Statistics Commission

Revisions to Economic Statistics

Review by National Institute of Economic and Social Research

Statistics Commission Report No 17, Volume 2
April 2004

Statistics Commission

Report No. 17, Volume 2

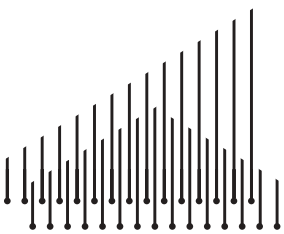
Review of Revisions to Economic Statistics

A report to the Statistics Commission

by Dr James Mitchell

National Institute of Economic and Social Research

Statistics Commission
10 Great George Street
London
SW1P 3AE
020 7273 8008
www.statscom.org.uk



National Institute of Economic and Social Research

2 Dean Trench Street
Smith Square
London SW1P 3HE
Tel +44(0) 207 222 7665
Fax +44(0) 207 654 1900
http://niesr.ac.uk

Richard Alldritt,
The Chief Executive,
Statistics Commission,
10, Great George Street,
London SW1

Dear Richard,

I enclose the National Institute's report to the Commission on recent revisions to economic data. It has been prepared by James Mitchell with my assistance. I should particularly like to draw to your attention the very high level of co-operation from the Office for National Statistics. The detailed work they have done for the review is included in Annex 3* of the report.

Yours sincerely,

Martin Weale

The Director
National Institute of Economic and Social Research

27 February 2004

*See Statistics Commission Report No. 17, Volume 3

Acknowledgements

This Review has benefited from multiple information sources. We should like to thank the following:

The Statistics Commission and the Statistics Commission Project Board: Richard Alldritt, Allen Ritchie, Janet Trewsdale, Derek Wanless (chairman of the Project Board), Martin Weale and Lovedeep Vaid. Colin Mowl attended Project Board meetings as an observer on behalf of the ONS.

ONS: David Caplan, James Ebdon, Robin Lynch, Rob Pike and especially Colin Mowl and Craig Richardson.

DTI: Glenn Everett, Keith Folwell, Frances Pottier and Geoff Reed.

Bank of England: Charles Bean, Ian Bond, Spencer Dale, Melissa Davey, Simon Whitaker and Chris Young.

HMT: Simon Brooks and Toby Fenwick.

City economists and journalists: Simon Briscoe, Financial Times; Ed Crooks, Financial Times; Kevin Daly and David Walton, Goldman Sachs; Geoffrey Dicks, The Royal Bank of Scotland; Gary Duncan, The Times; Faisal Islam, The Observer; Adam Law, Barclays Capital; Michael Saunders, Citigroup; David Smith, Williams de Broë; Philip Thornton, The Independent.

CBI: Ian McCafferty.

Academic users: Tony Garratt, University of Leicester and Shaun Vahey, University of British Columbia.

Review of Revisions to Economic Statistics

Detailed Index

Detailed Index

1. Preface

1.1 During 2003 the ONS and other Government Departments made a number of important revisions to major economic statistics which attracted a great deal of attention. In response, on 9th October 2003 the Statistics Commission announced that it would undertake a review of these recent revisions and what lay behind them.

1.2 The National Institute of Economic and Social Research was awarded the contract for the review on 4 December 2003. The National Institute is carrying out the review with the support of the Statistics Commission. The following sections constitute the National Institute's submission to the Statistics Commission.

1.3 The ONS, and other Government Departments, have co-operated fully with the Review and their input is gratefully acknowledged. This has involved face-to-face discussions plus the provision of written material. The ONS's written material is brought together in Volume 2 (Annex III). This includes written answers to specific questions put to the ONS by the National Institute.

1.1 Terms of Reference

1.4 To identify the key revisions to official statistics in the UK since June 2000 and agree with the Commission those that should be examined further.

1.5 To consider with key users of these statistics the impact of the identified revisions on decision making and on confidence in the evidence base.

1.6 To categorise the root causes of each of the revisions identified into the three categories listed below and address the questions posed against each of the three categories:

(i) Consequences of pre-announced changes in methodology. Prime question to be addressed: were the changes adequately flagged in advance?

(ii) Consequences of unforeseeable additional information. Prime question to be addressed: why could the additional information not have been predicted?

(iii) Avoidable circumstances. Prime question to be addressed: what happened and what could have been done to avoid them – and what lessons for the future?

1.7 To research more deeply any revisions that seem to fall into the 'avoidable circumstances' category, exploring the nature of any systematic or individual failing that may have caused each revision and the steps that might be required to prevent an equivalent failing in the future.

1 Preface

1.8 To assess whether there is any evidence of systematic bias in sets of revisions.

1.9 To comment on the reactions of commentators and the media to all the revisions identified – were those reactions reasonable and balanced; and in any case where they were not, to consider what prompted the reactions.

1.10 To check whether the revisions identified were managed in accordance with the National Statistics Code of Practice.

1.11 To identify any lessons for the future, including any steps that would lead to a systematic reduction in the size and frequency of controversial revisions.

2. Summary including principal conclusions and recommendations

2.1 The ONS performs well relative to statistical offices in other countries, in terms of the speed of delivery of its estimates of key macroeconomic variables such as GDP.

2.2 Its revision pattern, at worst, is similar to that of other offices and over the 1990s has been better.

2.3 But recent revisions have damaged the reputation of the ONS, in the sense that press comment in early October 2003 became extremely critical. A greater public understanding of revisions, and their causes, is required. We welcome the recent article by the National Statistician [13/1/04].[1] The ONS needs to build on this and consider routine means of "education" that help to build trust in national statistics. For example, it is very important that the ONS educates its users about the methodological improvements associated with the forthcoming re-engineering of the national accounts in the UK; see Tuke and Aldin (2004).

2.4 Discussion with the ONS during the course of this Review has led to an improvement in the amount of material in the public domain. This is to be welcomed, and we hope that they will build on this enhanced transparency.

2.1 Summary of Conclusions

C.1 There is an understandable demand for prompt economic data and this in turn means that data are bound to be revised.

C.2 Large revisions are often taken to be indicators of poor quality of initial data; they may of course instead indicate final data of particularly high quality. Series may not be revised because nothing new is learned rather than because they are initially accurate.

C.3 Revisions are caused by (i) pre-announced changes in methodology; (ii) the availability of unforeseen additional information and (iii) avoidable circumstances.

[1] This article is available at the ONS web-page: http://www.statistics.gov.uk/about_ns/downloads/economic_revisions_article_len_cook.pdf. More recently, it was published in the February 2004 edition of *Economic Trends*; see Cook (2004).

2.1.1 Interviews with users

C.4 From the many of the thousands of data that are revised from one period to the next we identify those revisions in the last few years that have attracted the most attention by interviewing a range of different types of user of economic statistics.

C.5 The following three revisions were identified consistently by the twelve key users of economic statistics; they represent the most important revisions from their perspective:

a. The upward revisions to GDP growth in 2003 Q1 and Q2 between the m1 (month one or 'preliminary') and m3 (month three) estimates. It should be noted that revisions continue for long after the m3 estimates.

b. The revisions to estimates of activity and the main macroeconomic aggregates over the 1998-2002 period that occurred at the Blue Book 2003.

c. Revisions to import and trade estimates due to the effect of VAT missing trader intra-community fraud.

C.6 Other revisions mentioned, but not always with specific examples in mind, were those to retail sales, the Index of Production, the average earnings index, population estimates related to the 2001 Census, employment, business investment, the capital stock, pensions and regional gross valued added (GVA). The revision to the average earnings index was discussed by the Turnbull-King report in 1999. Aspects of the revisions to the population estimates before and after the 2001 Census are under investigation by the Statistics Commission separately of this review. Neither of these is considered substantially.

C.7 A range of questions were asked of the key users designed to determine the effect of the revisions on the users, and decisions they may have taken.

a. The interviews revealed that revisions are expected. Users know that they will happen and moreover have an understanding of why they happen. Revisions are seen to be, in general, the result of methodological change and/or the arrival of new, more complete information understandably not available at the time the first estimate is produced given how quickly the first estimates, for example of GDP, are produced by the ONS.

b. Nevertheless, some users were surprised by the scale of the revision to quarterly GDP growth between m1 and m3 for 2003q2. Others were less surprised. Contributing to some users' lack of surprise was the observation that the construction output growth estimate used by the ONS in the m1 and m2 GDP growth estimates was "too low". They felt that published qualitative information about construction (at m1) and quantitative information (at m2) conflicted with other signals and they also noted the imbalances between supply and demand in the economy. Either way, the q2 revision clearly had an impact on the financial markets.

c. Although the majority of revisions were not seen by users to be due to 'avoidable circumstances', revisions to pensions, regional GVA, population estimates (related to the Census), the average earnings index and the capital stock were viewed by many as avoidable.

d. Users expect the preliminary estimate of quarterly GDP growth to be revised upwards with the passage of time. Such a bias, whether statistically significant or not, is seen as a cause for concern.

e. Substantial revisions to time-series, like those associated with methodological changes such as re-basing, are seen as inevitable, indeed desirable, but they do impose a burden on users. The implied re-writing of economic history necessitates extra work, such as re-estimation of economic models with the revised data.

f. The ONS performs well relative to statistical offices in other countries. The ONS is commended for the speed of delivery of its GDP data. Users, in general, are happy with the current trade-off between the speed and accuracy of estimates. In particular users feel that the ONS did a good job at warning them in advance about the introduction of annual chain linking.

2.1.2 GDP growth estimates

C.8 The revision of 0.3% points to the 2003 Q2 estimate of GDP growth is large compared with previous revisions between the m1 and m3 GDP growth estimates. Over the previous 20 quarters there has only once before been a revision as large as 0.3%.

C.9 Although revisions have always occurred at re-basing, the re-basing in September 2003 (also coinciding with a move to annual chain-linking) is slightly larger than the revisions associated with previous re-basings. It is, however, always difficult to isolate revisions due to re-basing alone as other changes occur at re-basing.

C.10 Studies have repeatedly found that the preliminary estimates of quarterly GDP growth in the UK are biased downwards. However, there is some evidence that the size of the bias varies over the economic cycle; e.g. see Richardson (2002).

C.11 The overall revision to GDP growth, as indicated by the root mean squared error (RMSE), has got markedly smaller since 1989.[2] This holds for both quarterly and annual growth. It is consistent with the view that both methodological improvement and improved data sources (both new surveys and larger samples for existing surveys) introduced by the ONS in the late 1980s, in the aftermath of the Pickford Review, have delivered more accurate estimates.

[2] The RMSE is a measure of accuracy that considers both the bias and dispersion of the estimate.

C.12 A comparison of revisions to quarterly GDP in the UK with those in the US, the only other country that we are aware of that also produces an early estimate of GDP growth, reveals that the overall revision over the period 1993q1-1999q4, as indicated by the RMSE, is lower in the UK than in the US.

C.13 Revisions to annual GDP growth estimates in the UK have a similar RMSE to annual growth estimates in Canada, Germany, Sweden and the US using data from 1980-1998 and a lower RMSE if data for the 1980s are excluded.

2.1.3 Causes of the revisions: quarterly GDP growth revisions

C.14 The causes of the revisions to these key statistics were investigated with the help of the statisticians concerned. It is concluded that:

a. Revisions to quarterly GDP growth typically are explained by pre-announced changes in methodology and the arrival of additional information. Given the speed at which the preliminary estimates are produced revisions are to be expected as more detailed information becomes available over time. The ONS does indicate to users that data are liable to revision.

b. The revision of -0.3% in 2002 Q2 has a specific 'special' explanation, namely the Jubilee holiday.

c. The upward revision to GDP growth for 2003 Q2 was caused by the large upward revision to construction output between the first and third months (m1 and m3, respectively). When the DTI provide their m1 estimate of construction output growth to the ONS they typically have no more than 5% of survey returns and so forecasting models are used to meet the ONS's deadline. When survey information eventually become available in m3 revisions should be expected.

d. Although revisions to construction estimates between m1 and m3 should be expected, it is important to test whether the revision could have been anticipated *ex ante*. If it could then the revision was at least partly due to 'avoidable circumstances'. As outlined below in paragraph 3.7 we make the distinction between four types of 'avoidable circumstance'. It is considered whether avoidable circumstances are explained by: (i) substantial mistakes in early processing; (ii) the models used to compute early estimates are not 'best practice'; (iii) processing timetables could be more rapid than they actually are or (iv) when the method used to obtain the estimates is "best practice" and theoretically appealing, but the method is implemented without sufficient resources to ensure adequate assessment of its practical performance so that unforeseen problems occur. Although users cannot expect the method used to obtain the m1 (and m2) construction estimates to be perfect, they should expect the method used to be the best in the circumstances.

2.1.4 Construction growth estimates

C.15 We examined the forecasting models used by the DTI for construction at m1 in some detail, albeit with modest resources, not just because construction is of interest *per se* (construction is responsible for just over 5% of GDP – however, it is a volatile series and its contribution to the variation of GDP growth is much larger than 5% suggests, indeed perhaps around three times) but also because it illustrates how we think forecasting models in general should be used and also assessed.[3]

C.16 We note that it may be difficult for the ONS to exercise adequate quality control when information is provided by third parties, like the DTI, however good the links with the third party. Moreover, the lack of information in the public domain on the construction model means the model has not been subject to wider peer review.

C.17 We conclude that if the appropriate forecasting approach had been followed, it would have helped identify shortcomings in the primary model used by the DTI to forecast construction output growth at m1. While the DTI and ONS also appear to have been aware of these shortcomings, reflected in the development of an alternative forecasting model in 2002, they have nevertheless continued to base their forecasts at m1 largely on the inadequate model rather than the alternative better one.

C.18 The history of the construction forecast used in 2003q2 is confused. The DTI model produced an estimate of -0.4% and the DTI passed this to the ONS with the comment that construction output was broadly "unchanged" in the quarter. ONS then used a figure of 0.3% in the production of the m1 GDP estimate. This figure was a combination of the DTI information and ONS's own "judgement" based on qualitative information; see Annex 3: paragraphs 86-88 for ONS's own description of this process.

C.19 We find if appropriate methods had been in place to forecast construction output at m1, construction output growth would have been forecast at 0.9% instead of at -0.4%, which was the DTI's forecast. It is impossible to say with certainty, given the use of "judgement" by both DTI and ONS, what effect this would have had on the construction estimates used by the ONS at m1. In turn it is therefore impossible to say what effect this may have had on the GDP estimate at m1. However, if the m3 construction figure of 5.3% had replaced 0.9%, instead of the 0.3% figure used by the ONS at m1, then the revision to GDP growth would have been about 0.04% smaller. Subject to rounding this may have led to the publication of an m1 estimate for GDP growth in 2003q2 of 0.4% rather than 0.3%. Therefore, the revision to growth for 2003q2 between m1 and m3, at least in part, was due to 'avoidable circumstances'.

[3] A previous research project by the Statistics Commission [see Statistics Commission (2003a)] undertook a scoping study into forecasting in the national accounts. This project considered forecasts made by the ONS only, and therefore did not consider construction. It is not clear whether the contractors were aware that forecasting work for the national accounts was carried out outside the ONS.

C.20 We are concerned about the routine use of "judgement" by the ONS when making forecasts without an attempt to assess whether this judgement, in general, improves forecasts. While we know that it did deliver improvements for the specific case of construction growth in 2003q2 the necessary data are not available to assess the ONS's use of judgement more generally. Judgement is used widely in making economic forecasts since the models by construction are designed to meet theoretical criteria at the expense of statistical performance. Statistical forecasts, by contrast, should be carried out using all the information, combining forecasts from a number of different models if necessary. Therefore judgement should normally have nothing further to add. The only exception to this is in the aftermath of rare events, such as the foot and mouth outbreak, when judgemental revision is required since such events are not accommodated by statistical models.

2.1.5 Causes of the other revisions

C.21 Revisions to economic time-series when detailed annual national accounts are published in the "Blue Book" each year (e.g. on 24th October in 2003) are caused by pre-announced changes in methodology and the arrival of additional information. More specifically, the changes to real growth that occurred at the 2003 Blue Book are explained by: (i) annual chain linking; (ii) rebased deflators; (iii) rebased lower level volume series and other methodological improvements such as to the deflation of imported capital goods.

C.22 The ONS should be congratulated for the job it did at warning users in advance of the likely effects of the introduction of annual chain-linking.

C.23 The arrival of additional information, namely awareness of criminal activity, explains revisions to import estimates.

C.24 The revisions to the regional GVA estimates and pensions (specifically those to pension funds assets not pension contributions; see paragraph 6.10g) were due to "avoidable circumstances". More specifically, GVA and pension fund assets revisions were due to processing errors.

C.25 The majority of revisions, however, arise due to pre-announced changes in methodology and the arrival of additional information.

