Assessing the efficiency of early release estimates of economic statistics

By Richard McKenzie, Elena Tosetto (OECD) and Dennis Fixler (United States Bureau of Economic Analysis)

Introduction

Accuracy and timeliness are two fundamental dimensions of statistical quality, but to some extent they are mutually exclusive. Therefore preliminary estimates published for many macroeconomic variables are frequently revised at a later stage, sometimes substantially. While the users of official statistics may consider revisions a nuisance that must be taken into account in their analysis, the producers consider revisions an inevitable consequence of the fact that timely data are often of lower accuracy and more sparse than more complete data that become available later. Nonetheless, both users and producers share a common interest in minimizing revisions without sacrificing the timeliness of the initial estimates or the accuracy of the later estimates.

Ideally, revisions should only represent the 'news' of incorporating more accurate and detailed information which becomes available for the compilation of the later estimates. So the real question is: can we evaluate whether successive revisions of the first estimate bring additional information, or whether they are due to the correction of earlier inaccuracies or bias, which does not add any useful information to economic analysts?

A statistical technique, known in economic literature as 'news' vs 'noise' analysis, can be used to answer this question, i.e. to give insights as to whether first published estimates are "efficient". Should early estimates be shown to be inefficient, it implies they contain systematic measurement error or 'noise' that could be eliminated in order to become a better forecast of the later published estimates. This "Statistics Brief" aims to explain the concept of 'news' vs 'noise' analysis and presents an approximate analysis of this concept for revisions to seasonally adjusted quarterly growth rates of GDP across eighteen OECD countries.

The goal of this article is to motivate both producers and users of economic statistics to understand the purpose of the 'news' vs 'noise' analysis and undertake more detailed studies of this kind using their own sources of officially published data. The example of the national statistics office of the Netherlands, as outlined later in this article, shows that this process has already started and can be successful.

Possibly a simpler way for the reader to understand the 'news' vs 'noise' analysis is to envisage that a revision of an economic variable may be of two different types. Either the revision is more related (i.e. correlated) to the earlier first estimate and therefore can be seen as 'noise' as revisions should ideally be independent of the earlier estimate, or the revision is more related to the later estimate and as such can be seen as 'news'. This analogy is almost an oversimplification but could help the reader grasp the basic premise.

How to determine whether revisions to economic variables reflect 'news' or 'noise'

Over the last few years the OECD has been very active in promoting the importance of performing revisions analysis within the international statistical community and has presented the results of three comprehensive revisions analyses for a range of OECD member countries and selected non-member economies for gross domestic product (GDP), index of industrial production (IIP) and retail trade volume (RTV). These analyses have been performed using data available in the OECD Main Economic Indicators Original Release Data and Revisions Database[1].

A key objective in the analysis of revisions from the perspective of producers of official statistics is to better understand the characteristics of the statistical compilation process, enabling potential problems to be identified and improvements to be made. One way to evaluate the quality of the early vintage estimates is to study the information content embodied in succeeding revisions. Basic summary statistics usually presented from revisions analyses, such as mean revision, look at the statistical pattern of the revisions and thereby focus on the distributional properties of revisions. For example, a mean revision that is statistically significant from zero could indicate a bias in the estimation process. Though mean, variance and assorted other commonly computed revision statistics provide information about the revision in a statistical sense, they do not provide inferences about how the information available is used and whether early estimates are compiled in an 'efficient' manner. Efficiency in this sense can be defined as ensuring that all available information at a particular time is being used in the most efficient way to compile an estimate of a macroeconomic variable. Having efficiently derived estimates implies that subsequent revisions are due solely to the incorporation of new information.

