_1}$ and $\mathbf{S_2}$ are consistent if one doesn’t care about possible differences in their trend levels. If the overall movement distributions are consistent, then it's very likely that the parent series are related in some way (e.g., CPI versus CVM measures, a state and national aggregate, or monthly versus quarterly, etc…). This assumption makes consistency checks between series pairs very general and is essential, given the myriad of flavours and units of series we wish to compare.

The null hypothesis defined by eq. (4) is very general and we can be more specific on the "level of consistency" we wish to acheive. We introduce the concepts of "weak consistency" and "strong consistency". Weak consistency is where we require the series to move consistently in the same direction (i.e., same signs in eq. 3), irrespective of the relative magnitudes in the movements. Strong consistency (as presented by eq. 4) is more stringent, and this is where we require consistency in both movement direction and magnitude. The latter imposes too much of a strong constraint and will miss a majority of series pairs that we regard as being consistent regardless of their magnitude (e.g. level in an aggregation structure; see "Issues" section). A series pair which satisfies the "weak" consistency principle are very likely to be related in some way.

The "strong" consistency hypothesis (as presented by eq. 4) will be more difficult to test. We will need a statistical test which is sensitive at detecting differences in volatility and extreme events in the tails of the distributions $\mathbf{D(M_1)}$ and $\mathbf{D(M_2)}$. This is because differences in the location of trend breaks or large extremes in two series will manifest themselves as unusually large movements (i.e., outliers). A test which is only sensitive around the median may miss "inconsistency effects" that only leave a signature in the tails. Furthermore, the distributions $\mathbf{D(M_1)}$ and $\mathbf{D(M_2)}$ are unlikely to be normal since $\mathbf{M_1}$ and $\mathbf{M_2}$ involve ratios of quantities which are correlated and whose distributions are unpredictable and non-standard. Thus, a parametric statistical test may be difficult to formulate for the "strong" hypothesis. The most general tests are probably:

- Kolmogorov-Smirnov (K-S) test, or one of its more robust variants: e.g. the Anderson-Darling or Kuiper statistic. This will test for differences in means, variances and distribution shape in general.
- "Mann Whitney U test" on the absolute values of the movements (since this statistic only tests for differences in medians, hence "volatility" and magnitudes in $\mathbf{M_1}$ and $\mathbf{M_2}$) and combine it with a "runs-in-signs test" to test for consistency in movement direction (see below).
- Binomial (sign) test – tests the null hypothesis that Pr($\mathbf{M_1} > \mathbf{M_2}$) = 0.5. i.e., that both $\mathbf{M_1}$ and $\mathbf{M_2}$ are equally likely to be larger than the other and is consistent with a binomial distribution. If this hypothesis is rejected, then the two parent series are declared "inconsistent".

The "weak" consistency hypothesis is based on ensuring "consistency in movement direction" and is much simpler to test. We can even get away with using a *parametric* test. The null and alternative hypotheses are defined as follows:

$$H_0 : \mathrm{sgn}(M_1) \neq \mathrm{sgn}(M_2) \quad \text{versus} \quad \mathrm{H}_1 : \mathrm{sgn}(M_1) = \mathrm{sgn}(M_2) \qquad (5)$$

Broadly speaking, we test the hypothesis that the signs of $\mathbf{M_1(t)}$ and $\mathbf{M_2(t)}$ are overall consistent by attempting to reject the null hypothesis that they are due to a random process. This can be accomplished by using a **"run-in-equal-signs"** test (or more generally a "runs test"). Here, one tests the null hypothesis that the number of "runs" of equal sign in the movements are consistent with randomness. If so, then this rejects the alternative hypothesis that the movement directions are consistent.

The "runs test" is particularly attractive for its simplicity and we expand it further. For movements at a given time t, we assign a "1" (yes) if $\mathbf{sgn[M_1(t)]}$ = $\mathbf{sgn[M_2(t)]}$ and a "0" (no) otherwise. For all times, we therefore end up with a sequence of 1's and 0's, e.g. 111110001111000011111111110. Is this sequence consistent with randomness? Lucky for us, the hard work has been done by Wald & Wolfowitz (1940). Let N_0 and N_1 represent the number of 0's and 1's respectively. The 'null' distribution for the number of runs N_R (number of consecutive sequences of either 0's or 1's) is approximately normal when the total number of observations $N = N_0 + N_1$ is relatively large (say, $N > 20$):

$$Z = \frac{N_R - \mu(N_R)}{\sigma_{N_R}} \sim N(0,1) \qquad (6)$$

where

$$\mu(N_R) = \frac{N_0 + N_1 + 2N_0N_1}{N_0 + N_1},$$

$$\sigma_{N_R} = \sqrt{\frac{2N_0N_1\left(2N_0N_1 - N_0 - N_1\right)}{\left(N_0 + N_1\right)^2\left(N_0 + N_1 - 1\right)}}$$