2.2 Summary of Recommendations

R.1 **The ONS and the Statistics Commission should think about how best to keep users informed about revisions to economic statistics and their causes.** Although serious users are aware that economic data are revised, ONS communication about the likelihood of revisions should be improved. While we welcome the National Statistician's article of 13th January 2004 perhaps a similar exercise should be undertaken on a routine basis. The ONS should explore means of changing the climate in which statistics are used, so that the public comes to understand that revisions to economic statistics are the norm, and should not be

seen as corrections of previous mistakes. Revisions might then generate less controversy than has sometimes arisen in the past.

R.2 The ONS should provide more information about past revisions in its first releases. The Statistics Commission has recommended this in the past. The ONS have announced plans to comply; see Annex 3: paragraph 332.

R.3 Following the introduction of information about past revisions into the first releases of time series the ONS should devote resources to assessing user comprehension and reaction.

R.4 The ONS could go further and should consider grading data again according to their perceived reliability. To provide a simple, easy-to-read, indicator of the reliability of the data the ONS could consider grading data, say *A*-*C*, according to their reliability along the lines of the system of reliability gradings introduced in 1956.[4] An *A* grading indicates (with 90% confidence) that the reported figures are correct subject to a margin of error less than, say, 3% in either direction; *B* indicates an error of 3%-10% in either direction and *C* indicates that the error is greater than 10%. It is, however, essential that when this is done distinction is made between reliability of level estimates and reliability of growth estimates.

R.5 The Statistics Commission should consider whether when the ONS believe data are liable to substantial revision they should make public their concerns immediately or they should wait until the revisions can be quantified. The ONS were aware of 'errors' in the import data, due to the VAT fraud, well in advance of informing the public. We feel that the ONS did have good reasons for keeping this information private, and for sharing it with selected users, such as the Bank/HMT, only – the ONS action can be seen consistent with their obligations as stated in the Code of Practice and its relevant protocols. There is nevertheless an issue of whether the ONS might be expected to publish on its website a "running commentary" on those points that are of sufficient importance to justify letters to the Bank/HMT or whether it is better for the ONS to wait until the evidence has accumulated and then publish both the revisions and the reasons for them.

R.6 Routine revisions analysis should be undertaken not just of aggregate GDP but also of its components on both the output and expenditure sides. Without such analysis it is impossible to know whether apparent biases in GDP always originate in the same components of expenditure and output. It is thus impossible to know how to address the issue.

R.7 The ONS should keep copies of real-time data sets (i.e. of revision triangles) and make these publicly available via a dedicated revisions web-site.

[4] For further discussion of the British system, references and an excellent methodological discussion see Morgenstern (1963), p. 273.

2 Summary including principal conclusions and recommendations

R.8 Forecasting models should be assessed systematically along the lines we suggest in Annex I, notwithstanding the fact that in the specific case we examined one of the models used by the ONS/DTI, although not the primary model, performs well. Had this been done as a matter of course the DTI/ONS would have discovered problems with its principal forecasting model earlier; see C17-C19.

The remaining recommendations listed below relate to construction but many are more general. If recommendations R9-R11 had been followed this would have helped identify shortcomings in the model used to forecast construction output growth.

R.9 It should be ensured that appropriate resources are directed to the methodological development of techniques used to produce data, such as construction output, that are an input into the national accounts. The resources needed are small relative to the importance of the data.

R.10 For transparency and to promote use of best-practice tools, the ONS/other data producers should make details of the models they use publicly available.

R.11 Additionally, it would be sensible to invite peer review by the ONS Methodological Division or experts from outside the GSS. This should promote use of “best-practice” techniques.

R.12 Production of construction estimates at month one and month two should be handled by the ONS rather than the DTI. This would facilitate quality control similar to that carried out for other national accounts data. We also draw the Statistics Commission’s attention to the fact that some other components of the national accounts, including the month three construction estimates which unlike the month one and two estimates are not model but survey based, are also compiled outside the ONS. Analysis of the desirability of this is outside the scope of this review.

R.13 In periods of concern about data the ONS/other data producers should ensure they indicate their lack of confidence in their estimates to the public. The ONS needs to decide how best to do this.

R.14 Various means of using information from qualitative surveys should be explored. The DTI has produced a model which uses qualitative information. This gives much better results than the model traditionally employed. But the study did not conduct a review of different ways of exploiting the available data.

R.15 **The ONS should consider whether the quality of early GDP estimates would be helpfully improved if the UK followed Germany in producing a monthly index of construction output.** This would put estimates of construction output on a basis similar to industrial production and would be likely to deliver much better m1 estimates of quarterly GDP. The DTI/ONS would no longer have to rely on model-based estimates for construction at m1. Given the volatility of construction this would be likely to deliver better estimates than those obtained at m1 not just from the DTI's primary model but even the 'best' forecasting model. For example, even if appropriate forecasting methods had been in place at m1 to forecast construction output for 2003q2 there would have been a revision against the outturn (at m3) of 4.4% [see paragraph C.19]. Against the production of a monthly index must be offset the cost to the Government and the extra burden imposed on business. We welcome the fact that, as we understand it, the second stage of the DTI/ONS joint review into the early estimates of construction output for GDP currently under way [see paragraph 9.9], amongst other options, is considering the production of a monthly index of construction output. Indeed, this was one of the recommendations of the first stage of the joint review that is published in Annex 3: Section B3.3.[5]

[5] See also: http://www.statistics.gov.uk/about/Methodology_by_theme/constructionstats/default.asp.

3. The Problem of Revisions

3.1 There is an understandable demand for prompt economic data and this in turn means that data are bound to be revised. Data may be required before processing can be completed, late respondents tracked down and queries investigated. Initial sources of information used to provide quarterly data may be less satisfactory than final sources based on annual surveys and in some areas, such as the five-yearly population study and the ten-yearly Census, information may be based on benchmarks which are established less frequently than once a year. Large revisions are often taken to be indicators of poor quality of initial data; they may of course instead indicate final data of particularly high quality. Series may not be revised because nothing new is learned rather than because they are initially accurate.

3.2 Nevertheless, recent revisions have attracted lots of attention. Press comment has perhaps become increasingly critical; the press certainly have made numerous attacks on the ONS in the last couple of years; see Section 7 below. This Review was announced by the Statistics Commission in response to concerns about the effect of revisions on key users of the statistics.

3.3 The Review hopes to raise awareness about revisions and increase our understanding of them.

3.1 The scope and aims of this Review

The scope of this Review into recent revisions to economic statistics is as follows:

3.4 The Review identifies the key revisions to official statistics in the UK in the last few years. There have, of course, even over just the last few years, been a large number of revisions. Many of the thousands of data in the Blue Book, for example, are revised from one year to the next and no useful purpose would be achieved in investigating all of these revisions. Therefore, we identify those revisions in the last few years that have attracted the most attention by interviewing a range of different types of user of economic statistics. We acknowledge that, of course to some extent, the most recent revisions are always the most important and memorable. We focussed the attention of users on national rather than regional data; some regional issues were nevertheless raised.

3.5 The Review considers the impact of these revisions on key users. By consulting key users we discover the effect of these revisions on decision making and their confidence in the statistical system.

3.6 The Review explains the causes of these revisions. A decision is made about whether revisions are caused by (i) pre-announced changes in methodology; (ii) the availability of unforeseen additional information or (iii) avoidable circumstances. However, this distinction may not always be possible because on occasion additional information may have been incorporated at the same time as methodological changes were made, making it difficult to unravel such changes without substantial input from the ONS.

3.7 The Review investigates further any revisions that are seen to be due to avoidable circumstances. It is considered whether the avoidable circumstances occurred because: (i) there are substantial mistakes in early processing; (ii) the models used to compute early estimates are not 'best practice'; (iii) processing timetables could be more rapid than they actually are or (iv) when the method used to obtain the estimates is "best practice" and theoretically appealing, but the method is implemented without sufficient resources to ensure adequate assessment of its practical performance so that unforeseen problems occur.

3.8 We see no need to replicate the regular analysis by the ONS itself looking for bias in key series, but we comment on the evidence for bias taking account of revisions to data more recent than those they consider; this can of course be done only for the part of the revision process which has already taken place. The magnitudes of recent revisions to key variables are compared with the history of revisions in the past. We also explore the issue of bias in revisions to the index of industrial production and construction output. Information is also presented on the revisions process in other countries so as to establish whether there is anything unusual about what happens in Britain.

3.9 Newspaper comment following on from key revisions is explored. The comment will be assessed in the light of the magnitudes of the revisions which took place.

3.10 The Statistics Code of Practice Protocol on Revisions was in force only at the end of the period of interest; in fact, it was only a draft that had been published for consultation. It is nevertheless established whether the key provisions of the protocol were followed ahead of its implementation.

3.11 Conclusions are drawn about ways of reducing the magnitude of revisions and means of ensuring that, when they do take place, they generate less controversy than has sometimes arisen in the past.

4. The "List of Key Revisions"

4.1 How the list was constructed

4.1 In consultation with the Statistics Commission Project Board a list of "key" users of economic statistics was drawn up. The users fell into three camps: (i) public sector, such as the Bank of England and HMT; (ii) private sector, including leading City economists and (iii) economic and financial journalists.

4.2 The following key users were interviewed:

a. Bank of England[6]

b. HMT[7]

c. Simon Briscoe; The Financial Times

d. Ed Crooks; The Financial Times

e. Kevin Daly and David Walton; Goldman Sachs

f. Geoffrey Dicks; The Royal Bank of Scotland

g. Gary Duncan; The Times

h. Feisal Islam; The Observer

i. Adam Law; Barclays Capital

j. Michael Saunders; Citigroup

k. David Smith; Williams de Broë

l. Philip Thornton; The Independent

4.3 The key users were first consulted in order to build up a list of "major revisions". When this list was completed interviews were carried out to establish the impact of these revisions and their effect on confidence in the statistical system. It was established what sort of expectation the users have of revisions and whether they were surprised by the magnitude and direction of the key revisions. It was also determined how far they are aware of ONS plans and timetables for methodological changes. An attempt was made to establish the nature of any reliability/timeliness tradeoff from the perspective of key users and whether this might be affected by the provision of better information about past revision patterns.

[6] The representatives from the Bank of England were: Charles Bean (Chief Economist & Executive Director, Monetary Analysis & Statistics); Ian Bond (Head, Structural Economic Analysis Division); Simon Whitaker (Manager, Structural Economic Analysis Division); Spencer Dale (Head, Conjunctural Analysis & Projections Division); Melissa Davey (Economist, Conjunctural Analysis & Projections Division); Chris Young (PS to Charlie Bean)

[7] The representatives from HMT were Simon Brooks (Head of Macroeconomics) and Toby Fenwick (PS to Simon Brooks).

4.2 Identifying the list

4.4. Table 4.1 summarises the list of key revisions identified by each of the key users. The following three revisions were identified consistently by the users; they represent the most important revisions from their perspective.

a. The upward revisions to GDP growth in 2003 Q1 and Q2 between the m1 (month one or 'preliminary') and m3 (month three) estimates, which have been revised up from 0.1% and 0.3% to 0.2% and 0.6% respectively, and the general tendency for the preliminary estimate of quarterly GDP growth to be revised upwards.

b. The revisions to estimates of activity and the main macroeconomic aggregates over the 1998-2002 period that occurred at the Blue Book 2003. Real GDP growth, for example, was revised up at Blue Book 2003 from its previous estimates by 0.4% and 0.7% in 1999 and 2000, respectively.

c. Import and trade estimates. On 9th July 2003 the ONS published revised estimates for imports of goods to the UK from the rest of the EU that accounted for the effect of VAT missing trader intra-community fraud. As a result the UK's trade deficit was much larger than previously thought. Specifically, there were upward revisions to imports of £1.7 billion in 1999, £2.8 billion in 2000, £7.1 billion in 2001 and £11.1 billion in 2002.

4.5 HMT chose not to provide their own list of specific revisions but shared other users' view that the three identified in 4.4 were important. HMT noted that revisions due to 'avoidable circumstances', such as mistakes, could cause difficulties for users. But HMT added that revisions of this type were thankfully rare; indeed they should become rarer with the forthcoming re-engineering of the national accounts. HMT also stressed that bias (ie. predictable revision) in a statistic's revision history can cause problems for users.

4.6 Other revisions mentioned, but not always with specific examples in mind, were those to:

a. Retail Sales.

b. Index of Production.

c. The suspension of the average earnings index; subject to its own Review in 1999. This is not considered further here except when reporting user views.

d. Census 2001 related. Revised mid-2001 population estimates were published on 26th September 2003. About 190,000 people were added to the mid-2001 Census based estimate of the population of England and Wales. Most of the extra people are young adult men aged 25-34.

4 The “List of Key Revisions”

Further revisions are possible when the ONS completes studies for Manchester and Westminster that match addresses used in the Census enumeration with those from local administrative sources. In fact, on 4 November 2003 ONS published a new provisional population estimate for Manchester, based on the findings to that date from the Census matching Project for Manchester. As a result the population estimates for Greater Manchester, the North West Region, England, England and Wales and the UK were amended. All these estimates are provisional in that they could be revised when the next population estimates are published in August 2004. The Statistics Commission is carrying out a separate investigation into the Westminster revision. This too is not considered further here, except when reporting user views.

e. Employment. This refers to revisions to employment measured on both the demand side (the number of jobs) and supply side (the number of people in employment) using the Annual Business Inquiry or Labour Force Survey (LFS), respectively. The re-benching of workforce jobs to the 2001 Annual Business Inquiry, published in December 2002, led to upward revisions to the number of jobs. Additionally, the LFS estimates were revised in the light of new population estimates from the 2001 Census. More generally, users spoke of the impact of these revisions on the reconciliation of the two measures and the delays in grossing up the LFS data using the new population data.

f. Business investment. Users spoke of repeated upwards revisions to business investment in 2002 and 2003. This derives from the observation that over the period 1998q1-2003q3 there has been a positive mean revision to quarterly business investment growth (in percentage points) between the provisional and revised estimates of 0.74%.

4.7 A number of other revisions were mentioned, all earlier than 2003, such as those to the capital stock (resulting in its suspension), pensions (in January 2002 processing errors led to a revision of over £100 billion to the estimate of pension funds assets; users also spoke of “double-counting” of pension contributions) and regional gross valued added [GVA].

Table 4.1: Summary Table of Key Revisions Identified by Users[8]

	GDP Q1 Q2	Blue Book	Trade VAT fraud	RS	IoP	AEI	Census	Emp	Bus Inv.	Other
Bank of England	x	x	x			x	x	x	x	Capital stock
Briscoe	x	x				x	x	x		
Crooks	x	x	x							Pensions; Regional GVA; Mix of growth
Dicks	x	x								Public spending
Duncan	x		x		x					Deflator Gov. C
Islam	x	x	x							Prod.
Law	x		x	x	x					
Saunders	x	x				x				
Smith		x								ESA 95 Definitional
Thornton	x	x	x				x			Pensions; Network Rail
Walton/ Daly	x			x	x			x	x	
Total	10	8	6	2	3	3	3	3		2

[8] GDP Q1 Q2 refers to the revisions to quarterly GDP growth in 2003 Q1 and Q2; Blue Book refers to revisions to the annual national accounts usually the annual GDP estimates; RS refer to retail sales; IoP refers to the Index of Production; AEI refers to the Average Earnings Index; Emp. refers to employment measured on either the demand side (the number of jobs) or supply side (employment) using the Annual Business Inquiry or Labour Force Survey, respectively; Bus Inv refers to business investment and Network Rail refers to the Network Rail statistical classification. It should be added that HMT chose not to submit formally to this Review its own list of specific revisions; however, discussions we did have with HMT do feed into Section 5. See also paragraph 4.5.

5. The impact of revisions: a summary of the interviews with the users

5.1 A range of questions were asked designed to determine the effect of the revisions on the users, and decisions they may have taken. These questions included: what impact have these revisions had and have they affected policy/analysis; what has been their effect on your confidence in the statistical system; what sort of expectations do you have of revisions; were you surprised by the magnitude and direction of the key revisions; how aware are you of ONS plans and timetables for methodological changes; what needs do you have of economic statistics – reliability versus timeliness – and might this be affected by the provision of 'better' information about past revision patterns or other information on reliability?

5.2 Subject to the permission of the user transcripts of the interview were prepared. These transcripts are included in Annex 1. Importantly, they summarise, deliberately in note form, the interview in the words of the interviewer not the interviewee. The users were given the opportunity to correct and/or add to the transcript prior to publication in Annex 1.

The results of the interviews can be summarised as follows:

5.1 Summary of Users' Conclusions

5.3 Revisions are expected. Users know that they will happen and moreover have an understanding of why they happen.[9] Revisions are seen to be, in general, the result of methodological change and/or the arrival of new, more complete information understandably not available at the time the first estimate is produced given how quickly the first estimates, for example of GDP, are produced by the ONS.

