Revisions information in ONS First Releases

Graham Jenkinson and Nigel Stuttard
Office for National Statistics

The Office for National Statistics (ONS) has started to provide users with more information about revisions to time series in its First Releases. The changes will help users see the statistics in context to aid their interpretation. In addition to clearer identification of revisions and the reasons for them, the main change will be a new table in the background notes to releases giving summary information about past revisions to key indicators.

Introduction

Starting with the UK Output, Income and Expenditure First Release (25 February 2004), ONS will be providing users with more information in its First Releases about revisions to time series. This will help users interpret and use the latest estimates as they will know how estimates have been revised in the past. The new information will appear in a standard form in each First Release, accompanied by explanations of any special features if necessary.

These changes will contribute towards the implementation of the National Statistics Protocol on Revisions. Principle 3 of the Protocol says "A statement explaining the normal effect of revisions will accompany the release of all key outputs subject to regular revisions." The changes will put the revisions to key series in an historical context, will clearly identify revisions being made, and will explain the status of the figures and identify any planned future revisions.

Summary of changes

The main change will be a new table in the background notes to releases giving information about past revisions to key indicators. The table will usually be in this form for releases giving quarterly series. There will be a slightly different form for releases of monthly series.

		Revisions between first publication and estimates three years later	
	Value in latest period	Average over the last 5 years (bias)	Average over the last 5 years without regard to sign (average absolute revision)
Key indicator 'a'			
Key indicator 'b'			
Key indicator 'c'			

The second column shows the bias over the last five years – revisions are biased if they are consistently different from zero in either direction. The information in this column will be tested using a modified t-statistic to see whether there is statistical evidence that the bias is significantly different from zero for any of the indicators. (If the test is not significant this implies that the observed bias might have occurred by chance when there is no bias in the revisions.) The third column in the table shows the average absolute revision to the key variables over this period – this gives the average size of revisions over the last five years as an indication of the reliability of the latest figures.

The ONS recognises that revisions are only one indicator of statistical quality and revisions do not capture the complete picture. The revised data may itself, be subject to sampling or other sources of error. Other aspects of quality in statistics are set out in the Eurostat framework and include accuracy, timeliness, and robustness.

Other developments to implement the protocol on revisions include:

- a table, or tables, of revisions to the main series published in the release with an explanation of the reasons for them. These tables appear after the other tables and are designated with an R in their numbering
- in most cases, identification of the periods subject to revision in the current edition of the release
- an extra note to explain the status of the figures and planned future revisions
- in some cases a periodic article about the revisions performance of the indicators. For example an article about GDP revisions appeared in *Economic Trends*, December 2003.

When will the new material appear?

It is being introduced over the next four months and should be in all releases published after the end of June 2004.

Which releases will carry the new information?

All those which carry long runs of time series information.

Will the information be in other ONS publications?

The immediate priority will be its inclusion in First Releases. But other publications giving long run time series will follow where appropriate.

What is the modified t-statistic?

Revisions are considered to be biased if the mean revision is statistically significantly different from zero. A t-test compares the calculated bias in the series (that is, the mean revision) with the variability of the revisions to test whether the bias is significant. However, the standard t-test is based on the assumption that the revisions are independent of each other. This is not true for a time series as revisions made for one period may be associated with revisions made to previous periods. The modified t-test corrects for this lack of independence by adjusting the estimate of the variability of the revisions to take into account the serial correlation, that is the extent of the association between successive revisions. A technical description of the modified t-statistic and its calculation is given in the Appendix.

Appendix A

The modified t-statistic

For the purpose of the revisions information contained in the First Releases, the revision to a economic series, r_t, is defined as the difference between the initial estimate of that series and the updated estimate, made twelve months later (for monthly series) or 3 years later (for a quarterly series). The revisions to a series are considered to be biased if the mean revision is statistically different from zero. A modified t-test is used to test the significance of the mean revisions. The rationale for using a modified t-test is that successive revisions in a series may not be independent. If they are not independent (i.e. there is serial (or auto-) correlation in the revisions) the standard t-test would overstate the significance of the results.

We assume that the revisions fit a regression model of the form:

$r_t = \mu + \varepsilon_t$ for t=1 to n i.e. with no explanatory variable.

If the errors are thought to be serially correlated, they follow an autoregressive model of order one, AR(1):

$$\varepsilon_t = \alpha\varepsilon_{t-1} + u_t$$

where the u_t are independent and the auto-correlation coefficient, α, is between –1 and 1.

The standard t-statistic is:

$$t = \frac{\bar{r} - \mu}{\sqrt{\sigma^2 / n}}$$

where $\bar{r}$ is our sample mean revision, μ is the population mean revision (which we will assume is zero and test for as our null hypothesis), σ^2 is the variance and n is the number of observations.

Priestly (1981) has suggested that, where auto-correlation is present, the equivalent number of independent observations for estimating the mean should be reduced to:

$$n\frac{(1-\alpha)}{(1+\alpha)}$$

and thus the variance of the mean should be adjusted by increasing it to:

$$\frac{\sigma^2(1+\alpha)}{n\ (1-\alpha)}$$

Under these circumstances, our modified t-adjusted statistic will be:

$$t\text{–adj} = \frac{\bar{r}}{\sqrt{\text{adjusted variance}}} \quad \text{[using n* degrees of freedom]}$$

with the null hypothesis that the population mean is zero and n*, the equivalent number of independent observations for estimating the variance, is:

$$n\frac{(1-\alpha^2)}{(1+\alpha^2)}$$

The calculation follows the steps below:

1. **Calculate the sample mean**

$$\bar{r} = \frac{\sum_{t=1}^{n} r_t}{n}$$

2. **The coefficient α is estimated by $\hat{\alpha}$ where**

$$\hat{\alpha} = \frac{\text{Cov}(r_{t-1}, r_t)}{\sqrt{\text{Var}(r_{t-1})\,\text{Var}(r_t)}}$$

$$= \frac{\sum_{i=2}^{n}(r_{i-1}-\bar{r}_{t-1})(r_i-\bar{r}_t)}{\sqrt{\sum_{i=1}^{n-1}(r_i-\bar{r}_{t-1})^2\sum_{i=2}^{n}(r_i-\bar{r}_t)^2}}$$

$$\text{Where } \bar{r}_{t-1} = \frac{\sum_{i=1}^{n-1} r_i}{n-1} \text{ and } \bar{r}_t = \frac{\sum_{i=2}^{n} r_i}{n-1}$$

3. **The estimate for the variance of the sample mean is s where**

$$s^2 = \frac{\sum_{t=1}^{n}(r_t-\bar{r})^2}{n}$$

4. **The estimate of the adjusted sample variance is *s** where**

$$s^{*2} = \frac{s^2(1+\hat{\alpha})}{n(1-\hat{\alpha})}$$

5. **The adjusted degrees of freedom is *n** where:**

$$n^* = n\frac{(1-\hat{\alpha}^2)}{(1+\hat{\alpha}^2)}$$

6. **Calculate modified t-statistic, t-adj**

$$\text{t-adj} = \frac{\bar{r}}{s^*} \text{ [using n* degrees of freedom]}$$

7. **Compare the t-adjusted value with the critical t-value**

Compare the absolute t-adjusted value against the critical t value at 95 per cent significance (2-tailed) and reject the null hypothesis if |t-adj| > t-critical, i.e. if rejected the test statistic is statistically significant.