Therefore if $Z > 1.96$, we <u>reject</u> the null hypothesis (at the 5% significance level) that the observed number of equal-sign runs is consistent with randomness in favor of the alternative hypothesis that the signs (hence movement directions in the two series) are consistent. The latter implies that the two series are somewhat related.

Another test of the <u>"weak" consistency hypothesis</u> could involve testing for a significantly positive correlation between the actual movements. This is similar to the "cross-correlation method" involving stationarised series proposed above, i.e., we would test the following:

$$H_0 : \rho(M_1, M_2)_h = 0 \quad \text{versus} \quad H_1 : \rho(M_1, M_2)_h > 0 \qquad (7)$$

If two series move consistently in the same direction, then their movements are expected to be correlated. As discussed above, this measure does not depend on the differences or ratios of the movement magnitudes (i.e. a correlation between two variables does not imply that a straight line fit between them has unit slope and zero intercept). This is a good thing since it allows us to probe a larger "consistency space" as opposed to enforcing strict equality in movements as a measure of consistency.

3. Issues

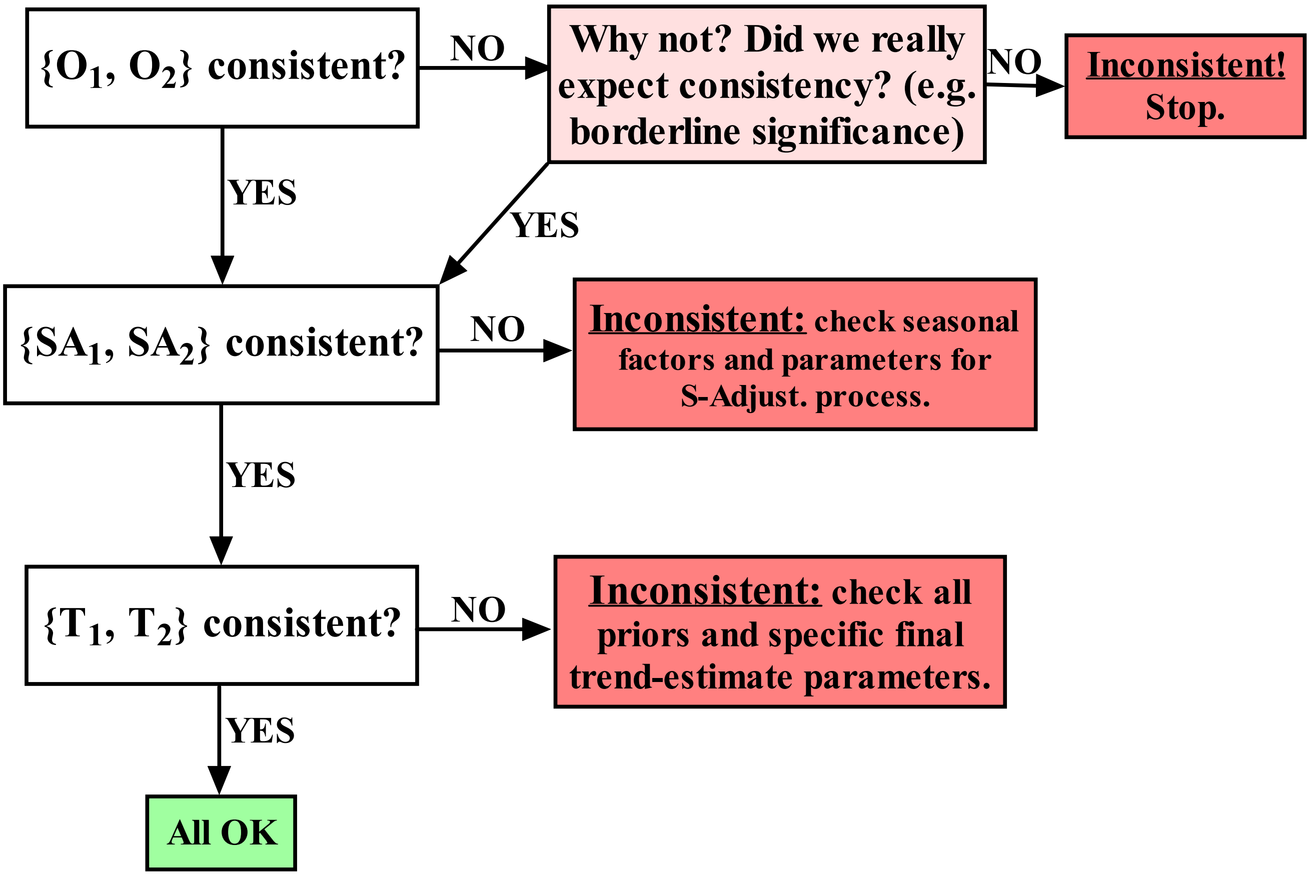
- The consistency measures above are <u>ignorant</u> of relative trend level differences between the input series. They only quantify similarities in structural changes and patterns. If their structure is consistent, then it's very likely that the parent series are related in some way (e.g., CPI versus CVM measures, a state and national aggregate, or monthly versus quarterly, etc…). This assumption makes consistency checks between series pairs very general and is essential, given the myriad of flavours and units of the series that we wish to compare.
- The correlation measures above require that we pair off values from $\mathbf{S}_1$ with those from $\mathbf{S}_2$ at equal times. This will present a problem when their periodicity (sampling) is different, e.g. monthly versus quarterly series. We can circumvent this by smoothing (averaging) over the monthly data to derive quarterly estimates, or, we can just use values for the middle month of each quarter from the monthly series for comparison with the input quarterly series. We will want to use the same quarterly-series derivation method as was used by the subject matter area (e.g. stock versus flow series).
- The input series may span different time ranges. It is expected that the input series will have some overlap and that this overlap is appreciable in order to get a robust correlation measure. The minimum overlap required is uncertain at this stage.
- It's not yet clear which of the above methods is more robust or sensitive at detecting inconsistencies between seasonally adjusted (or trend cycle) components that may arise from differences in the prior-corrections mentioned in section (1). E.g. inconsistencies due to the presence and non-presence of a trend break might be more easily picked up by methods which test for "strong consistency" (i.e. based on both magnitude and direction of movements) as opposed to "weak consistency" methods which only check for consistency in movement direction (overall movements have equal sign). One will need simulations to pick out the best method.