5.4 Revisions to pensions, regional GVA, population estimates (associated with the Census), the average earnings index and the capital stock were viewed to be due to 'avoidable circumstances', such as mistakes, in some sense and to some extent; see section 6 for further discussion.

[9] For example, City economists regularly write reports that anticipate revisions to national accounts data due to methodological changes and examine their implications; e.g. see Dicks and Walker (2003).

5.5 Despite their general expectation of revisions, some users were surprised by the scale of the revision to quarterly GDP growth between m1 and m3 for 2003q2. Revisions to the construction estimates supplied to the ONS by the DTI were known to be behind the GDP revision. Revisions to GDP growth of this scale do affect the financial markets.

5.6 Other users were less surprised by the upward revision to growth in 2003q2, anticipating an upward revision. The Bank of England, for example, was clearly not particularly surprised by the upward revision since in the August 2003 minutes of the MPC (paragraph 14) it was written that: "The preliminary estimate of GDP growth in the second quarter had been 0.3%, somewhat lower than expected…it was possible that the preliminary estimate might be revised up slightly". Indeed in the September minutes (paragraphs 13 and 14) it is noted that the MPC was surprised that Q2 GDP growth had not been revised upwards: "The implausibility of this pattern of demand, and the prospect of further revisions to net trade, implied either that there would be some large downward revisions to components of domestic demand or that output growth would be revised up".

5.7 Contributing to some users' lack of surprise that growth was revised up for 2003q2 was the observation that the construction output growth estimate used by the ONS in the m1 and m2 GDP growth estimates was "too low". They felt that published qualitative information about construction (at m1) and quantitative information (at m2) conflicted with other signals and they also noted the imbalances between supply and demand in the economy. Moreover, some users did not share the view that the construction sector had shrunk in the previous quarter (2003q1). This implies that these users were basing their forecasts for q2 on a higher level estimate for construction than the ONS/DTI. For example, as stated in the July 2003 MPC minutes (paragraph 15): "the estimate for construction [in Q1] showed a steep fall, which appeared to reflect private non-residential constriction in particular. The extent of the estimated weakness of the construction sector was difficult to reconcile with some other available information: business surveys suggested construction growth had remained strong; employment in construction had risen in Q1; and the Bank's regional agents had not seen evidence of a steep decline in construction activity, although their contacts had reported a reduction in the growth rate."

5.8 Decisions, in general, are not affected by revisions since, as indicated, users do expect them. In particular, the Bank of England (including the MPC) and HMT, given how early estimates first estimates are produced, are cautious in their use of them. As stated, for example, in the Bank of England's August *Inflation Report* in 2003: "The MPC takes account of the likelihood that GDP data will be revised when deciding how much weight to put on the latest data". The Bank and HMT use various data sources, and various variables, when forecasting. The Bank, in fact, informally grade data, *inter alia*, according to their reliability; they even produce their

own early estimates of GDP growth.[10] Nevertheless, revisions do contribute to increased uncertainty.

5.9 Some users felt that the media attention paid to recent revisions to GDP and trade estimates had become more acute because the climate in which statistics have recently been received had been influenced by recent mistakes, such as pensions and Census, and controversy about ONS statistics in general centering around, for example, the Network Rail statistical classification which some users have called "the network rail bailout"; see Weale (2003) for discussion of this classification.[11]

5.10 Revisions do affect the public's general perception about the quality of statistics since they may not appreciate that revisions do not necessarily constitute mistakes.

5.11 Users expect the preliminary estimate of quarterly GDP growth to be revised upwards with the passage of time. Such a bias, whether statistically significant or not, is seen as a cause for concern.

5.12 Some users, such as HMT, stressed the distinction between the bias of historical estimates and the bias of current estimates. For example, the evidence is that m1 GDP estimates offer an unbiased estimate of the m3 estimate. But there is evidence that the m3 estimates have been biased estimates of the 'final' figures that are produced a number of years later. Nevertheless, while accepting that the m1 estimates are on average unbiased in the short run, HMT does form views about the way a particular m1 estimate will be revised by m3. Users in general hoped that any biases would be eliminated via either methodological improvement at the ONS or the use of bias corrections. It must be added that some users were concerned about the possible use of bias corrections. For further discussion of the use of bias corrections see Annex 3: Section G4.

5.13 The revision to the trade estimates due to the VAT fraud was not seen to be the ONS's fault, in the sense that criminal activity was responsible. Both the Bank of England and HMT were warned about likely revisions in advance of the publication of the revised trade estimates in July 2003. They welcomed this advice. For example, reflecting this advice the June 2003 minutes of the MPC (paragraph 15) state: "There were … doubts over the reliability of these [import] data in the past few years, so it was difficult to be confident about the underlying trends in imports and exports".

5.14 The VAT revisions did contribute to increased uncertainty about the reliability of statistics; e.g. as stated in the July 2003 MPC minutes (paragraph 17): "The assessment of the growth of demand in recent quarters was complicated by major revisions to the trade data".

[10] NIESR also produce their own estimates of GDP growth three weeks ahead of the ONS's first estimates; see http://www.niesr.ac.uk/niesr/gdp.htm for an overview and Mitchell, Smith, Weale, Wright and Salazar (2003) for a formal analysis.

[11] It should be noted that the description of this classification as a "bail-out" is misleading since no large amounts of new money have been put into the new company.

5.15 More generally, questions were raised by users about the speed at which HM Customs and Excise and the ONS reacted to the discovery of criminal activity. See Section 10 for further discussion.

5.16 Substantial revisions to time-series, such as those associated with methodological changes such as re-basing, are seen as inevitable, indeed desirable, but they do impose a burden on users. The implied re-writing of economic history necessitates extra work, such as re-estimation of economic models with the revised data.[12]

5.17 The ONS performs well relative to statistical offices in other countries. The ONS is commended for the speed of delivery of its GDP data. Users, in general, are happy with the current trade-off between the speed and accuracy of estimates. Users feel that the ONS did a good job at warning them in advance about the introduction of annual chain linking.

5.18 Users expected the introduction of annual chain linking to depress growth. To some extent they were then surprised that, in effect, growth was not depressed as the effects of the re-basing were outweighed by other factors. They raised the issue of whether different changes should be introduced simultaneously or one-by-one.

5.19 Labour Force Survey – these figures are grossed up using population figures. Since November 2002, an interim LFS grossing procedure has been used which ensures that all the key aggregate LFS time series, needed by Treasury and others for macroeconomic policy purposes, incorporate the latest population estimates one month after publication of the population estimates. Revised LFS microdata take a long time to produce because the systems available are out-of-date and seem, according to some users, to be "rather ramshackle". This is being addressed by ONS's modernisation plans.

5.2 Recommendations based on our interviews with the users

5.20 These recommendations follow from our interviews with the users. They should be read in conjunction with our recommendations in Section 8 that consider revisions to GDP growth, where many of the recommendations are repeated in the specific context of GDP.

5.21 The ONS and the Statistics Commission should think about how best to keep users informed about revisions to economic statistics and their causes (recommendation R.1). The ONS should explore means of changing the climate in which statistics are used, so that the public expect revisions and know that they

[12] To quote David Smith (2003) of Williams de Broë in *the parliamentary monitor*: "the frequency with which the ONS is introducing major changes to the conventions under which it produces the official statistics risks swamping the ability of private-sector users to cope and may be one reason why the ONS has appeared to be more accident prone in recent years".

should be expected and are not necessarily explained by an 'avoidable circumstance'. This should mean revisions generate less controversy than has sometimes arisen in the past. While we welcome the National Statistician's article of 13th January 2004 explaining the causes of revisions perhaps a similar exercise should be undertaken on a routine basis. One step in the right direction is embodied in the following recommendation.

5.22 The ONS should provide summary information on past revisions with early estimates (recommendation R.2). Users felt that more could be been done to inform them about likely magnitudes of revisions by presenting summary information on past revisions in press releases.[13] ONS has already announced an intention to comply with this recommendation (which has also already been made by the Statistics Commission); see Annex 3: paragraph 332 for details. This information could mean revisions generate less controversy than has sometimes arisen in the past. The Bank of England commended the ONS for its labour market release; this was seen as a good example of how the reliability of statistics should be communicated. Since the measure of uncertainty relies on an estimate of the sampling error a similar approach cannot be adopted for GDP. As well as increased discussion of historical revisions the ONS should look at how to reduce the size of 'regular' revisions and get to the final estimate quicker. More resources will probably be needed for this.

5.23 The ONS could go further and should consider grading data again according to their perceived reliability (recommendation R.4). To provide a simple, easy-to-read, indicator of the reliability of the data the ONS could consider grading data, say *A*-*C*, according to their reliability along the lines of the system of reliability gradings introduced in 1956.[14] An *A* grading indicates (with 90% confidence) that the reported figures are correct subject to a margin of error less than, say, 3% in either direction; *B* indicates an error of 3%-10% in either direction and *C* indicates that the error is greater than 10%. It is, however, essential that when this is done distinction is made between reliability of level estimates and reliability of growth estimates.

5.24 The ONS should keep copies of real-time data sets (i.e. of revision triangles) and make these publicly available via a dedicated revisions web-site (recommendation R.7). We understand that the ONS is hoping to adopt this recommendation; we encourage the ONS to do so. Real-time data sets provide a valuable resource not just for the ONS itself, when undertaking revisions analysis and assessing forecasting performance [e.g. see Annex A to this Volume], but for users too. The Bank of England, for example, has recently made available a real-time data

[13] See Annex 3: Section G2 for ONS's account of how they currently keep users informed about revisions.

[14] For further discussion of the British system, references and an excellent methodological discussion see Morgenstern (1963), p. 273.

set for GDP(E) that has proved popular; see Castle and Ellis (2002). A dedicated revisions web-site could also bring together other relevant information, such as: (i) dates and timetables for methodological changes; (ii) explanatory notes such as those in *Economic Trends* and simpler box-type explanations; (iii) aims and purposes of the series; and (iv) information on the reliability of the first release of the time-series.[15] Although much of this information is available elsewhere putting it altogether could be convenient speeding up users' access of information. This could in turn help users' understanding of revisions.

[15] We thank, in particular, Ian McCafferty of the CBI for his suggestions about the role of a dedicated revisions web-site.

6. Reasons for revisions

6.1 Revisions are caused by (i) pre-announced changes in methodology; (ii) the availability of additional information and (iii) avoidable circumstances.

6.2 Avoidable circumstances may occur because: (i) there are substantial mistakes in early processing; (ii) the models used to compute early estimates are not 'best practice'; (iii) processing timetables could be more rapid than they actually are and (iv) when the method used to obtain the estimates is "best practice" and theoretically appealing, but the method is implemented without sufficient resources to ensure adequate assessment of its practical performance so that unforeseen problems occur.

6.3 It is not always possible to distinguish the effects of methodological changes from additional information and avoidable circumstances because on occasion additional information may have been incorporated at the same time as methodological changes were made, making it difficult to unravel such changes without substantial input from the ONS.

6.4 The ONS regularly publish articles in *Economic Trends* examining the revision history of important time-series, like GDP growth. Moreover, recently an article from the National Statistician ["Revisions to statistics: their role in measuring economic progress"; 13/1/04] provided a general explanation for why revisions happen. It is explained that revisions "are an expected and inevitable consequence of the explicit trade-off made between timeliness and reliability of statistics". The National Statistician continues by stating that only occasionally do revisions arise "as a result of mistakes or 'errors' in the popular sense of the word". These errors fall into our category of revisions due to avoidable circumstances.

6.1 The causes of the key revisions: a summary

6.5 A detailed analysis of the three main revisions follows in Sections 8 (the GDP revisions) and 10 (the import revision). Let us now summarise our findings about the causes of these revisions. We consider each in turn.

6.6 Revisions to quarterly GDP growth typically are explained by pre-announced changes in methodology and the arrival of additional information. Given the speed at which the preliminary estimates are produced revisions are to be expected as more detailed information becomes available over time. The ONS provide warnings about revisions in the background notes and notes to editors accompanying the press release, reminding readers that the preliminary estimate of GDP is computed using incomplete information and that as a result quantitative estimates for the categories of GDP(O) are not yet of publishable quality. Furthermore, there are regular articles on

the revision history published in *Economic Trends*. These reveal that most of the revision occurs after the m3 estimate; see Akritidis (2003a,b). Let us now consider the causes of some specific recent revisions.

a. The revision of 0.3% to the 2003 Q2 estimate of GDP growth is large compared with previous revisions between the month one and month three GDP growth estimates. Over the previous 20 quarters there has only once before been a revision as large as 0.3%; moreover, this particular revision of -0.3% in 2002 Q2 has a special explanation, namely the Jubilee holiday, the full effects of which perhaps could not reasonably have been predicted by statisticians; see Annex 3: paragraph 42 for details.

b. The revision in 2003Q2 does not have a similar 'special' explanation. The upward revision to GDP growth was caused by the large upward revision to construction output between m1 and m3. Although revisions to construction estimates between m1 and m3 should be expected, it is important to test whether the revision could have been anticipated *ex ante*. If it could then the revision was at least partly due to 'avoidable circumstances'. Although users cannot expect the method used to obtain the m1 (and m2) construction estimates to be perfect, they should expect the method used to be the best available in the circumstances. The CBI, for example, notes in its submission to this Review that: "the process of creating an estimate for construction output seems obscure. Press comment implied a faulty equation". Given such doubts we investigate construction further in Section 9.

6.7 Revisions to economic time-series at the Blue Book stage are caused by pre-announced changes in methodology and the arrival of unforeseen additional information. More specifically, the changes to real growth that occurred at the 2003 Blue Book are explained by (i) annual chain linking; (ii) rebased deflators; (iii) rebased lower level volume series and other methodological improvements such as to the deflation of imported capital goods. These revisions are explained in detail in Annex 3: Section B.

6.8 We feel that the ONS did a good job at warning users in advance of the likely effects of these changes. A press release, for example, on 15 April 2003, sought to anticipate the effect of its introduction on economic growth as did papers in *Economic Trends*; see Tuke and Reed (2001) and Tuke (2002). As indicated above, users congratulated ONS on the manner in which it advertised the change. The success was also reflected in newspaper coverage; see Section 7.

6.9 Import and trade estimates. On 9th July 2003 the ONS published revised estimates for imports of goods to the UK from the rest of the EU that accounted for the effect of VAT missing trader intra-community fraud. Clearly, the arrival of unforeseen additional information, namely awareness of this criminal activity, explains this revision. Further discussion follows in Section 10.

6 Reasons for revisions

6.10 Other revisions:

a. Retail sales. Methodological improvement introduced in October 2003 has caused larger revisions and indeed "widened the scope for potential revisions"; see Annex 3: paragraph 147. These methodological changes were well flagged up in advance by the ONS via articles in *Economic Trends*, press releases and bilateral meetings with key users such as the Bank/HMT. Otherwise, revisions are caused by the arrival of additional information, reflecting the trade-off between timeliness and accuracy of data. In Annex 3: Section D the ONS provide further details about revisions to retail sales.

b. Index of Production; see Section 12 and Annex 3: Section C.

c. Census related: revisions to mid-2001 population estimates in October 2003. The 2001 Census in Westminster and related issues are the subject of another Statistics Commission Review; see Statistics Commission (2003b). See also the ONS's web-site [http://www.statistics.gov.uk/census2001/implications.asp] for discussion of revisions to population estimates.

d. The average earnings index. Subject to its own review in 1999; see Turnbull and King (1999). Since the Review there have been substantial methodological improvements; indeed the ONS now presents estimates of the precision of these estimates based on estimates of the sampling errors; see Youll (2002).

e. Employment.[16] The arrival of new information, such as the 2001 Annual Business Inquiry and the 2001 Census explain revisions to the number of jobs and LFS estimates.

f. Business investment. The upward revisions are largely explained by the arrival of better information one month after the provisional estimate. Later revisions are made as part of the Blue Book balancing process; see Annex 3: Section E and ONS Annex 1 for more details.

g. Pensions. In January 2002 avoidable circumstances, namely processing errors, led to a revision of over £100 billion to the estimate of pension funds assets and the decision to withdraw the series for several months. A subsequent review made a number of recommendations designed to prevent a recurrence. These recommendations have been applied widely across the ONS. Users also spoke of revisions due to "double-counting"

[16] The number of jobs in the economy can be estimated both from the Labour Force Survey, that measures the number of people in employment, and from the Annual Business Inquiry and employer surveys that measures the total number of jobs. Since people can hold more than one job the two measures need not be the same.

of pension contributions. However, in fact, to-date there have been no revisions related to double-counting of pension contributions. Recently, in January 2004 the ONS did acknowledge that there was probably double-counting but the effect will remain unquantified until new survey results become available; for details see http://www.statistics.gov.uk/about/Methodology_by_theme/downloads/Discussion_note_on_insurance_companies'_recording_of_pensions_transactions.pdf.

h. Capital stock. Errors in the compilation of the capital stock resulted in its temporary suspension in 2001. Users of the capital stock data, in this case the Bank of England, noted unusual features of the data in the course of their research into the measurement of productivity and capital services. The errors were avoidable in the sense that the published data used inputs (investment, prices) which were inconsistent with the National Accounts, though that would not have been obvious to a casual observer. Capital stock data are used widely by economists in measuring, for example, potential output via the production function approach. Some users regarded the ONS as treating the data as of secondary importance. However, the ONS have devoted more resources to the data since a consultation exercise in 1997 revealed that there was more external interest in the data than hitherto they had perceived.

i. Regional GVA. On 10 December 2002 the ONS withdrew the regional GVA data published on 21 November 2002 and published revised data on 22 January 2003. This followed on 3 December 2002 the ONS announcing an error within the regional Annual Business Inquiry published on 18 September 2002. The revisions were due to avoidable circumstances, namely substantial mistakes in early processing. In the regional Annual Business Inquiry data public sector organisations had been incorrectly included in the population in Section M (education) and Sector N (Health and Social Work). The ONS kept the public informed via press releases. While our focus is on national data this revision does revision does raise an interesting question discussed in depth in Section 11.

