Interpreting the analytical spreadsheets for the revisions analysis of retail trade volume for OECD countries and selected non-member economies

This document explains how users should interpret data contained in the 3 revisions analysis spreadsheets for retail trade volume provided for each country in the zip files at: http://www.oecd.org/document/7/0,2340,en_2649_34239_37088263_1_1_1_1,00.html. The period of this study relates to data published each month in the OECD Main Economic Indicators CD Rom from February 1999 to May 2006. Note that due to suspensions in the transmission of data for some countries during parts of the analysis period the results contained in the revisions analysis spreadsheets should be treated with caution for Austria, France, Italy and the Russian Federation.

1. Revision triangles spreadsheet

The first spreadsheet in the zip file for a country is labelled Country_Revision_Triangles.

1) Organization of the Revision triangles spreadsheet

The Revision triangles spreadsheet contains data for the level estimates, their month-on-previous-month [(M_t/M_{t-1}) -1)] or quarter-on-previous-quarter [(Q_t/Q_{t-1}) -1)] and year-on-year [(M_t/M_{t-12}) -1 or (Q_t/Q_{t-4}) -1] growth rates and monthly revisions to these.

The Revision triangles spreadsheet consists of 5 worksheets:

a. ***Levels***: This worksheet contains the level estimates for (seasonally adjusted) retail trade volume as published in each successive monthly edition of the OECD Main Economic Indicators CD Rom starting from February 1999 (later for some countries) to May 2006. The values in the following worksheets will be derived from the data contained in this worksheet;
b. ***MoM Growth Rates***: In this worksheet are the month-on-previous-month growth rates of the variable based on the level estimates (or quarter-on-previous-quarter growth rates in the case of quarterly data);
c. ***Revisions to MoM Growth Rates***: This worksheet calculates the difference between the month-on-previous-month (or quarterly-on-previous-quarter) growth rates for the same reference period in two successive monthly editions of the MEI. Each value in the lower block is printed in bold if it is different from the previous value in the same column – that is if a revision was made to the data point from one monthly edition of the MEI to the other.
d. ***YoY Growth Rates***: In this worksheet are the year-on-year growth rates of the variable based on the level estimates.
e. ***Revisions to YoY Growth Rates***: This worksheet calculates the difference between the year-on-year growth rates in the same way as described in c. above.

Each worksheet is divided into two blocks:

i. The first block of four rows of each worksheet give the reference period and summary values for the comparisons, i.e. first published estimates and estimates published 3 months, one year, and 2 years after the first publication in the case of monthly source data. For quarterly source data the values are first published estimates and estimates published 5 months, one year and 2 years after the first publication.

ii. In the second block of data, each row represents the month of publication in the MEI CD Rom as labelled in column 2 and contains the values of the variable that was published for each reference period at that time. The data in this block should in theory look triangular (i.e. one additional monthly data point of the series for each successive monthly publication in the MEI). However, they often take the form of a distorted triangle because of the periods in which more than one or no new data point was published. Of course where the

input data is of quarterly frequency a new data point is only likely to appear every third month of the MEI publication on average.

2. Revision analysis spreadsheet

The second and third spreadsheets in the zip file for a country contain the revisions analysis statistics. One spreadsheet analyses revisions for month-on-previous-month (or quarter-on-previous-quarter for Australia and New Zealand) growth rates and the other for year-on-year growth rates. Their respective names are Country_Revisions_Analysis_MoM.xls and Country_Revisions_Analysis_YoY.xls.

1) Organization of the Revision analysis spreadsheet

The Revision analysis spreadsheet gives the final statistics for the revision analysis, using the growth rates calculated in the Revision triangles spreadsheet. It consists of the following worksheets:

a. ***Revision Spreadsheet***: In each column of this worksheet are the growth rates for the first published estimates(labelled *P*), estimates published 3 months, 1 year, and 2 years later(labelled *M3*, *Y1*, and *Y2* respectively), and the latest estimates(labelled *L*). The reference periods in relation are given in column 2. The only difference for quarterly source data is that *M3* is replaced with *M5* (estimate published 5 months later).

b. ***Summary Statistics****:* This worksheet contains the summary statistics for all the various comparisons, which are extracted from the summary statistics tables in the *X_Y* worksheets.

c. ***'X_Y'****:* The comparison *X_Y* worksheets evaluates the revision that occurred between periods *X* and *Y*, e.g. the comparison *M3_P* is an analysis of the first published estimates and revision 3 months later. These worksheets contain the intermediate values necessary for the computation of all statistics, for which the relevant summary values are shown in the table of Summary Statistics, a bar chart of revisions(*X-Y)*, and a line graph of the time series *X* and *Y*.