4. How do we use the Above Measures?

Let us assume here that we adopt a "cross-correlation method" as a measure of the overall consistency between the *stationarised* components (Δ_1 and Δ_2) of two series $\mathbf{S}_1$ and $\mathbf{S}_2$ with measure $\rho(\mathbf{S}_1, \mathbf{S}_2)$ at some lag h . This is just a working measure to assist in the discussion below. We could just have used a statistic from one of the "movement methods" (ii) above.

Given the three components for this series pair (see above processing chain): originals, $\mathbf{O}_1$ and $\mathbf{O}_2$; seasonally adjusted estimates, $\mathbf{SA}_1$ and $\mathbf{SA}_2$; and the final trend estimates, $\mathbf{T}_1$ and $\mathbf{T}_2$, we will want to quote three consistency measures: $\rho(\mathbf{O}_1, \mathbf{O}_2)$; $\rho(\mathbf{SA}_1, \mathbf{SA}_2)$; $\rho(\mathbf{T}_1, \mathbf{T}_2)$ along with their significance levels. As an aside, we may allow the user to modify the critical significance level for accepting "consistency" on each of these levels. Having computed these measures and their significance, we could then use the logic below at identifying (in)consistency for a series pair.

Note that by "consistent" at any level of processing, we mean it in the "statistically significant sense".



Points to note:

- Declaration of complete consistency between a series pair would entail satisfying all three levels in this flowchart. However, we may want to omit the final trend consistency check to speed up the process.

- In the first consistency check, if the originals are declared to be inconsistent at first go, then we suggest that the user peruse all three consistency measures to see if it's a borderline case (light-red coloured box), i.e. their significance may be close to the ciritical level(s). If so, or if the user is really adament that the series pair should be consistent, then they can change the consistency acceptance critical level(s) and rerun the overall consistency check.

- In the production environment, an "inconsistent" series pair can be indicated by a flag in the consistency table, highlighted by a different colour (with a different colour assigned according to the processing level above), or, we can just omit all the consistent series from the table output so it's easier to focus on the inconsistent cases.