7. Newspaper comment: has it become more critical?

7.1 This section considers how the newspapers have reacted in the last eighteen months to recent revisions to economic statistics. A selective review of the newspaper coverage follows. We then make some observations.

7.1 Nov. 2002

7.2 The impact of the Jubilee celebrations on quarterly GDP growth is discussed. The reason why the Jubilee celebrations have contributed to revisions seems to be well understood. As the FT writes on 27 November ["UK growth revised up ahead of pre-Budget"]: "the ONS said that Jubilee celebrations during June pushed down output in the manufacturing and service sectors during the second quarter. Without this effect, growth would have been higher in the second quarter and lower in the third quarter, it said".

7.2 Dec 2002

7.3 In The Times on 12 December Patience Wheatcroft ["Time the ONS figured it out"] attacks the ONS over its withdrawal of revised regional GDP data. The ONS are accused of not being able to add up ["the poor things there [at the ONS] are having yet more trouble with their sums"] and of being subject to political pressures.

7.4 In The Times, 12 December, Gary Duncan ["ONS withdraws growth data after third blunder"] also mentions other revisions such as those to pensions and the census and talks of a "year of woes".

7.5 24 Dec 2002. The Independent [Philip Thornton: "Growth reaches highest level in three years"], The Financial Times [Scheherazade Daneshkhu and Anna Fifield: "Chancellor on target to beat growth forecast"] The Guardian [Charlotte Denny: "Royal flush to the economy"] and The Telegraph [George Trefgarne: "Britain shrugs off world gloom as growth surges"] consider the upward revision from 0.8% to 0.9% for Q3 GDP. The ONS is not attacked. The reason for the revision appears to be well understood. The articles consider the revision to construction output and the effect of the Jubilee celebrations. The effect of the revisions on the behaviour of the MPC are discussed often.

7.6 Downward revisions to consumer debt as a share of income are also discussed; as Philip Thornton writes: "At the stroke of a statistician's pen…the UK looks considerably less unbalanced than it did".

7.3 Jan 2003

7.7 On 24th Jan The Financial Times [Anna Fifield: "UK annual growth worst since 1992"] quotes an analyst who believes that the first estimates of GDP for 2002q4, based on output approach, to be too high. In fact the preliminary estimate of 0.4% was not revised at either month two or three.

7.4 Feb 2003

7.8 On 4th Feb Simon Briscoe in the Financial Times ["Growth rates of recent years face reduction"] notes that recent growth rates may be revised downwards with the introduction of chain-linking later in the year. The effect on forecasts is discussed and David Smith of Williams de Broë is quoted as saying: "forecasts made on the present basis will have little meaning in the strange, post modern world of the methodology". Mr Briscoe's analysis is based on preliminary work by the ONS. He notes that the actual results will be published in September, and anticipated in April.

7.5 April 2003

7.9 On 16 April 2003 The Scotsman ["NS to revise GDP figures downwards"] again talks of the fact that the ONS is expected to revise downwards its GDP growth estimates when chain-linking is introduced in summer. The article speaks of this being worrying news for Gordon Brown and notes that the revisions will damage the credibility of Mr Brown's growth forecasts.

7.10 On 16 April The Financial Times [Scheherazade Daneshkhu and Anna Fifield: "Economic growth to be cut when new measuring method begins"] says ONS predicts 2001 growth will be revised from 2.1% to 1.9% and estimates for 1999 and 2000 will be cut by 0.1% when annual chain linking is introduced.

7.6 July 2003

7.11 A letter to the Financial Times on 17 July by Alex Izurieta and Jonathan Ward explores the implications of the VAT fraud on policy making. Noting that the fraud meant imports were lower than expected, they argue that if the ONS is correct in their belief that the fraud will have little effect on GDP, then some other component of GDP must rise to compensate.

7.7 August 2003

7.12 On the 23 August 2003 the Western Mail ["Economic growth still failing to take off, reveals statistics"], Daily Record ["Stagnant economy"] and The Herald [Catherine MacLeod: "Growth figures fall short of Brown forecast"] consider the low growth

estimates for Q1 and Q2 of 0.1% and 0.3%, respectively, and the failure of the second estimate for Q2 to revise up the preliminary estimate. The implications for Mr Brown are examined.

7.8 September 2003

7.13 On 13 Sept 2003 Ed Crooks in The Financial Times ["From here to recovery"] reminds us of Lord Lawson's attack on economic statistics that were "little more than a work of fiction" for leading him astray. This is in response to ONS's upward revision to GDP growth between 1998 and 2000 amounting to 1.3%. As Ed Crooks writes: "The ONS, however, has discovered 1.3 percentage points of additional GDP growth between 1998 and 2000. Other things being equal, the output gap that supposedly exists at the moment would be wiped away at a stroke". The implications of these revisions on policy-makers are then discussed further; the revisions mean that policy errors are possible as policy makers are unsure of the position of the business cycle. Ed Crooks writes: "the revised ONS figures are likely to change the way the Bank now views the balance of risks". The FT also has an article entitled "An exceptional revision" that notes how the expected effect of the introduction of annual chain-linking on GDP growth estimates has been outweighed by other factors.

7.14 15 September 2003. The Guardian [Heather Stewart: "ONS replay suggests that Brown was not offside"] writes of an "ONS replay", since revisions due to chain-linking have altered the sectoral picture of the UK economy.

7.15 On the 28 September 2003 the Sunday Express [Megan Davies: "New growth stats give Brown reason to cheer"] expects an upward revision to GDP growth for 2002. In fact growth ends up being revised down from 1.9% to 1.7%. The expectation of an upward revision is based on the earlier upward revisions to GDP growth for 1998-2000. Philip Shaw is also quoted as saying: "The pattern of revisions over the past four to five years has been in an upward direction, so it would be sensible to suggest 2002 would go the same way".

7.16 On the 30 Sept. 2003 the Evening Standard [Steve Hawkes and Jane Padgham] lead with an article entitled "Doubling of UK growth estimate stuns the City". This is in response to the upward revision to 2003 Q2 growth from 0.3% to 0.6%, due to revisions to construction output. To quote: "The City was braced for volatility in today's growth data after a move by the ONS to change the base year for the series to 2000 from 1995 and re-base the series annually from now on. But the bulk of the upwards revision was put down to figures that showed construction output in the second quarter was 4.4% instead of a much more conservative earlier estimate of 0.8%. The ONS said the change reflected new data supplied by the Department of Trade and Industry but it will still be seen as a major embarrassment to chief executive Len Cook and add to doubts over the credibility of its numbers".

7.17 The Evening Standard also published a box called “The ONS foul-ups file” that catalogued recent troubles to ONS statistics. The troubles mentioned are: the census (the latest census figures may under-estimate UK population by up to one million), imports (due to the VAT fraud), the volatility of retail sales, the double-counting of pensions and the issue of whether the Network Rail bail-out should be classified as government debt.

7.9 October 2003

7.18 From 1 Oct. 2003 onwards, at least for a week, much was written in response to the 2003Q2 quarterly growth revision; the earlier Blue Book revisions to annual GDP growth back to 1998 [e.g. see paragraph 7.13] also contributed to the comment. For example, on 1 October The Independent [Philip Thornton: “ONS forced to double growth estimate”] writes that “government statistician’s admitted to errors in key economic figures stretching back over several years”. It is said that these revisions “will anger the Treasury and Bank of England”. The Independent also writes [Philip Thornton: “Errors deal fresh blow to a battered credibility”]: “The scale of yesterday’s revisions is a fresh blow to the credibility of the Government’s statistics service, which have been bedeviled by a string of problems”. This string of problems includes the census, the volatility of retail sales, the double counting of pension contributions and the average earnings index.

7.19 The Independent continues [Jeremy Warner: “Another right royal ONS screw-up has the Bank fuming”] its attack with the assertion that the doubling of growth in 2003q2 was “scandalous”. Mr. Warner continues: “The ONS has become a comedy of errors. It double counted the stock of pension wealth in the economy, it misstated the trade figures by failing to understand the importance of the carousel mobile phone fraud in the amount of imports coming into the country, the retail sales figures have fluctuated so widely that nobody can take them seriously any more, and it ridiculously sided with the Government in ruling that Network Rail loans should not count as part of the public finances, despite the fact that they carry a Treasury guarantee”.

7.20 A more understanding account of the revision is given in the Telegraph on 1 October (“Brown almost off the hook on growth”) where the revision is associated with the move to annual chain linking. The Financial Times on 1 October [Anna Fifield: “Revised figures show economic surge” and Ed Crooks [“Brown’s figures do add up but there’s a snag”] also presents a less critical account focussing on the implications of the revision for the economy. It is explained that revisions to construction output lie behind the GDP revisions.

7.21 The Herald on 2 October [Alf Young: “Vital statistics which quite simply do not measure up”] launches another attack: “professional integrity is at stake. If confidence begins to be lost in what National Statistics is telling us, how can other independent forecasters have confidence in their own predictions?” Mr Young also writes: “National Statistics is becoming dangerously accident prone”. He reminds readers of the suspension of key series and the regional GVA data revisions.

7 Newspaper comment: has it become more critical?

7.22 On 4 October The Times [Graham Searjeant: “Sorry, wrong numbers – I’ll say that again”] writes of “grinding teeth” at the Bank of England in response to the doubling of 2003Q2 growth. Mr Searjeant explains, however, that this may not have caused the MPC to misjudge the position of the economy since the MPC “feared that the [initial] figures might be wrong and made their own judgement” but private investors may have been fooled. He writes: “The many heavy revisions coming out of the ONS reflect an improved technical method as well as one-off mistakes, such as Whitehall underestimating smuggling. The new figures should be more reliable, but the lesson is not to rely too much on any kind of figures”.

7.23 On 5 October The Sunday Times continues with an article by Irwin Stelzer entitled: “Forecasters who don’t even know where they’ve been”.

7.10 Observations

7.24 Based on this selected survey of the newspapers’ reaction to revisions to economic statistics from November 2002 to October 2003 it seems fair to say that press comment has become increasingly critical.

7.25 This increasingly adverse comment reached its zenith with the 0.3% revision for 2003q2 GDP growth between the preliminary and third estimates. This particular revision, and the recent revisions to annual GDP at the Blue Book, sparked a series of attacks on the ONS itself. In the absence of this one revision to quarterly GDP growth we believe the press comment would not have become nearly as critical as it did.

7.26 These attacks on the ONS gained additional force by reminding readers of previous revisions, or what were often called “errors”, such as those to trade, the average earnings index, regional GVA estimates, population estimates related to the census, pensions and retail sales.

7.27 Certainly the revisions to the regional GVA estimates and the average earnings index were due to “avoidable circumstances”. Furthermore, as we argue below in Section 9, although the revision to the GDP growth estimate itself is explained by the arrival of additional information, namely updated construction estimates, we do believe the model used to estimate construction growth at m1 was not “best practice”, and so again appears to fall at least in part into the avoidable circumstances category. The newspapers also seemed surprised that the construction forecast produced by the DTI at m1 could be so off the mark.

7.28 However, in fact, to-date there have been no revisions related to double-counting of pension contributions. Recently, in January 2004, the ONS did acknowledge that there was probably double-counting but the effect has yet to be quantified since the ONS await the latest survey results. Therefore, the press claims of “double-counting” of pension contributions reflects speculation about the quality of the data [apparently prompted by a letter from David Willetts M.P. to The Times on 10 June 2002] rather than an actual revision made by the ONS.

7.29 The newspapers, as we should expect, have 'hyped' up these recent revisions to economic statistics, for example, by exploiting the fact that the revision to GDP growth in 2003q2 constituted a 'doubling' of the rate of growth. This although factually correct is misleading reporting. On the same basis a similar revision from 0.1% to 0.3% would be a trebling. Moreover, newspapers do not find it easy to distinguish those revisions that were "avoidable", such as those to regional GVA estimates, the average earnings index and pension assets (but not contributions), from revisions that were not, such as the Blue Book revisions to annual GDP growth and the revision to the import estimates due to the VAT fraud. This confusion obviously creates an unfairly bad image of ONS. Nevertheless, the newspapers appear to be, at least in general, aware of the reasons for revisions and have often reported them uncontroversially.[17]

7.30 It is likely that if someone were to form a view about the performance of the ONS based on examination of the newspapers only they would conclude that the ONS performs badly. This is obviously a matter for individual judgement. However, relative to offices in other countries the ONS performs well both in terms of the speed of delivery of its estimates and revision patterns.

7.31 Moreover, the newspapers' view that the Bank were "fuming" about the 2003q2 growth revision [see paragraph 7.19] is, at the very least, inconsistent with what the Bank has indicated to us in the course of this Review.

7.32 Increased public awareness about the reasons for revisions, explaining to them that the majority of revisions are not 'avoidable' but are a natural, even desirable, feature of data, could help create a more understanding climate in which revisions, including those that are due to avoidable circumstances, are received. ONS advertising of the introduction of annual chain linking was successful in the sense that the newspapers reported early ONS results on its likely effect on GDP estimates.

[17] Nevertheless, as footnote 12 indicates some users clearly feel that the public generally may be unable to distinguish "good" revisions from "bad" revisions (namely those due to avoidable circumstances). As a result the public may have felt that the recent methodological changes (such as chain-linking) were further examples of the ONS becoming "more accident prone", which of course they were not.

8. Revisions to GDP growth

8.1 In this Section we explain in detail revisions to both quarterly and annual (Blue Book) GDP growth estimates.

8.1 How and why do revisions happen?

8.2 GDP data, like many economic statistics, are continually revised. As recent studies by the ONS have explained, these revisions occur at different stages of the GDP compilation process; see Richardson (2002) and Akritidis (2003a,b). The preliminary estimate of GDP, produced just one month after the end of the quarter, is revised in month three as more complete survey data for the output, expenditure and income measures of GDP become available; see Reed (2000) for details of how the preliminary estimate is constructed. Annual figures, providing more detail than the quarterly releases are published in the September following the year in which they relate in the Blue Book. This provides an opportunity to consolidate sources which had not hitherto been available (such as information from the Inland Revenue for the fiscal year) and it is likely that revisions will be needed to the quarterly data at this stage. Thus this revision occurs six to eighteen months after the month three estimate has been published. In these data estimates of GDP on the income side do not agree with those on the expenditure side; a residual is included. A year later at what the ONS calls the "Blue Book Two stage", typically eighteen to thirty months after the preliminary estimate is published, these estimates are again revised. The revision makes use of input-output data and from the ONS Annual Business Inquiry together with revised Inland Revenue income estimates. At this stage the discrepancy between income and expenditure is removed. Annex 3: ONS Annex 1 explains this process in the clearest statement yet. Post Blue Book Two revisions occur too, as input-output supply and use balancing is run again, and methodological changes are introduced. Finally estimates of output growth change when data are rebased; this has happened every five years since 1970 and was less frequent before then. The nature of rebasing is discussed below. In September 2003 the ONS changed from producing data in constant prices to producing chain-linked volume indices. This change, discussed below, means that re-basing will now happen every year.[18]

[18] It should be added that despite the move to annual rebasing at a high level, price indices and weights aggregated below the level of chain linking will still be re-based on a five yearly cycle.