The concept of 'news' vs 'noise' analysis was first introduced in ground breaking articles by Mankiw and Shapiro in 1984 on revisions to money stock and in 1986 on revisions to Gross National Product in the United States. Put simply, when a revision to an economic variable is said to contain 'noise' it means that all or part of the revision does not contain any new information. Rather, the change in the estimate is due to systematic measurement error that could arise from several factors such as biased estimates due to estimation based on non-representative samples, use of non-optimal imputation methodologies for imputing missing data, etc. On the other hand, a revision is said to contain 'news' if it can be shown that the change in the estimate can be attributed to the incorporation of new information, implying the early estimates are efficient forecasts of the later ones. The incorporation of such new information from the ongoing flow of source data is important given the general goal of the sequence of estimates is to approach some "true" value.[2] This is also an important issue for users of the data, as they would expect that revisions are adding information to aid in their decision making processes, rather than confusing the picture by providing random changes in official estimates.

For determining whether revisions represent 'news' or 'noise' it is possible to use two methods: the correlation method and the regression method. The correlation method is somewhat more simplistic than the regression method. It gives informal evidence and can be seen as a preliminary

[1] This new database and the revisions analysis studies for GDP, IIP and RTV were presented in Statistics Brief No. 12, November 2006: http://www.oecd.org/dataoecd/46/48/37669085.pdf

[2] Underlying the estimation process is a targeted "true" value that derives from economic theory.

step to the regression method, although the two approaches are likely in most cases to lead to the same conclusions.

Correlation method

The correlation method is the original technique introduced by Mankiw and Shapiro, and is applied to the growth rate of a variable, such as GDP. First, we define for a reference period, denoted by t:

P_t = Preliminary or earlier estimate of growth
L_t = Later estimate of growth
$R_t = L_t - P_t$ = Revision to growth rate.

If we find that the correlation between the revision and the earlier estimate of growth over a given set of reference periods is statistically different from zero, we can conclude that the revision embodies 'noise'. That is, the earlier estimate could have been better in some sense because the correlation between the revision and the earlier estimate indicates that the estimate did not fully utilize the information available (i.e. the revision should be completely independent of the earlier estimate). For example, suppose there is a tendency for large initial quarterly growth rates to be revised down or falls to be subsequently revised up. This would be indicative of a systematic error in the estimation of the initial estimates and it would be reflected in a negative correlation between the revision and the first estimate. On the other hand, if the initial quarterly growth rate is an efficient ("rational") estimate, then the revision can be expected to be correlated with the later estimate. To appreciate the rationale for this statement, note that the correlation between two variables is linearly related to the expected difference between each variable and its mean. Suppose that the initial published growth rate is equal to the average growth rate over the analysis period, which means there is zero correlation between the initial estimate and the revision. Then, a positive revision will result in an above average growth rate in the later estimate and a negative revision will result in a below average growth rate in the later estimate. Therefore, a statistically significant correlation between the revision and the later estimate signifies the effective incorporation of the new information, i.e. that the revisions are 'news'. This finding would mean that there is merit in the revision process as it enables the estimates to capture new information and thereby allows the estimates to approach the true value.

Regression method

Regression techniques can also be used to determine whether revisions should be classified as 'news' or 'noise'.[3] Generally, one would form a regression equation:

$$R_t = \acute{\alpha} + \beta^* L_t + \varepsilon_t \qquad (1)$$

and test whether: $\acute{\alpha} = \beta = 0$, where ε_t signifies random error. If the parameters are found to be equal to zero, then one would conclude that the revision does not contain 'news' because the

[3] Faust, J., Rogers, J.H. and Wright, J.H. (2005), *News and Noise in G-7 GDP Announcements*, Journal of Money, Credit and Banking, 37, 403-417discuss the regression technique for identifying 'news' versus 'noise' in revisions and provide citations of other work.

revision is not dependent on the later estimate. This suggests the revision is likely to depend on the earlier estimate (i.e. revisions are 'noise'), but this proposition must be tested as well and so an additional regression equation has to be estimated:

$$R_t = \gamma + \delta^* P_t + \xi_t \qquad (2)$$

and we would expect to reject the hypothesis that $\gamma = \delta = 0$ (ξ_t is the random error term) and confirm that the revision depends on the earlier estimate and are thus 'noise'. The analysis could have started with the second equation and then proceeded to the first; the salient point being that both equations must be estimated to be certain of the designation of 'news' or 'noise'. Economists tend to focus on equation (2) in their assessments, because if revisions are shown to be 'noise' then equation (2) can be used to predict the revision.