5. Graphical Representations?

It would also be good to represent the above process graphically. Some ideas are:

- we can plot the (%) movements (eq. 3) in each of the three components for a series pair against each other. e.g. movements in $\mathbf{O_1}$ **versus** $\mathbf{O_2}$; $\mathbf{SA_1}$ **versus** $\mathbf{SA_2}$ and $\mathbf{T_1}$ **versus** $\mathbf{T_2}$. This will allow a user to see if a real correlation exists by eye and whether outliers are giving a misleading measure of the cross-correlation (and hence a false indicator of (in)consistency). This will result in three plots for a series pair.

- plot the (%) movements in each of the above three components versus time, just like in Lisa A's coherence charts. This will also result in three plots: e.g., movements in $\mathbf{O_1}$, $\mathbf{O_2}$ versus t; movements in $\mathbf{SA_1}$, $\mathbf{SA_2}$ versus t; and $\mathbf{T_1}$, $\mathbf{T_2}$ versus t.

- plot the cross-correlation in movements: $\rho(\mathbf{M_1}, \mathbf{M_2})_h$ at the six non-zero lags: h = -3, -2, -1, 0, 1, 2, 3 for each of the above three components. In a monthly series, these represent "business cycles" of interest. For a series pair, this can be represented as one plot.

- as a variation to the above, we can explore consistency on seasonal scales. e.g. for a monthly series, we take all Januarys across the whole span from series 1, then all the Januarys from series 2 and correlate them againt each other. We then repeat this process for each month until we end up with 12 correlation measures (one for each month) between the two series. We can then plot the measures $\rho(\mathbf{O_1}, \mathbf{O_2})$; $\rho(\mathbf{SA_1}, \mathbf{SA_2})$; $\rho(\mathbf{T_1}, \mathbf{T_2})$ versus month on one plot. The only limitation here is that we'll need at least 15 years of data in each series to reliably compute a correlation coefficient for each month. The benefit of this is that only some periods (months or quarters) could be (in)consistent between two series. This information is washed out in the methods above where we attempt to correlate the total series spans to derive a global consistency measure for a series pair.

6. Summary of popular methods and direction

This is nothing final. It only represents a glimmer of consensus picked up by the author from discussions with the team and refinements from Mark Zhang. The following four "consistency" diagnostics are proposed.

TEST 1 - "Equal Sign Test": As a first summary measure, check to see if a series pair move consistently in the same direction, i.e. overall movements have equal sign, irrespective of their magnitudes. This is very general but essential given the diversity of series we wish to test for consistency. This can be tested by comparing the signs of the movements at matched times and testing for the absence of structure in the run of sign differences:

- For example, suppose we have the sequence of movements for series 1 and series 2 at matched times: **{M_1(t)}, {M_2(t)}**, if sign[**M_1(t)**] = sign[**M_2(t)**] then we set an indicator variable **i(t)** = 1, or 0 if sign[**M_1(t)**] .NE. sign[**M_2(t)**].

- Given a sequence (for example) **i(t)** = 11111110111100111111111111, we can then test for complete uniformity (i.e. a structureless sequence of purely 1's) by computing the ACF at different lags and testing for:

$$H_0 : ACF[i(t)]_{lag\ h} = 1 \quad \text{versus} \quad H_1 : ACF[i(t)]_{lag\ h} = 0$$

- Rejection of the null hypothesis (at some significance level) implies that the series pair <u>do not move consistently</u> in the same direction.

- This test can be applied to movements in each of the three components: **{O_1, O_2}**; **{SA_1, SA_2}**; **{T_1, T_2}**.

- Plots will be generated showing **i(t) versus t** for each of these components and also its **ACF** for various lags with confidence limits.

<u>TEST 2 - "Cross-correlation Magnitude Tests":</u> Supplement the above with a cross-correlation measure between the movements in each paired component: **{O_1, O_2}**; **{SA_1, SA_2}**; **{T_1, T_2}**, i.e. we would compute ρ**(M_1, M_2)$_h$** for the interesting lags h = -3, -2, -1, 0, 1, 2, 3. These lags are of interest because they reflect patterns in business cycles. More specifically,

- <u>For lag h = 0:</u> since we expect series pairs to be "conceptually" consistent (on apriori grounds) <u>at lag 0</u> and since the maximum value of the absolute value of the cross-correlation coefficient is "1", we will be testing the following null hypothesis. This results in a stronger test for consistency.

$$H_0 : \rho(M_1, M_2)_{h=0} = 1 \quad \text{versus} \quad H_1 : \rho(M_1, M_2)_{h=0} < 1$$

Rejection of the null hypothesis (at some significance level) implies that the movement magnitudes in a series pair are not correlated or overall "inconsistent".