8.3 Revisions to economic statistics are not a new phenomenon and are not undesirable *per se*. In the early 1990s the ONS rationalised its revisions timetable in order to make it easier for users to understand when revisions were likely. Nevertheless, recent revisions between the first (preliminary) and third month estimates of GDP growth have attracted much comment. This is in part because revisions due to rebasing came at much the same time as the large revision to the 2003Q2 estimate of GDP growth, creating a general impression that data were subject to large revision. This section places these recent revisions in their context. This is achieved in the following four ways:

(i) Examining the evidence for whether revisions between the first and third month have got greater over time. We see whether users should have been surprised by the revisions in 2003q1 and 2003q2. This involves calculating the *ex ante* probability of there being a revision error as large as 0.3%.

(ii) Considering revisions to GDP growth estimates beyond the third month.

(iii) Examination of revisions at re-basing.

(iv) Revisions to GDP growth in the UK are placed in an international context.

8.1.1 ONS studies

8.4 ONS carries out regular assessments of the nature of revisions to its estimates of GDP and its components. We focus on aggregate quarterly GDP growth figures (percentage points at a quarterly rate). Recent studies of GDP growth estimates are Richardson (2002, 2003) and Akritidis (2003a,b), published in *Economic Trends*. Akritidis (2003b), updating Richardson's (2002) work, looks at revisions by stages of the GDP compilation process, as discussed above. Focus is on the period 1993q1-2000q4. The starting point is dictated by the fact that the first preliminary (month one) estimate of GDP, corresponding to 1993 Q1, was published in April 1993. Before then initial estimates were published after a longer lag.

8.5 The latest data used were published in September 2003 and therefore include those revisions caused by the introduction of annual chain linking and the correction of VAT fraud. We should therefore expect revisions to be larger than those arising when we consider a data vintage prior to these changes. This is indeed found to be the case by Akritidis (2003b); see Sections 8.1.3-8.1.5 below for further analysis of the effects of re-basing and annual-chain linking.

8.6 Table 8.1 summarises the findings by reporting the mean and variance of the revision, "t-test" which is a statistical test for bias, and the RMSE. The t-test is adjusted to allow for serial dependence in the pattern of revisions. The 95% critical value is just over two; see Akritidis (2003b) for details.

8.7 It should be added that ONS test for bias, that is whether the mean is statistically significant from zero, using a t-test that allows for first-order dependence in the revisions process. Below, we use a different form of the t-test that does not require an assumption about the parametric process governing the dependence. This is attractive if one is unsure about the precise nature of the dependence; using our non-parametric test results using the t-test remain 'valid' (strictly the estimator is consistent) even if the true revisions process is not a first-order auto-regression. Specifically, we use the non-parametric test of Newey and West (1987). For a review of some more recent tests in a similar vein see Hamilton (1994). The ONS should keep abreast of the latest developments to ensure they are happy with the specific tests they employ.

Table 8.1: Revisions at each stage of the quarterly constant price GDP growth estimate; adapted from Akritidis (2003b). Percentage Points at a Quarterly Rate

1993q1-2000q4	Mean	Variance	t-test	RMSE
Month three less month one	0.013	0.010	0.728	0.100
First edition of Blue Book (BB1) less month three	0.059	0.026	1.808	0.170
Second edition of Blue Book (BB2) less BB1	0.013	0.039	0.466	0.197
Revisions after BB2	0.103	0.122	2.171	0.365
Total	0.188	0.112	3.322	0.383

8.8 Table 8.1 shows that consistent with the earlier findings of Richardson (2002), "the bulk of the upward revision to the month one or preliminary estimate occurs after the first three monthly releases". Indeed, there is statistical evidence for bias at all stages of the GDP compilation process aside from the month three less month one revision and BB2 less BB1 as evidenced by t-values greater than the critical value of about two.

8.1.2 Updated analysis of the revision between the first and third month estimates: have revisions got larger?

8.9 Although it is not possible to update Akritidis' analysis for all stages of the GDP compilation process, as one needs to leave, say, about three years at the end of the sample to give the data time to 'mature', we can update analysis of the revisions between the preliminary (or month one) estimate and the month three estimate. We update Akritidis' data to 2003q2 using consecutive press releases from the ONS. This is particularly relevant given the recent revisions at this stage of the GDP compilation process seen in the Q1 and Q2 revisions in 2003.

8.10 Figure 8.1 plots the time-series of revisions between the month three and month one GDP growth estimates. In the recent past two larger revisions have been observed. The -0.3% revision in 2002 Q2 we know has a specific cause due to the Jubilee effect. Holiday patterns and therefore working day and seasonal adjustments were changed by the Jubilee holiday. Statisticians did not, and probably could not have reasonably, anticipated the effects of this fully; see Annex 3: paragraph 42 for details. There does not seem to be any similar 'special' explanation for the revision to the 2003Q2 data.

8.11 Table 8.2 confirms the impression given by Figure 8.1. The mean revision, or bias, between the month three and month one estimates has risen from 0.005 percentage points using data up to 2003q1 to 0.012 percentage points using data up to 2003q2. But these figures are very small and we cannot sensibly reject the hypothesis that the bias is insignificantly different from zero as evidenced by p-values for the t-tests greater than 0.2.

8.12 The revision of 0.3% points happened in an environment where figures are rounded to one decimal place. Thus it would have happened with a true revision of 0.25% points or more. We therefore compute the probability of a revision error of greater than or equal to +0.25% and less than or equal to -0.25%. This probability is computed both fitting a normal density and a nonparametric kernel density to the time-series of revisions. The nonparametric kernel density may provide a better fit of the data if the revision errors are not normally distributed. Estimates are presented for the three sample periods; see Table 8.3.

Figure 8.1: Revision to GDP growth estimates between month three and month one [Percentage Points at Quarterly Rates]

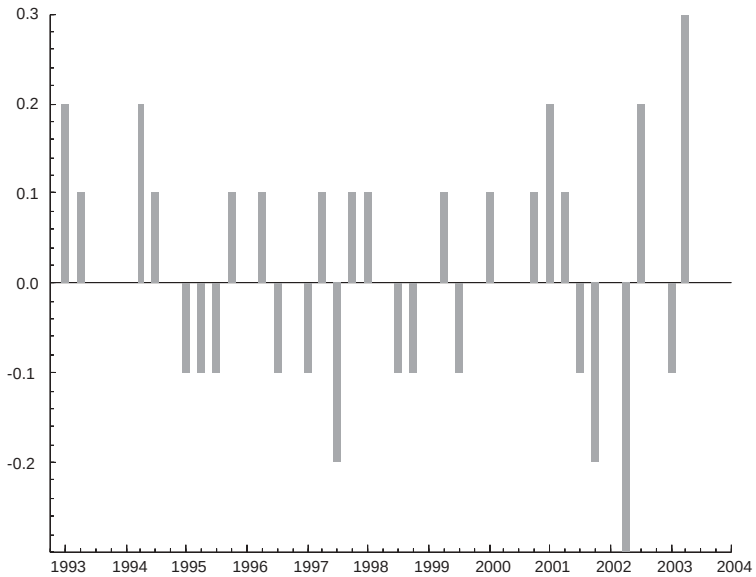


Table 8.2: Revisions to GDP growth estimates between the month three and month one estimates. (Percentage Points at Quarterly Rate). t-test (p-value) is the p-value for the statistical test of whether the bias is significantly different from zero[19]

	Mean	Variance	t-test (p-value)	RMSE
1993q1-2003q1	0.005	0.013	0.385	0.116
1993q1-2003q2	0.012	0.015	0.235	0.124

Table 8.3: The probability of a revision error at least as large as ±0.25% occurring between the month one and month three GDP growth estimates (Percentage Points at Quarterly Rate)

	Prob(revisions) greater or less than ±0.25%	
	Normal	Kernel
UK 1993q1-1999q4	2.14%	1.41%
UK 1993q1-2003q1	3.15%	5.09%
UK 1993q1-2003q2	4.37%	7.13%

8.13 Table 8.3 shows the probability of a revision error at least as big as 0.25% increasing over time. However, the probability is low. It appears users were justified in feeling surprised by the 2003q2 revision. Before this revision, there was no more than a five percent chance of it occurring.

8.14 This result is consistent with the finding of Drs. Garratt and Vahey in their submission to this Review. They find the revision of 0.3% lies outside the predicted 95% confidence band given an initial estimate of 0.3%; see Annex 2 for further details.

8.1.3 Rebasing and Chain-linking

8.15 Economic data can be collected reasonably unambiguously in current prices. However users of data need to know how the volume of output or the volume of activity is changing, and this means removing the effects of price changes from changes in value measured at current prices. Since a large and changing mix of disparate goods and services is produced by the economy it is a priori unlikely that there will be a single satisfactory way of measuring changes in the volume of output (and equally that there will be a single satisfactory way of measuring changes in "the price level"). Perhaps the most obvious way of identifying volume changes is to ask how much the value of output would be changing at constant prices. In other words, if one chose the prices of a base year and applied these to all the transactions in the

[19] A Newey-West type robust estimator is used; see Newey and West (1987).

economy (the Laspeyres index) it might be thought that the effects of price changes would be removed. In some sense this is true but it is also easy to see the pitfalls. The more stale are the prices, the less relevant they are to the modern economy. If we measured the changes in the volume of output using prices from 1948 we would not expect to obtain the same answer as if we used prices from 1995; we may nevertheless want to know how the volume of output has changed since 1948. A good data system should enable us to answer long-term questions of this sort as well as short-term questions about what has happened in the last quarter. Prices may not only become stale; there may be some goods which did not exist even a relatively short time ago and therefore a single price base is likely to be misleading.[20]

8.16 It is not only that stale prices are misleading; it is also likely that the goods whose output has expanded most in volume terms are those which have become cheaper. Thus a measure of output at constant prices is likely to overstate the "true" growth of the economy. This problem has been particularly acute with computing equipment in the last twenty years and ad hoc solutions have been adopted to deal with this. The conventional solution, employed up until 2003, has been to change the price base every five years. Until 2002 estimates were in 1995 prices. In 2003 the ONS moved to annual chain linking. This means the price base will never again be out-of-date.[21]

8.17 From 1993 to 1997 the ONS worked to 1990 prices. In 1998 they had enough information on the economy in 1995 to be able to move to 1995 prices, and they published data in constant 1995 prices until September of this year. However, for the distant past prices fairly close to the present are as inappropriate as are 1948 prices to today's economy. Thus, to provide the long runs of data which are essential for many users, the ONS uses a link year. They calculate the value of 1993 output in both 1990 prices and 1995 prices. The figures for 1988 to 1992 calculated in 1990 prices are multiplied by the ratio of 1993 output in 1995 prices to 1993 output in 1990 prices. This gives the figures for 1988 to 1992 which are described as being in 1995 prices, although that is obviously not in fact the case. Similar procedures are used to go further back. The figures for 1983 to 1987 are evaluated in 1985 prices and then scaled up to convert them to 1995 prices. Thus there are a number of linking years involved in the construction of the series for the volume of output back to 1948.

8.18 This linking process has one unfortunate consequence. Output measured in constant prices should equal expenditure in constant prices because everything which is produced is sold (or added to stock which is also allowed for). This is true for the data which are genuinely in 1995 prices. The linking process means, however, that this adding up property is lost in the data for the earlier years.

[20] For a helpful introduction see Cahill (2003).

[21] Strictly, the price base will still be out-of-date since it will be always out-of-date by up to three years (i.e. 2003 numbers were at 2000 prices). However, annual chain linking means the prices will be less out-of-date than before: at the level of annual chain linking the price base will be a maximum of three years out-of-date compared with the eight years previously.

8 Revisions to GDP growth

8.19 The process of linking and rebasing has been justified on the *ad hoc* grounds that stale prices are not likely to provide a sound basis for measuring volume changes. There is, however, a powerful index number result which should be mentioned. Samuelson and Swamy (1974) showed that a satisfactory volume indicator of changes in output can be produced if and only if output is produced with constant returns to scale (doubling all inputs doubles output). In this case the appropriate index is one in which the weights are continuously updated. The percentage change in this ideal volume index is equal to the sum of the percentage change in the volume of each output multiplied by the share of the value of each output in the total value of output (the Divisia index). Because current value shares are always used, the base never becomes stale and rebasing is never needed. The index can be thought of as a chain comprised of successive percentage changes and is therefore often described as chain-linked. In the ideal world of continuous updating the length of each link would shrink to zero; the old system of rebasing can be thought of as chain-linking with links five years long. Nevertheless the term chain-linking is usually adopted to describe indices in which the links cover no more than one year.

8.20 The percentage growth in the index between 1998 and 1999 is measured using the percentage change in the volume of each output weighted together by the share of each output in the total. But should one use the 1998 share or the 1999 share. The recommended procedure (the Törnqvist index) is to use the arithmetic average of the shares in the two years between which output growth is measured and is justified as the best practical approximation to the Divisia index. Output growth between 1999 and 2000 is calculated by weighting the percentage change in the volume of output of each good or service by the average of the shares of each good in the total for 1999 and 2000. Compounding the percentage change from 1998 to 1999 with that from 1999 to 2000 allows one to calculate the percentage change in output between 1998 and 2000. In September 2003 the ONS applied this procedure to recalculate an index for the volume of output in every year back to 1948. The index is set to 100 for 2000 purely as a statistical convenience.

8.21 At present the ONS does not have the information it needs to calculate the value shares for 2001. Thus output growth since 2000 is calculated using 2000 prices and can be viewed as a traditional fixed base (Laspeyres) tail to the annually chain-linked index. This means that output growth up to and including 2001 is calculated using previous years prices (PYPs). Next year, value shares for 2001 will be available, and ONS will be able to calculate growth using PYPs up to 2002. Figures for 2002 onwards will be at 2001 prices. The ONS will also move forward the reference year so that indices will be published at 100 for 2001.

8.22 This change means that the quinquennial rebasing we have seen in the past will no longer happen. Instead there will be steady (and probably smaller) changes as for example, the estimate of growth in 2002 relative to 2001, currently calculated at 2000 prices, is replaced first by a figure at 2001 prices and then by one using the average of the 2001 and 2002 value shares.

8.1.4 A Comparison of the 2003 and earlier rebasing

8.23 While rebasing has happened for the last time, we nevertheless explore how the changes to output growth which occurred at the time (but which were not necessarily due to rebasing) compare with those implemented in previous rebasing exercises.

Figure 8.2: Size of revision in annual growth rate (in percentage points) when the base period changes as a function of the number of years before the change

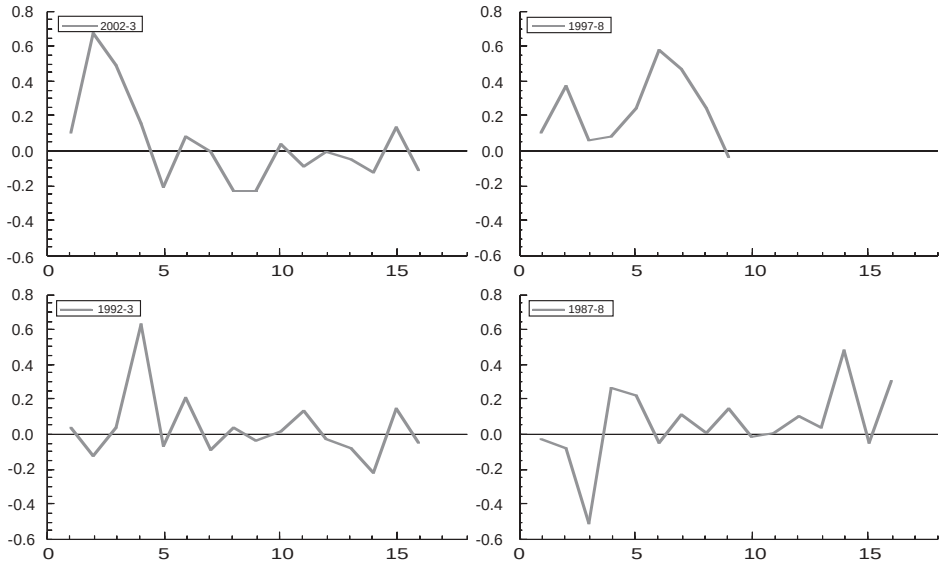
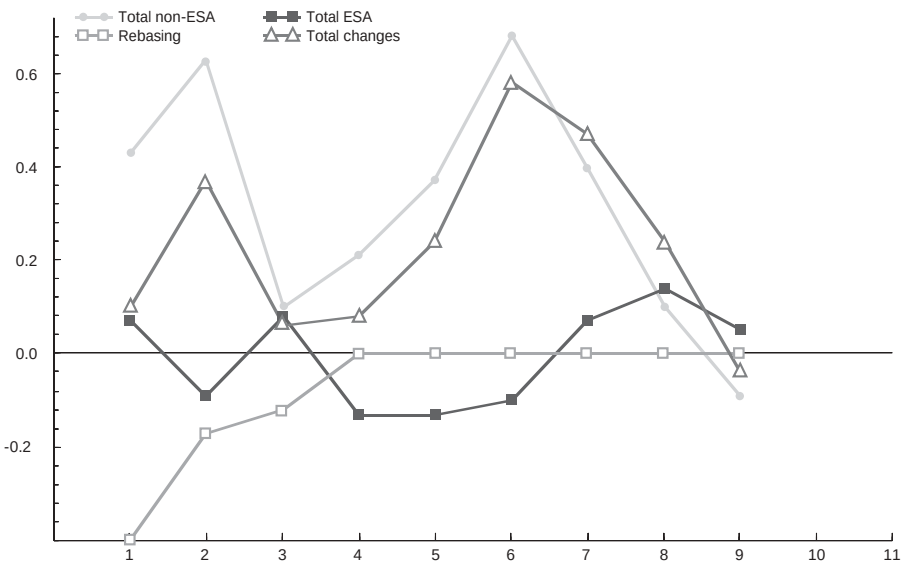


Figure 8.3: Decomposing the source of revisions: evidence from the 1998 Blue Book



8 Revisions to GDP growth

8.24 Figure 8.2 plots the revision between the annual GDP growth estimates presented in Blue Book estimates either side of the last four changes in base. We see that the 2002-3 revisions (associated with annual chain-linking) are far from atypical. Indeed it looks quite similar to the revision in 1992-3. There is perhaps some indication, however, that the revisions associated with the re-basing in 2003 are slightly large than the other re-basings as the area under the line (the integral) appears larger in 2002-3 than certainly 1992-3 and 1987-8.