Importance of 'news' vs 'noise' analysis as a diagnostic tool for quality assurance

A national statistical institute (NSI) should strive in principle to ensure that revisions to their published data represent the addition of 'news' rather than the removal of 'noise'. NSIs continually receive new data that are often used to replace earlier judgments or assumptions about missing data. An evaluation of the production process would then focus on how well the agency is using the information available at the time an estimate is made. The 'news' versus 'noise' assessment is one tool that can be employed for such an evaluation; that is it can determine whether some revisions may not be the consequence of additional information and are potentially avoidable. If so, the statistical agency may conclude that the production process is not using all of the available information as well as it could and improvements must be sought. For example, the agency may want to reassess the methods used for imputing the value of missing data. If the revisions are found to consist of 'news' then the agency could conclude that it is indeed successfully using all available information.

Case study: do revisions to GDP for OECD countries represent 'news' or 'noise'?

In practice, the evaluation of whether revisions constitute 'news' or 'noise' may not be straightforward, as the results from statistical tests may not be clear cut in that there could be evidence of the existence of both 'news' and 'noise'. This also depends on the revision interval considered for analysis, e.g. one may be interested in evaluating revisions to subsequent releases of GDP, or alternatively some longer time interval, such as revisions between first published estimates of GDP and those released one year later.

Using data derived from the latest revisions analysis of seasonally adjusted quarterly GDP growth rates for eighteen OECD countries it is possible to perform an initial indicative study of whether revisions to first published GDP growth rates across these OECD countries constitute 'news' or 'noise'. Two revision intervals have been targeted[4] for this analysis:

[4] We chose these two revision intervals to restrict the size of the analysis, thus other revision intervals could have been considered with equal validity. For example, one may be interested in the revision between first published estimates and those once all normal benchmarking and balancing processes have been completed which in many countries occurs after about three years.

1. A 'first update' revision, which aims to capture the revision for each data point in the time series that occurs between its first published value and the value published with the first release of the next data point[5] in the time series.

2. The revision between first published data and that published one year later.

The reference periods over which this study has been performed relate to first published GDP seasonally adjusted quarterly growth rates from the fourth quarter of 1994 through to the third quarter of 2006 for most[6] of the eighteen OECD countries included in this study.

Data issues and limitations

Due to an occasional lack of consistency in the way data were published over the years in the OECD Main Economic Indicators (MEI) publication from which the revisions database is built, it has not been possible to precisely capture the 'first update' revision as described above for all periods due to the following reasons:

- A break in data transmission to the OECD, resulting in no data being published for several months followed by the subsequent publication of two or more new data points at the same time.
- A break in the consistency of publication of data points that leads to two new data points appearing within a five month period (which is the interval within which data have been captured for this project), such that the captured revision is in fact the 'second update' revision rather than the first.
- A zero revision caused by the fact that data have been updated manually at the end point of the time series only, which occurred from time to time prior to improved automated updating procedures implemented in 2003.

To overcome these problems, all affected reference points have been removed to create a more precise dataset of first update revisions. The penalty is a disjoint time series of revisions and a reduction in the sample size available to perform the analysis - implying a loss in precision for the test of statistical significance. The number of reference points from which the analysis of 'news' vs 'noise' for the first update revision is shown in the table below, and from this it can be deduced how many data points have been removed (the maximum number being 48, i.e. all data points from 94Q4 to 06Q3 inclusive).

The problems just described are of lesser importance when analysing revisions between first published data and those published one year later, and so no reference points have been removed for this analysis[7]. Whilst the above problems affect the preciseness of this revision interval (i.e. it may not be exactly one year between nationally published vintages), the purpose of this analysis

[5] In many countries this will also represent the first revision of the previously published reference point. However, some countries may release an updated GDP estimate for the latest published quarter without releasing a new data point. For example, this has been the case for many years in the United States and a similar approach has been adopted more recently in some European countries with the release of GDP 'flash' estimates.