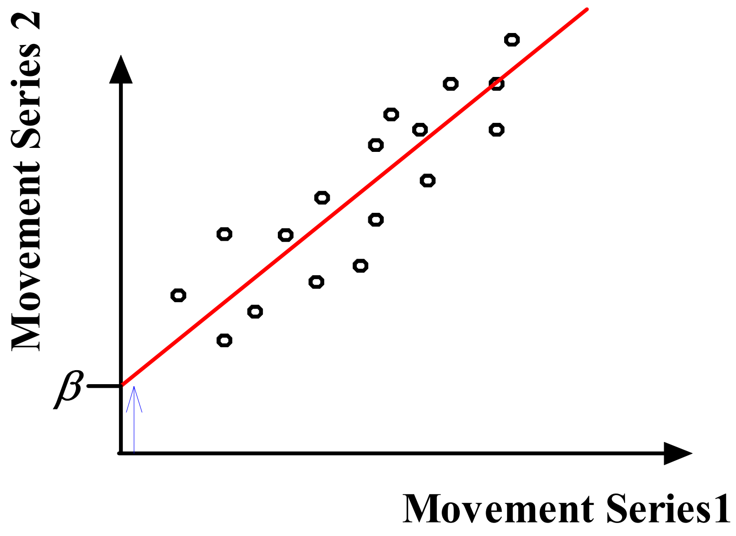
- <u>For lags h .NE. 0</u>, correlations are expected to be small (or insignificant most of the time). In this case we will test the following:

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

Rejection of this null hypothesis (at some significance level) implies that the movement magnitudes <u>at the non-zero lag</u> are correlated to some degree or overall, there is some level of "consistency" when the series are shifted with respect to each other.

- Probability (significance) levels will be computed for each correlation measure ρ**(M_1, M_2)$_h$** . These measures will be plotted against the lag h.

TEST 3 - "Equal Movement Magnitude Test": The above cross-correlation measures, on their own, are independent of any relative differences in the overall movement magnitudes. This can be seen by noting that a correlation between any two variables does not imply that a straight line regression fit has zero intercept and unit slope:



The straight line regression fit can be parameterised as:

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

where α is the slope and β is the intercept or distance from the horizontal axis along $\mathbf{M_1 = 0}$. A significantly non-zero value for β implies an overall difference (hence inconsistency) in the movement magnitudes for a series pair. Thus, we can test for the following:

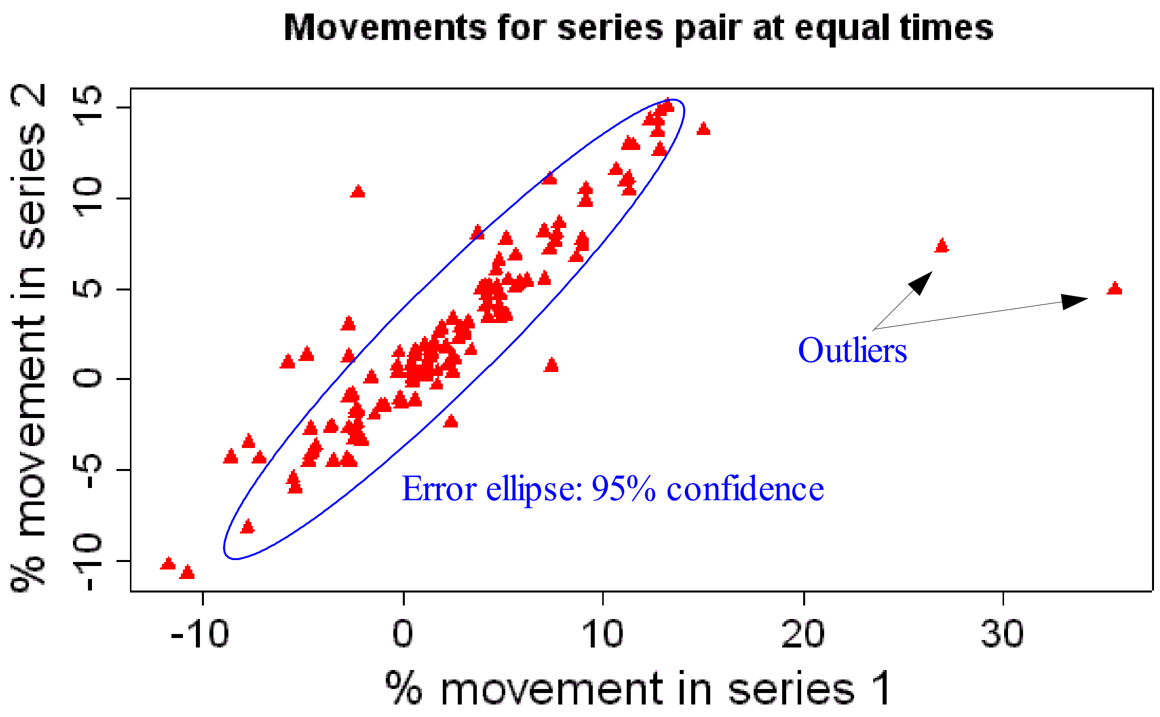
$$H_0 : \beta = 0 \quad \text{versus} \quad H_1 : \beta \neq 0$$