8.25 Unsurprisingly Figure 8.2 suggests that the larger revisions are for the more recent years. Some indication of how much of this revision is due to re-basing is presented in the 1998 Blue Book where the revisions are decomposed by source. Figure 8.3 plots this decomposition. It reveals that re-basing, as economic theory predicts, does cause the more recent estimates to be down-weighted. However, other changes are far more important.

8.1.5 The effect of the introduction of annual chain linking on quarterly GDP growth estimates

8.26 As expected, annual chain-linking also caused larger revisions to quarterly GDP growth. Comparing the results of Table 8.1 with those from Table 8.4, that uses data prior to the introduction of annual chain-linking as the latest data (i.e. data for June 2003), we see that the introduction of annual chain-linking increased the RMSE at the post BB2 stage from 0.179 to 0.365. The total revision therefore increased from a RMSE of 0.308 to 0.383.

8.27 In the long run we should expect annual chain-linking to decrease the typical size of revisions that used to occur at the five yearly re-basing.

Table 8.4: Revisions at each stage of the quarterly constant price GDP growth estimate: using June 2003 as the latest data; adapted from Akritidis (2003b). Percentage Points at a Quarterly Rate

1993q1-2000q4	Mean	Variance	t-test	RMSE
Revisions after BB2	0.078	0.026	3.309	0.179
Total	0.163	0.068	2.554	0.308

8.1.6 International Comparison

8.28 It is important to place recent revisions to quarterly GDP growth in the UK into context by considering GDP growth revisions for other countries. To this end we survey two existing studies by Faust et al. (2000) and Statistics Sweden. Thirdly, we use a publicly available data-set for the US (available from the Federal Reserve Bank of Philadelphia's web-page) to provide some updated analysis that, as far as possible, focusses on the revisions at the same stage of the compilation process as the month three minus month one revision in the UK.

8.29 Annex 3: Section G1 also provides a detailed and thorough response to our question put to the ONS about international studies of GDP revisions. The ONS share our view that it is important to place revisions to UK statistics in their international context. The ONS also point out two dangers of international comparisons, both of which we agree with. First, different countries produce data at different speeds; one therefore has to be careful that one is comparing like with like. Secondly, different countries will have made improvements to their national accounts at different points in time. A fair comparison of different countries' performance rests on being able to find a sample period that ensures like is also compared with like within countries.

8.30 The results below are therefore subject to the criticism that historical analysis of revisions is a poor indicator of current performance.

8.1.6.1 Study 1: Faust et al.'s (2000) comparison of the G7 economies

8.31 Faust et al. take their data from consecutive issues of the OECD's Main Economic Indicators. The final growth rate is taken from the April 1999 publication. The revision is defined as the difference between the final and preliminary estimates. It should be acknowledged that some countries, such as the UK, produce their data more quickly and this may in part explain the higher mean revision for the UK in Table 8.5. Indeed the UK is the only country where there is statistical evidence (at a 95% significance level) of bias. This finding of bias is consistent with the ONS's own results; see Akritidis (2003a,b).

Table 8.5 Study 1: International comparison of revisions to quarterly GDP growth estimates (percentage points at a quarterly rate): 1988q1-1997q4. Note t-tests are based on Newey-West robust standard errors (results taken directly from Faust et al.)

1988q1-1997q4	Canada	France	Germany	Italy	Japan	UK	US
Mean	0.01	-0.02	0.20	-0.01	0.05	0.22	0.02
t-stat	0.21	-0.28	1.61	-0.19	0.61	2.08	0.41
RMSE	0.34	0.34	1.08	0.52	0.80	0.73	0.33

8.32 It should be noted that the finding in Table 8.5 of a RMSE of 0.73 for the UK, although not too unflattering in the sense that two other countries have a higher RMSE, does appear out-of-line with the findings of the ONS. The RMSE for the US in Table 8.5 is, however, broadly similar to that we find below (see Table 8.7; for the US we find a RMSE over the period 1993q3-1999q4 for the total revision of 0.361).

8.33 Akritidis (2003b), as shown in Table 8.1, found a RMSE for the total revision in the UK of 0.38. The higher RMSE of 0.73 found by Faust et al. for the UK is explained, at least in part, by the different sample period. This is seen in Figure 8.4. Figure 8.4 re-considers the Faust et al. data set but plots the RMSE as a function of the starting date of the sample period. The end point is kept fixed at 1997q4. We see for 1988q1, the first starting date, a RMSE of 0.73, consistent with Faust

et al.'s results presented in Table 8.5. The RMSE is then seen to decline rapidly. For example, starting the analysis just four quarters later in 1989q1 the RMSE is 0.313.

8.34 This finding is consistent with the results presented in Annex 3: paragraph 187 using forecast efficiency tests; see also Richardson (2003).

8.35 Both our results and those of the ONS suggest that the UK has improved its performance dramatically since the 1989 Blue Book. Following the Rayner Review in 1981 the production of data in the UK ran into substantial difficulties; data possibly became less reliable. In the late 1980s there were some substantial revisions that were a source of concern, culminating in publication of the Pickford Review in 1989. Since then there have been a number of methodological changes as well as improved data sources (both new surveys and larger samples for existing surveys). These changes have delivered more accurate estimates.

8.36 These estimates for the UK also compare favourably with those for the US, the only other country that produces GDP estimates at a comparable speed to the UK; see Section 8.1.6.3 below.

8.37 Despite this apparent improvement in the UK estimates since 1989, Garratt and Vahey (2003) do not find, using formal statistical tests, evidence that the Pickford Review coincided with a structural break and the transition to a 'regime' with lower bias. This suggests that further work on this issue may be required before it can be completely resolved.

Figure 8.4 Re-considering the Faust et al data for the UK. The RMSE as a function of the starting date.

8.1.6.2 Study 2: Statistics Sweden

8.38 The second study we examined was carried out by Statistics Sweden; see Öller and Hansson (2002). We use the raw data from Table A5 in their report that provides the preliminary and final estimates of growth for some selected countries for the years 1980 to 1998. Statistics Sweden focus on short-term revisions by examining revisions between, for the UK what amounts to, the third Blue Book and the first time the annual data are published. They thus try to avoid, as much as they can, revisions arising due to methodological changes. Table 8.6 then summarises the revisions properties of these growth estimates.

Table 8.6 Study 2: International comparison of revisions to annual GDP growth estimates (percentage points at an annual rate): Note t-tests are based on Newey-West robust standard errors

1980-1998	Canada	Germany	UK	US	Sweden
Mean	-0.047	0.000	0.263	0.126	0.358
Variance	0.188	0.437	0.261	0.268	0.159
t-stat (p-value)	0.363	0.500	0.069	0.162	0.010
RMSE	0.436	0.661	0.575	0.533	0.536
RMSE: 1990:1998	0.545	0.838	0.268	0.559	0.579

8.39 Table 8.6 indicates that the UK has a relatively large mean revision to its annual GDP growth estimates over the full sample period. Indeed there is statistical evidence for bias although Sweden's estimates are more biased. However, looking at the overall reliability of the estimates, as measured by the RMSE, the revision for the UK appears to be of a similar order to the other countries.

8.40 Consistent with the findings using the Faust et al. data, the performance of the UK estimates improves if the revisions in the 1980s are excluded. Over the period 1990-1998 the UK estimates have a RMSE of 0.268. This is both lower than over the period 1980-1998, when there was a RMSE of 0.575, and lower than the RMSE of any other country over the period 1990-1998. Indeed, the RMSE of the UK over the period 1990-1998 is half that of any other country.

8.1.6.3 Study 3: Updated analysis for the US using the "Real Time Data Set for Macroeconomists" compiled at the Federal Reserve Bank of Philadelphia

8.41 Using the "Real Time Data Set for Macroeconomists" compiled at the Federal Reserve Bank of Philadelphia we examine the revisions between the month one (called the advance estimate by the BEA), month three (somewhat confusingly called the final estimate) and final estimates of quarterly GDP growth, computed after an annual or comprehensive revision; for more details see, e.g., Dynan and Elmendorf (2001). These estimates broadly correspond with the month one, month three and final estimates of quarterly GDP growth produced by the ONS considered in Table 8.1 and Table 8.8 below.

8 Revisions to GDP growth

Table 8.7 Study 3: Revision analysis of US GDP growth (pp. at a quarterly rate)

	Mean	Variance	t-test (p-value)	RMSE
1968q1-1999q4: m3-m1	0.046	0.039	0.001	0.203
1968q1-1999q4: final-m3	0.110	0.259	0.007	0.521
1968q1-1999q4: total revisions	0.157	0.272	0.000	0.545
1993q1-1999q4: m3-m1	0.044	0.030	0.044	0.179
1993q1-1999q4: final-m3	0.040	0.077	0.225	0.280
1993q1-1999q4: total revisions	0.084	0.123	0.052	0.361
1993q1-2003q2: m3-m1	0.030	0.031	0.099	0.178

8.42 Results using this data set are summarised in Table 8.7. It finds a larger bias in the total revision than indicated by Faust et al. Indeed, in contrast to the results of Faust et al. the results in Table 8.7 indicate that the bias in US GDP growth revisions is statistically significant. The size of the bias found in Table 8.7, however, is broadly consistent with the findings of Fixler and Grimm (2002) at the BEA who find, over the period 1983-2000, a mean revision for the total revision of 0.46% at an annual rate (which corresponds to roughly to 0.12% at a quarterly rate). Moreover, there is a 21.61% (16.08%) chance of a revision between the month one and month three at least as big as 0.25% (in absolute value) using the kernel (normal) density.

Table 8.8: UK results. Revisions at each stage of the quarterly constant price GDP growth estimate; adapted from Akritidis (2003a). Percentage Points at a Quarterly Rate

1993q1-1999q4	Mean	Variance	t-test	RMSE
Month three less month one	0.007	0.010	0.372	0.100
First edition of Blue Book (BB1) less month three	0.071	0.028	2.127	0.182
Second edition of Blue Book (BB2) less BB1	0.021	0.044	0.719	0.210
Revisions after BB2	0.089	0.029	3.346	0.192
Total	0.189	0.071	2.778	0.327

8.43 To facilitate comparison with the UK Table 8.8 examines revisions to UK quarterly GDP growth over the period 1993q1-1999q4; the results are adapted from Akritidis (2003a).

8.44 Comparing Tables 8.7 and 8.8, and focussing on the common sample period 1993q1-1999q4, we find that while the bias to total revisions is smaller for the US than the UK, revisions between the month one and month three estimates have a larger mean for the US. However, importantly US revisions, both total and between month one and month three, have a higher RMSE than the UK. For example, for the UK there is a RMSE for total revisions of only 0.327, while in the US there is a RMSE of 0.361. Note that if we consider the results from Akritidis (2003b), see Table 8.1, with the 'final' data set incorporating the effects of the introduction of annual chain linking, we find a higher RMSE for the UK of 0.383. This reminds us of the sensitivity of revisions analysis to the sample period.

8.45 The UK also has a lower RMSE between its month three and month one estimates than the US over the period 1993q1-2003q2; the UK has a RMSE of 0.124, the US a RMSE of 0.178.

8.2 Additional Observations

8.46 The ONS intends to provide more information about past revisions in its first releases of time series [see Annex 3: paragraph 332 for details]. More does need to be done to inform users about likely magnitudes of revisions by presenting summary information on past revisions in press releases. While the regular articles in *Economic Trends* are helpful they do perhaps attract the attention of sophisticated users only, so we welcome this ONS initiative that should reach out to a wider set of users.

8.47 It must be added that the introduction of summary revision information is over-due. The Statistics Commission has called for such information to be provided. Not providing it has perhaps invited some of the adverse press comment recent ONS revisions have received; see Chapter 7.

8.48 The ONS resists the use of bias corrections. Key users also share this view, as do we. Instead it is important to "drill down" through the data to see whether there are significant biases to components of output and expenditure. Once identified the causes of such biases can be removed. Nevertheless, as the user interviews demonstrated, many users remain confused about why bias in the GDP estimates appears to persist, i.e. about why the preliminary estimate appears to be systematically too low. This leaves users in the position at each point in time of having to take a view on whether the causes of biases to the historical time-series have been removed via methodological improvement at the ONS. They have to make a decision about whether the current (not historical) GDP estimates are unbiased or not. The introduction of revision history information measures into the first releases should help users understand that revisions are to be expected, and appreciate why this is not necessarily a bad thing.

8.49 The following two questions were also asked of the ONS seeking (i) to clarify the ONS position on the use of bias corrections and (ii) to explain whether the key provisions of the Statistics Code of Practice Protocol on Revisions protocol were followed ahead of its implementation. Note that the Protocol on Revisions was in force only at the end of the period of interest.

(i) *We understand from you [the ONS] that there used to be bias corrections present in early estimates of GDP. Is this still the case? If not when and why were they dropped?* The ONS explain that bias corrections have only ever been included in the index of manufacturing, not in early estimates of GDP. The motivation for the use of bias corrections was the fact that the survey underlying the manufacturing estimates in the mid 1980s did not include new and growing firms. The bias correction did reduce the mean revision of the series.[22] It appears from the ONS's response in Annex 3: Section G4 that the correction was in place before the cause of the bias, namely the poor survey design, was identified. It was an *ad hoc* measure introduced on a temporary basis to improve the accuracy of, what must have been otherwise unsatisfactory, manufacturing estimates. After the Pickford Review the sample design was improved and the bias adjustment was no longer needed. For details see Annex 3: Section G4.

(ii) *Can you explain whether, and if so how, the revisions post-publication of the Revisions Protocol draft in February 2003 have been handled in a manner consistent with the Protocol?* We believe that the ONS have handled revisions since February 2003 in accordance with the Protocol. Annex 3: Section G8 details the principles of the Protocol and outlines how the ONS have handled revisions in a manner consistent with them.

8.3 Recommendations

8.50 Routine revisions analysis should be undertaken not just of aggregate GDP but also of its components (recommendation R.6). Although the ONS [Akritidis, 2003b] recently undertook a disaggregate examination for GDP(E), i.e. GDP measured on the expenditure side, no such analysis has been undertaken on the output side. Given the evidence for bias in quarterly GDP growth it is important to know where the bias is coming from. While the ONS resists the implementation of bias corrections, perhaps users would react differently if it was clear that the bias was explained by just one or two components, rather than being spread across many. In any case this seems to be a necessary step for the identification of sources of bias.

[22] This does not imply that the range of revisions was also increased, which it was not, just that large positive revisions were offset by large negative revisions.

8.51 The ONS should keep copies of real-time data sets (i.e. of revision triangles) and make these publicly available via a dedicated revisions web-site (recommendation R.7). We understand that the ONS is hoping to adopt this recommendation; we encourage the ONS to do so. Real-time data sets provide a valuable resource not just for the ONS itself, when undertaking revisions analysis and assessing forecasting performance [e.g. see Annex A to this Volume], but for users too. The Bank of England, for example, has recently made available a real-time data set for GDP(E) that has proved popular; see Castle and Ellis (2002). A dedicated revisions web-site could also bring together other relevant information, such as: (i) dates and timetables for methodological changes; (ii) explanatory notes such as those in *Economic Trends* and simpler box-type explanations; (iii) aims and purposes of the series; and (iv) information on the reliability of the first release of the time-series.[23] Although much of this information is available elsewhere putting it altogether could be convenient speeding up users' access of information.

8.52 Following the introduction of information about past revisions into the first releases of time series the ONS should devote resources to assessing user comprehension and reaction (recommendation R.3).