[6] The main exceptions are for Korea, Portugal and Belgium where the references periods covered commence in 99Q4, 00Q3 and 96Q3, respectively.

in comparison to the 'first update' is to consider revisions over a longer interval so the same degree of precision was not applied. However, this could affect the consistency of the results of the two analyses.

The table below gives the results of the analyses for the two revision intervals considered using the more simplistic correlation tests as originally performed by Mankiw and Shapiro. The correlation of early estimates with the revision **Corr (P_t,R_t)** is used for the 'noise' test and the correlation of the later estimates with the revision **Corr (L_t,R_t)** is used for the 'news' test. We have used as an indicative test a t-test[7] based on the implicit assumption of Normal distributions of all the series (i.e. the revision, early estimates and later estimates)[8]. The lower the p-value the more confident one is that there is a correlation between the revision and the estimate, and the higher the p-value the more confident one is that there is no correlation between the revision and the estimate. One can then assess the existence of 'news' or 'noise' separately in the respective tests.

Given the interpretation of the concept, it is conceivable that a revision may contain elements of both 'news' and 'noise' and that different results may be obtained for different revision intervals. In the case of this analysis, it also needs to be taken into account that the data source is not perfect due to issues as outlined earlier in this article. In addition, the correlation method and p-values generated can only be expected to give preliminary results that should be the subject of more

[7] To perform a t-test of statistical significance one must first define a starting hypothesis, referred to as the 'null' hypothesis. For both the test of 'news' and 'noise' we begin with the 'null' hypothesis that the respective correlations defined above are not statistically significant from zero. That is, we start by assuming there is no 'news' or 'noise' and look for statistical evidence against this which will depend on the size of the observed correlation and the sample size. If there is no "news" or "noise", one expects that the observed correlations from a set of data would deviate from zero by chance according to a particular probability distribution, referred to as the t-distribution.

For a statistical test of this kind one generally needs to choose a level of significance 'α', where α is the probability, on the basis of the t-distribution, of making what is known as a type one error, i.e. to reject the null hypothesis when it is true. If the probability of the observed correlation happening by chance (called the p-value) is less than α then the null hypothesis is rejected. However, choosing a significance level α is rather arbitrary. A value of α = 0.05 is commonly used for statistical tests of this kind as a default setting. However, more stringent values of 0.02 or 0.01 or lenient values of 0.1 are not uncommon. An alternative is to simply report the p-value from the t-test. The lower the p-value the more confident one is that the null hypothesis is incorrect and there is a correlation between the revision and the estimate, and the higher the p-value the more confident one is that the null hypothesis is correct and there is no correlation between the revision and the estimate. One can then assess the existence of 'news' or 'noise' separately in the respective tests.

For example, in the case of New Zealand, we see that for the first update revision the correlation of the revision with the later estimate is 0.25 with a p-value of 0.13. This gives some weak evidence that the correlation is in fact different from zero which would imply that the revisions are 'news'. However, in the case of the revision between first published estimates and those published one year later (denoted as 'One yr later' in the table) the correlation between the revision and the later estimate is 0.35 with a p-value of 0.02, thus providing strong evidence that the correlation is different from zero and implying that the revisions are 'news'. Conversely, we see very low correlations between the revision and the early estimate for both the first update (-0.06) and one year later (-0.10) revision intervals and the high associated p-values of 0.73 and 0.49 confirm that there is no evidence that the revisions represent 'noise'.

[8] To "relax" this assumption it is possible to use a Spearman's rank correlation coefficient, although such a test has less statistical power.

formal analysis using the regression approach. Therefore, countries are encouraged to undertake this analysis with their own more precise data sources. In the case of the United States, where both first update and one year later revisions have a clear 'news' signal, these results are consistent with a recent study[9] performed by the Bureau of Economic Analysis that updated the original study of Mankiw and Shapiro performed in 1986.