- Acceptance of the null hypothesis therefore implies that the intercept is consistent with zero and hence movement magnitudes are overall of the same magnitude.

- The significance for zero slope will come from the standard errors on the parameters of a linear regression fit.

TEST 4 - "Spot-Outlier Movement Test": The basis of this test involves searching for outliers in a two-dimensional movement versus movement plot (see figure below).

- More specifically, we will attempt to seach for outlying, incompatible movements at specific time points. These are represented by points lying outside some standard error ellipse defined by some pre-determined confidence level in $\mathbf{M_1}$ **and** $\mathbf{M_2}$ jointly:



- Outliers in this plot indicate unequal movement magnitudes and could arise from intrinsic differences in the original series (e.g., outliers, breaks etc..) or, if they occur in movements between the seasonally adjusted or trend component series, could indicate differences in prior correction factors.

- This test can be automated in the sense that if any outliers are detected, a flag is set to indicate that this test "failed" and that the user should review the plot.

- In the production environment (e.g., under SEASABS), the goal would be to click on an outlier in the movment plot and all prior factors and series values would pop up for the two series.

Tying it all together..?

- So, for a given pair of series that we suspect to be "consistent", we can apply the above four tests to potentially each of the component pairs:

  Originals: $\{\mathbf{O_1}, \mathbf{O_2}\}$;
  Prior-corrected B1 originals: $\{\mathbf{B1_1}, \mathbf{B1_2}\}$;
  Seasonally adjusted: $\{\mathbf{SA_1}, \mathbf{SA_2}\}$;
  Trends: $\{\mathbf{T_1}, \mathbf{T_2}\}$

The prior-correct B1 pair was just recently added to this list as it represents a different but important step in the processing chain. An inconsistency in this component pair implies a possible inconsistency in the prior corrections early on in processing.

- Consistency checks in each of the above component pairs can be used to quantify the accuracy of different assumptions and methodologies. E.g., revisions and conceptual differences (originals - good for the client), priors and corrections (B1 and Trends), seasonal factors and filter properties (SA) and trend estimation parameters/properties (Trends).

- Note that if the originals are out-right inconsistent, this does not necessarily mean that everything downstream is also inconsistent. Why?

- How do we represent all of the above in a "consistency summary" table? A suggestion is below.

- On the <u>group level</u>, we can have a table stored in the series knowledge that lists all potentially correlated series pairs (first column). Clicking on any of these entries will bring up a consistency summary table with priors and parameters.

- In the subsequent columns, we summarise information on the above four tests for each of series component. "P" stands for "pass" and under this column, we list the above tests 1,2,3 or 4 that <u>passed</u> in confirming consistency. "F" stands for "failed" and under this column, we list those tests that failed.

- Clicking on any of these "test" numbers will bring up results of statistical tests and associated diagnotic plots.

- The last column may list a quality flag that assesses the overall level of consistency for the series pair, i.e., a number which encompasses the results of all tests across all four series components. This is still to be devised and we may leave it out since the "red danger flags" will be represented by the "test" numbers in the "F" columns.

	Originals		B1's		SA's		T's		Final consistency quality flag?
	P	F	P	F	P	F	P	F	
pair 1	1, 2	3, 4	etc...	etc...					?
pair 2	2, 4	1, 2							
pair 3	1,2,3,4	.							
pair 4	3	1, 2, 3							
.									
.									
.									

A simple tool is currently under development for TSA to use under production. Example output is as follows:
S:\data\BOP\coherence\Frank'sCode\results1.html