2) Description of Summary Statistics

Summary statistics	
n	74
mean absolute revision	1.1871
mean revision (Rbar)	0.3034
st. dev(Rbar) - HAC formula	0.2193
mean squared revision	2.7642
relative mean absolute revision	0.2261
t-stat	1.3839
t-crit	1.9930
Is mean revision significant?	NO
Correlation between L and P	0.9583
Min Revision	-6.3
Max Revision	4.3
Range	10.6
% L > P	51.4
% Sign(L) = Sign(P)	95.9
Variance of L	30.1158
Variance of P	32.7492
UM %	3.33
UR %	7.77
UD %	88.90

Revision is defined as $R_t = L_t - P_t$ where L_t is the later estimate and P_t is the preliminary (or earlier) estimate. *n* is the number of observations.

The following are the relevant formulas for the main statistics included in the summary statistics tables.

- *mean revision*:

$$\bar{R} = \frac{1}{n}\sum_{t=1}^{n}\left(L_t - P_t\right) = \frac{1}{n}\sum_{t=1}^{n} R_t$$

- *mean absolute revision*:

$$MAR = \frac{1}{n}\sum_{t=1}^{n}\left|L_t - P_t\right| = \frac{1}{n}\sum_{t=1}^{n}\left|R_t\right|$$

- *relative mean absolute revision:*

$$RMAR = \frac{\sum_{t=1}^{n}\left|L_t - P_t\right|}{\sum_{t=1}^{n}\left|L_t\right|} = \frac{\sum_{t=1}^{n}\left|R_t\right|}{\sum_{t=1}^{n}\left|L_t\right|}$$

- *mean squared revision*:

$$MSR = \frac{1}{n}\sum_{t=1}^{n}\left(L_t - P_t\right)^2 = \frac{1}{n}\sum_{t=1}^{n} R_t^2$$

In order to test whether mean revision is significantly different from zero, we perform a *t* test with the test statistic

$$t = \frac{\bar{R}}{\text{st.dev(Rbar) - HAC Formula}}$$

where *st.dev(Rbar)-HAC Formula*, or the *heteroscedasticity and autocorrelation consistent standard deviation of mean revision* is defined as the square root of:

$$\text{var}(\bar{R}) = \frac{1}{n(n-1)}\left\{\sum_{t=1}^{n}\hat{\varepsilon}_t^2 + \frac{3}{4}\sum_{t=2}^{n}\hat{\varepsilon}_t\hat{\varepsilon}_{t-1} + \frac{2}{3}\sum_{t=3}^{n}\hat{\varepsilon}_t\hat{\varepsilon}_{t-2}\right\},$$

with $\hat{\varepsilon}_t = R_t - \bar{R}$.

Other statistics useful for evaluation of various aspects are

- *Range=Max Revision-Min Revision*

Max Revision: value of the highest revision
Min Revision: value of the lowest revision

- *% Later>Earlier:*

The percentage of observations where the later estimate is larger than the earlier estimate, i.e. revision is greater than 0.

- *% sign(later)=sign(earlier):*

The percentage of observations where the sign of later estimate and the sign of earlier estimate are the same.

Decomposition of the mean-squared revision

Mean squared revision measures the variance of revision based on a symmetric and quadratic loss function. *MSR* decomposed and divided by itself gives:

$$1 = UM + UR + UD$$

with

$$UM = \frac{\bar{R}^2}{MSR}$$

$$UR = \frac{\left(S_p - \rho S_l\right)^2}{MSR}$$

$$UD = \frac{(1-\rho^2)\,S_l^2}{MSR}$$

where S_p is the variance of earlier estimates, S_l the variance of the later estimates, and ρ the correlation between them.

UM is the proportion of *MSR* due to mean revision not being equal to zero. It is thus also known as mean error.
If we consider a linear regression model of the earlier and later estimates $L_t = \alpha + \beta P_t + u_t$, *UR* is the proportion of MSR due to the slope coefficient β being different from 0, or the slope error. *UD* is the disturbance proportion of MSR, i.e. the proportion of MSR that is not caused by systematic difference between earlier and later estimates.

Good preliminary estimates have low values of *UM* and *UR* and a high value of *UD*.

More detailed information on these statistics can be found in a paper written by Professor Tommaso DiFonzo, available at http://www.oecd.org/dataoecd/55/17/35010765.pdf . In particular Appendix A and B contain comprehensive technical details on the test statistic.