'News' vs 'Noise' assessment of revision to first published GDP quarterly growth rates

Country	Revision interval	Sample (ref points)	Corr (P_t,R_t)	p-value	Corr (L_t,R_t)	p-value	Comment on news vs noise
Australia	First update	37	-0.16	0.34	0.25	0.14	Weak evidence of news
	One yr later	45	-0.25	0.10	0.50	0.00	Clear news signal
Belgium	First update	28	-0.21	0.28	0.28	0.15	Inconclusive
	One yr later	39	-0.44	0.01	0.44	0.01	Clear news and noise signals
Canada	First update	41	-0.08	0.63	0.15	0.36	Neither news nor noise
	One yr later	45	-0.21	0.18	0.25	0.10	Weak evidence of news
Denmark	First update	40	-0.15	0.34	0.37	0.02	Clear news signal
	One yr later	45	-0.29	0.05	0.44	0.00	Clear news signal but also some noise
Finland	First update	35	0.05	0.79	0.38	0.02	Clear news signal
	One yr later	45	-0.50	0.00	-0.06	0.68	Clear noise signal
France	First update	41	-0.13	0.43	0.25	0.12	Weak evidence of news
	One yr later	45	0.18	0.23	0.59	0.00	Clear news signal
Germany	First update	37	-0.01	0.97	0.31	0.07	Evidence of news
	One yr later	43	-0.29	0.06	0.18	0.26	Evidence of noise
Italy	First update	37	0.14	0.40	0.32	0.05	Evidence of news
	One yr later	45	-0.47	0.00	0.15	0.32	Clear noise signal
Japan	First update	41	-0.13	0.42	0.28	0.08	Evidence of news
	One yr later	45	-0.28	0.07	0.25	0.09	News and noise
Korea	First update	22	-0.44	0.04	-0.20	0.37	Clear noise signal
	One yr later	26	-0.27	0.18	0.22	0.29	Inconclusive
Netherlands	First update	43	-0.42	0.01	0.16	0.30	Clear noise signal
	One yr later	45	-0.65	0.00	0.12	0.42	Clear noise signal
New Zealand	First update	38	-0.06	0.73	0.25	0.13	Weak evidence of news
	One yr later	45	-0.10	0.49	0.35	0.02	Clear news signal
Norway	First update	33	-0.39	0.02	0.27	0.12	Clear noise signal
	One yr later	45	-0.16	0.31	0.58	0.00	Clear news signal
Portugal	First update	20	0.12	0.60	0.36	0.11	Weak evidence of news
	One yr later	22	-0.42	0.05	-0.02	0.92	Evidence of noise
Spain	First update	38	-0.34	0.04	0.32	0.05	News and noise
	One yr later	45	-0.28	0.06	0.43	0.00	Clear news signal (but also some noise)
Switzerland	First update	38	0.07	0.66	0.51	0.00	Clear news signal
	One yr later	45	0.18	0.24	0.77	0.00	Clear news signal
United Kingdom	First update	46	-0.28	0.06	0.29	0.05	News and noise
	One yr later	45	-0.44	0.00	0.33	0.02	Clear news and noise signals
United States	First update	42	0.05	0.77	0.40	0.01	Clear news signal

[9] See Fixler, D and Grimm, B., "Revisions, Rationality, and Turning Points in GDP" BEA Working Paper WP2003.1 which can be found at http://www.bea.gov/papers/pdf/RevRationality-abs.pdf

	One yr later	45	-0.05	0.75	0.43	0.00	Clear news signal

Notwithstanding the caveats of this analysis as listed above, from the results presented there is evidence for Australia, Denmark, France, New Zealand, Switzerland and the United States that revisions over both the first update and one year later intervals represent 'news' and not 'noise'. Therefore, this would support the assertion that the estimation process for GDP in these countries is efficient. In the case of Belgium, Japan, Spain and the United Kingdom we see evidence of 'news' for one or both revision intervals but 'noise' is also present, and in the case of the Netherlands and to a lesser extent Korea revisions appear to be solely 'noise'. For other countries we see mixed results, in that either the first update revision signals 'news' but the one year later revision 'noise' (Finland, Germany, Italy, Portugal) or vice versa (Norway). For Canada, the results are inconclusive.