8.53 The ONS should consider grading data again according to their perceived reliability (recommendation R.4). To provide a simple, easy to read, indicator of the reliability of the data the ONS could consider grading data, say *A*-*C*, according to their reliability along the lines of the system of reliability gradings introduced in 1956.[24] An *A* grading indicates (with 90% confidence) that the reported figures are correct subject to a margin of error less than, say, 3% in either direction; *B* indicates an error of 3%-10% in either direction and *C* indicates that the error is greater than 10%. It is, however, essential that when this is done distinction is made between reliability of level estimates and reliability of growth estimates.

8.54 The ONS and the Statistics Commission should think more generally about how best to keep users informed about revisions to economic statistics and their causes (recommendation R.1). While we welcome the National Statistician's article of 13th January 2004 perhaps a similar exercise should be undertaken on a routine basis.[25]

[23] We thank, in particular, Ian McCafferty of the CBI for his suggestions about the role of a dedicated revisions web-site.

[24] For further discussion of the British system, references and an excellent methodological discussion see Morgenstern (1963), p. 273.

[25] See Annex 3: Section G2 for ONS's account of how they currently keep users informed about revisions.

9. Construction Estimates

9.1 The importance of construction

9.1 Construction is responsible for just over 5% of GDP. However, construction is a volatile series and its contribution to the variation of GDP growth is much larger than 5% suggests, indeed perhaps around three times. This estimate is based on examination of the covariance matrix of 100 x the quarter-on-quarter log changes of each component of GDP with the variables themselves scaled by their shares of GDP. To first order, the variance measures are in effect contributions to the variance of the growth rate of GDP; see Mitchell, Smith, Weale, Wright and Salazar (2003) for details.

9.2 The DTI is responsible for producing official information on construction that then feeds into the ONS estimate for GDP. This explains why revisions to the construction estimates can translate into revisions to GDP estimates.

9.3 Revisions to quarterly GDP growth between m1 and m3 over the period 1993q1-2003q2 would have been almost 20% smaller if the revisions to quarterly construction growth between m2 and m3 had not occurred.[26] Looking at Table 9.1 we see the RMSE drop from 0.124 in the presence of revisions to construction to 0.101 without them.[27]

Table 9.1: Revisions to GDP growth estimates between the month three and month one estimates (Percentage Points at Quarterly Rate) with and without the revisions to construction growth between m2 and m3. t-test (p-value) is the p-value for the statistical test of whether the bias is significantly different from zero using a Newey-West type robust estimator

1993q1-2003q2. m3 – m1.	Mean	Variance	t-test (p-value)	RMSE
Actual revision	0.012	0.015	0.235	0.124
With construction revision stripped out	-0.003	0.010	0.422	0.101

[26] The ONS itself does not publish construction estimates at m1 hence our emphasis here on the revision between m2 and m3. In fact, as we see below, looking at the DTI's estimates of construction which are not published but made available to the ONS in confidence, there has been historically little revision between the m1 and m2 construction estimates.

[27] We calculated the effect the revisions to the construction growth estimates had on revisions to quarterly GDP growth as follows. The actual revisions to quarterly GDP growth estimates between the m1 and m3 estimates are compared with the GDP growth revisions there would have been if there had been no revision to the construction growth estimates as published by the ONS between m2 and m3. This is calculated by comparing with the m1 quarterly GDP growth estimate both the actual m3 GDP growth estimate and the m3 GDP growth estimate that would have existed if there had been no revision to construction growth between m2 and m3. This is derived by subtracting from the actual m3 GDP growth estimate the construction revision (construction growth at m3 minus construction growth at m2) scaled by the share of construction in GDP (5.2%). Note revisions to construction growth between m1 and m2 are still included; however, looking at the DTI's construction estimates there is little revision between m1 and m2.

9.4 The DTI supply the ONS with two early releases of construction output prior, in the first week of the third month following the end of the quarter, to publishing their m3 estimate that feeds into the ONS's *Quarterly National Accounts First Release*. The two early estimates, m1 and m2, are published in the first and second months following the end of the quarter. Only at m3 do DTI have full survey information to-hand.[28] The estimates they produce at m1 and m2 for the ONS, and the ONS alone, inevitably involve elements of extrapolation, modelling and judgement, and revisions to them should be expected as more complete survey information is gathered and processed with the passage of time.

9.5 Reflecting this uncertainty the ONS in the press release for the preliminary estimate of GDP provides a qualitative indication only of the direction of construction output growth. It also reminds the user that full survey information is not yet to hand and that revisions should be expected. One month later in the notes to the *Output, Income and Expenditure* press release typically the ONS talks of "early estimates of construction output growth".

9.6 The ONS themselves publish estimates for construction at m2 and m3. These estimates differ from those estimates supplied by the DTI for two reasons. First, the ONS perform an adjustment for Northern Ireland. Secondly, definitional changes are made; this involves a balancing adjustment using weights from the Annual Business Inquiry. There is also an adjustment for evasion two years down the line.

9.7 As indicated in the *Quarterly National Accounts* press release on 30 September 2003, the upward revision between the m1 and m3 estimates of quarterly GDP growth in 2003q2 (from 0.3% in m1 to 0.6% in m3), was due to a large upward revision to the construction estimates supplied to the ONS by the DTI.

9.8 There were also historically large revisions to the construction estimates between the m2 and m3 estimates in the previous quarter, 2003q1, although the impact on GDP growth was attenuated by offsetting changes in the services sector.

9.9 In response to these revisions to construction output growth between the m1 and m3 estimates, in the *Output, Income and Expenditure* press release of 26 November 2003 the ONS indicated that a joint review with the DTI had been set up that would, in its first stage, examine the causes of these revisions and at a second stage consider ways of reducing them.

9.10 Stage 1 has been completed; the results were published on 12 February 2004 on http://www.statistics.gov.uk/about/methodology_by_theme/constructionstats/default.asp. The key finding is that the unforeseen fall in construction output in 2003q1 was responsible for the 2003q2 revision since it caused the DTI "to be more cautious in forecasting a return to positive growth in Q2 2003 than might otherwise have been the case" [see Annex 3: Section B3.3 for a copy of the Stage 1 report and paragraph 57 in particular].

[28] The m3 estimate is based on three sources: (i) the DTI Quarterly Inquiry of Construction Activity; (ii) a quarterly census of Direct Labour Organisations, mostly local authority maintenance teams and (iii) the unrecorded estimate that is an estimate of the output of self-employed construction workers and companies not registered for VAT.

9.2 How the construction estimates are produced

9.11 The following sections review the manner in which the m1, m2 and m3 construction output estimates are produced by the DTI. Again see Annex 3: Section B3.3 for a copy of the DTI/ONS's Stage 1 report.

9.12 Information on how the construction estimates are produced was not placed fully in the public domain until 12 February 2004. Even now all information is not available; in particular the Experian report discussed below is still unpublished. Before the 12 February 2004, background summary information on how construction output is estimated was published only on the 26th November 2003 in the *Output, Income and Expenditure* press release. Full details were not made available.

9.2.1 The Month 1 estimate

9.13 The m1 construction estimates are used by the ONS in calculating their preliminary GDP estimates one month after the end of the quarter. Therefore, the ONS require a timely estimate for construction output growth in the preceding quarter. The DTI supply their m1 estimate within two weeks of the end of the quarter. At this point, DTI typically have no more than 5% of survey returns and so forecasting models are used to meet the ONS's deadline.

9.14 Two forecasting models are used currently by the DTI, the *Orders* model and the *Activity Balances* model.

9.2.1.1 The Orders model

9.15 This model is the most established model used by the DTI at m1. It uses information from the DTI *Monthly Inquiry of Contracts and New Orders* (itself published as a monthly statistic) to model New Work output. New Work output is modelled at a disaggregate level by breaking it down into six sectors: public housing, private housing, infrastructure, public non-housing, private industrial and private commercial. The construction output estimate is then the sum of total new work output and repairs and maintenance (R&M).

9.16 When the ONS require the construction data only two months of data are available from the *Monthly Inquiry of Contracts and New Orders* for the relevant quarter. The third month for each of the 6 new work sectors is forecast. Judgement is used in forming this forecast.

9.17 Since orders figures are not available for R&M, it is extrapolated using the judgement of modellers at the DTI. Again models are not used and the judgement of the modeller is required.

9.18 A series known as "composite new orders" is then derived for each of the six new work sectors as the weighted sum of orders placed over the last few quarters, excluding orders over £150 million.

9.19 These weights are designed to reflect the fact that some contracts take longer than others.

9.20 The composite orders figures for each of the six sectors are then related to new work output via linear regression. This regression is then used to forecast new work output. Two forecasts are computed based on estimating the regression over two sample periods. The first regresses new work output on composite orders from 1983 yielding the so-called Long Term forecast. The second is from 1990 delivering the Short Term forecast.

9.21 Growth relative to the previous period is then computed for each sector. These are then summed to produce the new work estimate.

9.22 The DTI tend to concentrate on the Short Term model, only using the Long Term model as a guide if the Short Term model seems to be doing something unexpected. However, for private housing the Long Term model is the main model used as it seems to fit better.

9.23 The new work and R&M estimates are combined to produce an estimate of total construction output.

9.2.1.2 The Activity Balances model

9.24 This model was developed in response to increasing concern from the ONS and HMT about the performance of the Orders model and the reliability of the construction forecasts. This concern arose from routine analysis of the forecasts.

9.25 Following the recommendation in 2002 of an inter-departmental working group consisting of the ONS, DTI, HMT and Bank of England an “Invitation to Tender” was issued to Experian (since taken over by Business Strategies) and to two other companies to develop an alternative early estimate of construction output growth. Business Strategies were the only one of the three to respond and subsequently were awarded the contract.

9.26 Business Strategies developed the Activity Balances model.

9.27 The activity balances model is based on relating construction output growth to the results of a monthly survey of activity balances within the construction industry.[29] Specifically, the explanatory variable is the balance (percentage of firms reporting rising activity less those reporting falling activity; “ups” minus “downs”) for the first two months of the quarter and the last month of the previous quarter.

[29] The survey conducted by Construction Forecasting and Research (CFR) is part of Experian’s Business Strategies Division and is part of the European Commission’s harmonised series of business surveys undertaken on their behalf by CFR in the UK. The survey is conducted by monthly questionnaire among 800 selected firms. For further details see http://www.construction-forecast.com/links/cif.pdf and http://www.construction-forecast.com/html/publications.htm .

9.28 Dummy variables were also included for 1998 q1 and q2, and 2000 q1 and q2. This equation was found, for example, to deliver forecasts with a lower RMSE than the DTI's m1 (based on the Orders model) estimate over the periods 1997q1-1997q4 and 2001q1-2002q1. This involved estimation up to 1996q4 and 2000q4, respectively, and then producing forecasts for the next four periods.

9.2.1.3 The "combined" m1 estimate

9.29 Since late 2002 the DTI have produced forecasts of construction using both the Orders and Activity Balance models. The estimate given to the ONS by the DTI appears to have been based primarily on the Orders model.

9.2.2 The Month 2 estimate

9.30 Traditionally, this estimate was derived by re-running the Orders model using updated orders information. In particular, the Orders data for the third month in the quarter is now known and can replace the forecast used at m1.

9.31 Recently, the DTI has investigated the benefits of accessing early survey returns from the DTI Quarterly Inquiry of Construction Activity, which amounts to around 81% of total output. Working practices have been changed at the DTI with a view to validating as many returns as possible in time for the m2 estimate. This appears to be delivering promising results and we encourage the DTI to continue these developments.

9.3 Summary of past revisions to construction estimates

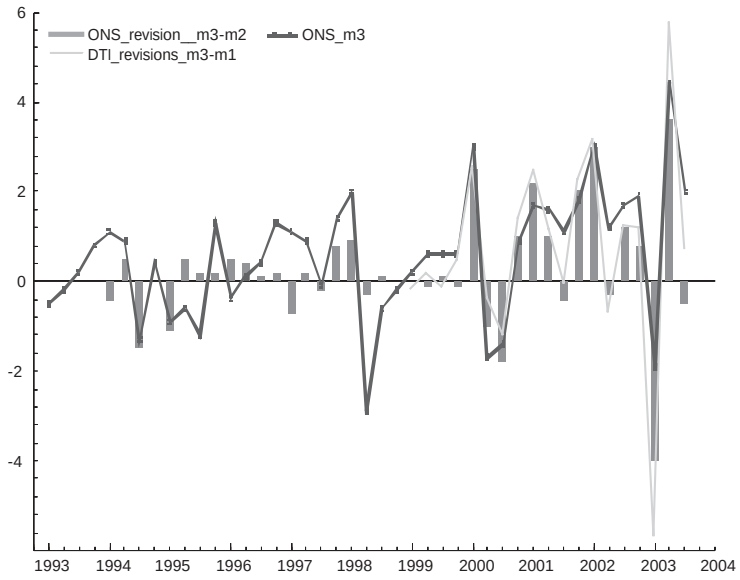
9.32 This section places the revisions to construction output growth in 2003q1 and 2003q2 in their historical context by considering the time-series of revisions to quarterly construction output growth estimates.

9.33 We consider revisions to both the DTI estimates between m1, m2 and m3, and the ONS estimates between m2 and m3. Figure 9.1 provides a visual impression of these revisions over the available sample periods; from 1993q1 for the ONS estimates and from 1999q1 for the DTI estimates. We see that from 1999 revisions to the DTI estimates between m1 and m3 share the same shape, although they are slightly more volatile, to those between the m2 and m3 ONS estimates. This suggests that there is little revision between the DTI's m1 and m2 estimates.

9.34 The large revisions in 2003 q1 and q2 are seen clearly. These revisions are the largest two to-date. Indeed, the volatility of the revisions has increased markedly since 1999/2000.

9.35 Figure 9.1 also shows the positive relation between the size of the revisions and the actual quarterly construction output growth rate published by the ONS at m3. The larger revisions seen recently in 2003 q1 and q2 have been associated with more extreme growth rates.

Figure 9.1: Revisions to construction output (at a quarterly rate in percentage points) between the ONS's m2 and m3 estimates alongside quarterly construction output growth as published by the ONS (m3) and from 1999q2 the revision between the DTI's m1 and m3 estimates.



9.36 Table 9.1 presents some summary statistics about these revisions to construction output growth. The overall accuracy of the estimates, as measured by the RMSE, has increased over time for the ONS estimates. There is also increasing statistical evidence for bias between the ONS's m2 and m3 estimates as evidenced by the decreasing p-values for the test of bias. Looking at the DTI estimates, consistent with Figure 9.1, we see that the bulk of the revision occurs after the m2 estimate; there is little revision between the m1 and m2 estimates. There is also clearer evidence for bias between the m1 and m3 revisions.[30]

Table 9.1: Examination of ONS and DTI revisions to quarterly construction output growth (in percentage points)

	mean	Variance	t-test (p-value)	RMSE
ONS: m3-m2				
1993q1-2003q3	0.223	1.584	0.074	1.264
1999q2-2003q3	0.511	3.365	0.070	1.854
2000q1-2003q3	0.620	4.007	0.058	2.031
DTI: 1999q2-2003q3				
m2-m1	0.122	0.189	0.076	0.440
m3-m2	0.673	4.862	0.047	2.246
m3-m1	0.794	5.334	0.016	2.381

[30] Note that since we have only 18 observations for the DTI revisions, statistical inference may be subject to small sample biases.

9.37 Using the ONS data we find that prior to 2003q1 there was a 5% probability of a revision to quarterly construction output growth at least as large as 2.5% in absolute value using a kernel density estimate (and a 1% chance using the normal density). Using data up to and including 2003q3 this probability rises to 10% (or 5% for the normal density).

9.4 Observations

9.38 The methodology underpinning the Orders model is inadequate in certain ways. For example, the regression of new work output on composite orders appears to be in the levels of the variables. Given the trending nature of new work output there is a danger of "spurious" relationships; e.g. see Hamilton (1994).

9.39 Judgement is used at various stages in constructing forecasts from the Orders model. To the best of our knowledge, no record has been kept of the historical, real-time, performance of these forecasts.

9.40 Information from the Euroconstruct survey could be used more efficiently. Rather than considering the balance statistic alone, the proportion of firms that replied "up" and "down" could be considered separately as explanatory variables in the regression. Information is lost when considering the balance alone. Regressions of the variable of interest on the proportions of firms replying "up" and "down" are used widely; see Pesaran (1987) and Mitchell, Smith and Weale (2002a,b).

9.41 Recent work has found that better signals can be extracted from the underlying panel of firm-level survey responses by considering alternative means of aggregating than the proportion of firms replying "up" and "down"; see Mitchell, Smith and Weale (2002a,b). To be made operational the co-operation of the survey data collectors is required.