An initial draft of this article, including the above results, was sent for comment to the institutes responsible for compiling the national accounts included in the analysis. A common issue raised in several countries was that the methodology used for compiling quarterly national accounts has changed[10] over the analysis period, and as such there is a break in the time series of revisions. This implies that it may not be appropriate to perform the tests of 'news' vs 'noise' over the entire interval considered in this analysis, and that it should be split to observe results before and after major changes in methodology. Other comments emphasised the importance of precisely defining the revision interval and the need to perform the more detailed regression analysis in order to make conclusions. Specific comments from countries who performed some initial analyses of their own were:

- In the Netherlands there were a number of significant changes in data sources and to the balancing process introduced over the course of the analysis period that influence the 'noise' correlations. The latter have progressively declined in recent years as a result of these improvements;
- Italy introduced 'flash' estimates for the first published growth rate of quarterly GDP in the third quarter of 2003. Analysing from this point onwards shows no evidence of 'noise' for the revision between first published and one year later estimates;
- Japan introduced major methodological changes in the third quarter of 2002 and also brought forward their first release estimates to 45 days after the reference period. Analysing revisions after this point shows stronger evidence of 'news' and no evidence of 'noise'.

Concluding remarks

Revisions are an essential part of the process of providing decision-makers with the most reliable picture of economic activity available at a particular point in time. It is therefore extremely important for statistical agencies to convey to these users that the revisions are adding information rather than confusing the picture by providing random changes in official estimates. If the latter case is true, then users have a right to question the efficiency of the compilation practices used by the statistical institution responsible. Determining whether revisions between earlier and later published estimates represent 'news' or 'noise' by the process described in this article can be an effective diagnostic tool in this assessment, and is therefore of interest to both producers and users

[10] In particular, several countries described their implementation of more timely 'flash' estimates for the first publication of GDP growth.

of official statistics. Of course the possibility of diagnostic revision tools giving clear 'news' and 'noise' signals for the same economic variable is possible and indicates that a deeper analysis is required.

The results from the 'news' vs 'noise' assessment shown in this paper should only be considered indicative for the countries presented due to the limitations in the data source used, thus strong conclusions cannot be drawn as to the efficiency of the GDP estimation process in the countries concerned. It should also be noted that the table presented in this article uses the more simplistic correlation test as an indicative pointer, presenting the full results to a satisfactory level from the more powerful regression models is not possible in a brief article like this. Nonetheless, the key purpose of this study is to encourage national statistical offices to undertake more comprehensive studies of their own and report the results, and more importantly seek to make improvements to their compilation processes should the existence of 'noise' in revisions be found.

Further Reading

Faust, J., Rogers, J.H. and Wright, J.H. (2005), *News and Noise in G-7 GDP Announcements*, Journal of Money, Credit and Banking, 37, 403-419.

Fixler, D. and Grimm, B. (2003), *Revisions, Rationality, and Turning Points in GDP*, BEA Working Paper WP2003-01.

Mankiw, N.G., Runkle, D.E. and Shapiro, M.D. (1984). *Are preliminary announcements of money stock rational forecasts?*, Journal of Monetary Economics 14, 15-27.

Mankiw, N.G. and Shapiro, M.D. (1986), *News or Noise: an Analysis of GNP Revisions*, Survey of Current Business 66, 20-25.

McKenzie, R (2006). *Performing Revisions and Real-Time Data Analysis.* OECD Statistics Brief, Issue 12, November 2006, Paris.

McKenzie, R & Adam, Z (2007). *Revisions in Quarterly GDP of OECD Countries: An Update*. Conference paper: OECD Working Party of National Accounts. Paris, October 2007.