9.42 The performance of models should be assessed on the basis of out-of-sample experiments. This involves the recursive application of the models in the out-of-sample period. This allows the parameter estimates to be updated as more information arrives and closely represents the position the modeller would actually have been in when making their forecast. Not allowing the parameters to be updated can affect inference. For example, re-examining the Activity Balances model, the RMSE against the m3 estimate falls, albeit in this case by 0.04% only, over the period 2001q1-2003q3 if we update the parameters each quarter rather than simply use the estimates based on estimation up to 2000q4.

9.43 As well as comparing the forecasting performance of the Activity Balances model against the Orders model other models should be considered also. Simple auto-regressive type models estimated in the growth rate, for example, are used widely as benchmark models in the forecasting literature.

9.44 There are many ways of combining the Orders and Activity Balance forecasts. It is well established that combining competing individual point forecasts of the same event can deliver more accurate forecasts, in the sense of a lower RMSE; see Bates and Granger (1969) and Stock and Watson (2003).[31] A popular combination method involves weighting the two forecasts according to their historical relationship with the outturn, i.e. the m3 estimate; see Granger and Ramanathan (1984).

9.45 A particular concern of this Review is the revision to the preliminary estimate of GDP growth for 2003q2. As discussed above this was largely due to poor performance by the model used to forecast output of the construction industry at m1. Although revisions to construction estimates between m1 and m3 should be expected, as the m1 and m2 estimates are calculated using incomplete, indeed certainly for m1 no, survey information, it is important to test whether the revision could have been anticipated *ex ante*. If it could then the revision was at least partly due to 'avoidable circumstances'.

a. Although users cannot expect the method used to obtain the m1 (and m2) construction estimates to be perfect, they should expect the method used to be the best in the circumstances.

b. Therefore in Annex A we examine whether the model used to forecast construction growth in 2003q1 and q2 was the "best in the circumstances". This same question could have been asked by the ONS/DTI. We examine this question in some detail not just because it is of interest *per se*, but because it illustrates how we think that forecasting models in general should be used and also assessed. We set out the sort of models we think should be used to estimate early estimates of national accounts.

c. We conclude that if the appropriate forecasting approach had been followed, it would have helped identify shortcomings in the primary model used by the DTI to forecast construction output growth at m1. While the DTI and ONS also appear to have been aware of these shortcomings, reflected in the development of an alternative forecasting model in 2002, they have nevertheless continued to base their forecasts at m1 largely on the inadequate model rather than the alternative better one.

[31] The success of combination follows from the fact that individual forecasts may be based on misspecified models, poor estimation or non-stationarities. Indeed recent work, see e.g. Hendry and Clements (2002), has begun to explore further why point forecast combination works through analytical and Monte-Carlo investigation.

d. The history of the construction forecast used in 2003q2 is confused. The DTI model produced an estimate of -0.4% and the DTI passed this to the ONS with the comment that construction output was broadly "unchanged" in the quarter. ONS then used a figure of 0.3% in the production of the m1 GDP estimate. This figure was a combination of the DTI information and ONS's own "judgement" based on qualitative information; see Annex 3: Section B3.3.4 for ONS's own description of this process.

e. We find if appropriate methods had been in place to forecast construction output at m1, construction output growth would have been forecast at 0.9% instead of at -0.4%, which was the DTI's forecast. It is impossible to say with certainty, given the use of "judgement" by both DTI and ONS, what effect this would have had on the construction estimates used by the ONS at m1. In turn it is therefore impossible to say what effect this may have had on the GDP estimate at m1. However, if the m3 figure of 5.3% had replaced 0.9%, instead of the 0.3% figure used by the ONS at m1, then the revision to GDP growth would have been about 0.04% smaller. Subject to rounding this may have led to the publication of an m1 estimate for GDP growth in 2003q2 of 0.4% rather than 0.3%. Therefore, the revision to growth for 2003q2 between m1 and m3, at least in part, was due to 'avoidable circumstances'.

f. We are concerned about the routine use of "judgement" by the ONS when making forecasts without an attempt to assess whether this judgement, in general, improves forecasts. While we know that it did deliver improvements for the specific case of construction growth in 2003q2 the necessary data are not available to assess the ONS's use of judgement more generally. Judgement is used widely in making economic forecasts since the models by construction are designed to meet theoretical criteria at the expense of statistical performance. Statistical forecasts, by contrast, should be carried out using all the information, combining forecasts from a number of different models if necessary. Therefore judgement should normally have nothing further to add. The only exception to this is in the aftermath of rare events, such as the foot and mouth outbreak, when judgemental revision is required since such events are not accommodated by statistical models.

9.5 Development of early survey response forecast at m2

9.46 As indicated above the early survey returns accessed at m2 appear to offer a potentially useful additional source of information. Although back or historical estimates cannot be analysed via recursive application in an out-of-sample period, the early survey estimates accessed at m2 appear in the last few quarters to have

been good. In 2003q1 they predicted a fall of -7.7% and in 2003q2 they predicted rise of 4.5%. Although the 2003q1 estimate was too extreme they did, and perhaps importantly, correctly identify the sign of the growth rate in contrast to the model based estimates.

9.6 Recommendations

9.47 Forecasting models should be assessed systematically along the lines we have suggested in Annex A, notwithstanding the fact that the ONS/DTI have ended up with quite a good forecasting model for construction output growth (recommendation R.8). We imagine that similar exercises could be carried out by a trained statistician in a matter of weeks. We examined the forecasting models for construction in some detail, albeit with modest resources, not just because construction is of interest *per se*, but because it illustrates how we think forecasting models in general should be used and also assessed.

9.48 The remaining recommendations are listed below. Some relate to construction only while others are more general. Adoption of the recommendations would have helped identify shortcomings in the model used to forecast construction output growth.

9.49 Forecasts should always be compared against simple naïve alternatives. Had this been done as a matter of course the DTI/ONS would have discovered problems with the Orders forecasting model earlier (part of recommendation R.8).

9.50 It should be ensured that appropriate resources are directed to the methodological development of techniques used to produce data, such as construction output, that are an input into the national accounts (recommendation R.9). This should help ensure that adequate methods are used.

> (i) The ONS explain how resources are currently allocated in Annex 3: Section G5 when they reply to our question: "*What procedures are in place to identify when and where methodological improvement is required; what mechanisms do you use to allocate resources to meet these needs?*"

9.51 For transparency and to promote use of best-practice tools, the ONS/other data producers should make details of the models they use publicly available (recommendation R.10).

9.52 Additionally, it would be sensible to invite peer review by the ONS Methodological Division or expects from outside the GSS (recommendation R.11). This should promote use of "best-practice" techniques.

9.53 Production of construction estimates at month one and month two should be handled by the ONS rather than the DTI (recommendation R.12). This would facilitate quality control similar to that carried out for other national accounts data. We also draw the Statistics Commission's attention to the fact that some other components of the national accounts, including the month three construction

estimates which unlike the month one and two estimates are not model but survey based, are also compiled outside the ONS.[32] Analysis of the desirability of this is outside the scope of this review.

9.54 The real-time track record of the forecasts, i.e. both the final forecast and its components, should be kept and analysed routinely (relates to recommendation R.7). This could be done semi automatically via a suitable Visual Basic macro. We feel that it would be useful to keep such a record to facilitate routine and regular *ex post* analysis of the forecasting errors. This should help highlight potential shortcoming in the process used to calculate the forecast.

9.55 At least alongside the judgement exercised by the modeller, it would be useful to forecast regularly using a range of (automated) models, such as ARIMA models and so-called FLASH models of the type regularly used to forecast quarterly GDP growth ahead of publication of official data by using two months of hard data but forecasting the final month in the quarter. Large discrepancies between forecasts may help identify periods of concern *ex ante* and warn the modeller that extra care should be taken.

9.56 In periods of concern about data the ONS/other data producers should ensure they indicate their lack of confidence in their estimates to the public (recommendation R.13). The ONS needs to decide how best to communicate this to the public.

9.57 Various means of using information from qualitative surveys should be explored (recommendation R.14). The DTI has produced a model which uses qualitative information, namely the Activity Balances model. This gives much better results than the model traditionally employed (the Orders model). But the study by Business Strategies that introduced the Activity Balances model did not conduct a review of different ways of exploiting the available data. Alternative sources of data should be considered also. Ian McCafferty (CBI) in his submission to this Review notes that: "one wonders if methods [for estimating construction] could be improved by using Builder Merchant Federation data and other surveys of construction firms and heavyweight building material producers and by cross-checking with employment data and industry sources".

9.58 The ONS should consider whether the quality of early GDP estimates would be helpfully improved if the UK followed Germany in producing a monthly index of construction output (recommendation R.15). This would put estimates of construction output on a basis similar to industrial production and would be likely to deliver much better m1 estimates of quarterly GDP. The DTI/ONS would no longer have to rely on model-based estimates for construction at m1. Given the volatility of construction this would be likely to deliver better estimates than those

[32] The workings of the present system, and how data supplied to the ONS from outside sources are quality controlled, are explained in Annex 3: Section G6 in response to our question: "*We understand that some inputs into early estimates of GDP are provided by other Government Departments. What procedures does ONS have for quality control of this work?*" See also Annex 3: Section G5.

obtained at m1 not just from the DTI's primary model but even the 'best' forecasting model. For example, even if appropriate forecasting methods had been in place at m1 to forecast construction output for 2003q2 there would have been a revision against the outturn (at m3) of 4.4% [see Annex A: paragraph A.23]. Against the production of a monthly index must be offset the cost to the Government and the extra burden imposed on business. We welcome the fact that, as we understand it, the second stage of the DTI/ONS joint review into the early estimates of construction output for GDP currently under way [see paragraph 9.9], amongst other options, is considering the production of a monthly index of construction output. Indeed, this was one of the recommendations of the first stage of the joint review that is published in Annex 3: Section B3.3.[33]

[33] See also: http://www.statistics.gov.uk/about/Methodology_by_theme/constructionstats/default.asp.

10. Trade estimates: VAT fraud

10.1 On 9th July 2003 the ONS published revised estimates for imports of goods to the UK from the rest of the EU that accounted for the effect of VAT missing trader intra-community fraud. As a result the UK's trade deficit was much larger than previously thought. Specifically, there were upward revisions to imports of £1.7 billion in 1999, £2.8 billion in 2000, £7.1 billion in 2001 and £11.1 billion in 2002.

10.2 Clearly, the arrival of unforeseen additional information, namely awareness of criminal activity, explains this revision.

10.3 The revisions to the import estimates affected national account estimates, such as those for GDP.

10.4 The effect on the national accounts was not incorporated into the Quarterly National Accounts until 30 September 2003.

10.5 This left users for over two months in the position of knowing the national accounts data were 'wrong' but not knowing the 'true' values. This is a product of the ONS's release calendar. The Bank of England produced interim estimates. Other users understandably could not do this.

10.6 The ONS advertised the revisions via press releases on 9th July 2003, including the publication of a letter from the National Statistician to the Chairman of the Statistics Commission [see Annex 3: Section F] and the publication of a paper in *Economic Trends*; see Ruffles, Tily, Caplan and Tudor (2003).

10.7 The letter from the National Statistician to the Chairman of the Statistics Commission (9 July) indicates that the ONS did provide early indications of the potential unreliability of the import data to selected users (such as the Bank of England and HMT), even in the absence of a quantitative indication of the effect on GDP. The question then arises why was this same information not shared with other users, the public, for six months – why did the public have to wait until July before they were told that the import data were unreliable?

10.8 The ONS have answered this question explicitly; see Annex 3: Section G7 for their written response. The ONS view, consistent with the National Statistics Code of Practice Protocol on Release Practices, is that the revisions could not be announced publicly until the effect of the fraud on the trade estimates had been calculated. Otherwise, given the sensitivity of the trade estimates, the release of the revisions

would of itself have created unnecessary uncertainty. To minimise uncertainty the ONS also wished to delay release of the revisions until they could provide a quantitative indication of their effect on GDP as well. This was not possible until late June 2003.

10.9 The ONS's explanation for the delay in publication appears to us to be perfectly consistent with their obligations.

10.10 Nevertheless, the more general question does arise of whether the ONS should wait until it has quantified the revisions and considered their likely effects before releasing them publicly or should it alert users straightaway about the potential unreliability of the data? Why should selected users only be trusted with the early information?

10.11 The issue is one of whether the ONS might be expected to publish on its website a "running commentary" on those points that are of sufficient importance to justify letters to the Bank/HMT or whether it is better for the ONS to wait until the evidence has accumulated and then publish both the revisions and the reasons for them.

10.12 The Statistics Commission needs to take a view on this, in the light of the above comments from the ONS on the Code of Practice; see recommendation R.5.

10.13 We contacted for a second time those users from outside the public sector who had listed VAT fraud as one of their "key revisions". They were asked what they thought of the fact that other users, those from the public sector, were informed privately about the revision six months in advance of them.

10.14 As expected, although appreciating that the financial markets may be unsettled, users felt that the public should be notified too. They felt that if there was sufficient cause to believe the numbers were misleading to raise the issue with these selected users from the Bank/HMT then perhaps the ONS should not let the public use the existing numbers without some indication of their unreliability. Certainly waiting six months before indicating their unreliability seemed excessive to them. They suggested that a more transparent approach could have been adopted by the ONS. This could have involved telling all users, not just the selected ones, that due to criminal activity (naturally something beyond their control) the import estimates were likely to be unreliable and that revised estimates would be published as soon as possible.

10.15 Some users also expressed concern about the speed at which the ONS detected the unreliability of the import estimates. The following question arose: Could the ONS have anticipated problems with the import data earlier? Could information on the discrepancy with Eurostat trade data, for example, have been exploited in this respect?

10 Trade estimates: VAT fraud

10.16 The ONS were asked these questions; their responses are in Annex 3: Section G7. They explain that HM Customs and Excise became concerned about VAT fraud during 1999. The first estimate of the tax loss caused by the fraud was published in November 2001. It then took from November 2001 until the end of June 2003 to establish the effect of the fraud on the trade estimates and provide some indication of its effect on GDP. ONS feel it would have been dangerous to use directly the discrepancies with the Eurostat data since all European countries had the conditions for the fraud to work.

10.17 While we cannot comment on the efficiency of the process the ONS carried out from November 2001 until the end of June 2003, it does seem noteworthy that the ONS were aware by spring 2002 that the fraud would have some effect on the trade estimates. But this information, according to the National Statistician's letter to the Chairman of the Statistics Commission, was not shared with even the Bank/HMT until December 2002. The ONS did not make the public aware of the problem until it had calculated revised data in July 2003.

11. Regional GVA

11.1 This section examines further the recent revision to the regional GVA estimates discussed above in paragraph 6.10(i). As indicated above the focus of this Review is on national data. However, since regional output estimates have attracted attention recently due to their use by the EU in allocating regional support, we felt that a more general question should be asked of the ONS.

11.2 The interesting issue we felt was not so much that the estimates were revised but whether users in drawing up their policy thought about this possibility and accordingly designed their policy? Did the users contact the ONS in advance of drawing up their policy to seek advice on the use of their data?

11.3 Therefore, we asked the ONS the following question: "*Do users of your data contact you in advance of drawing up their, say, policy to seek advice about the size and nature of likely revisions? In particular we need to know whether anyone in ONS was consulted before it was agreed to make EU regional support available to regions whose GDP per capita was less than 75% of the EU average.*" The ONS response is in Annex 3: Section G3.

11.4 The ONS state (Annex 3: paragraph 216) that: "users do not seek advance advice on normal revisions, i.e. those due to later information. This is neither practical nor consistent with the National Statistics Code of Practice. On the other hand there is an active dialogue on revisions expected as a result of planned methodological changes".

11.5 Specifically in the context of EU regional aid, the ONS note that the estimates used are Eurostat estimates based on them standardising (using purchasing power parities) each country's GDP estimates.

11.6 Notwithstanding our recommendations above (see recommendations R.1-R.4) suggesting that the ONS should consider alternative means of 'educating' users about revision, we note that policy-makers (what we might call 'sophisticated' users) do understand that revisions are inevitable; this is consistent with our user interviews; see Section 5.

11.7 Given that the ONS is adequately educating them about likely revisions, the responsibility for use of the data clearly then lies with the user. It is down to the user when setting policy to factor in the probability of a revision. Certainly, as indicated in Section 5, policy-makers in the UK do this. Of course, this particular revision to the

11 Regional GVA

regional GVA data was not a normal revision; it was explained by an avoidable circumstance. Policy-makers need to be aware than these types of revision happen also. It is difficult to give probabilities to rare events such as revisions due to processing errors. But unless such revisions are very different in character from other revisions, policy-makers can take account of their distribution when designing policies. An essential part of the current reliance on “evidence-based” policy is a proper assessment of the quality of the evidence. For example, where data are known to be imprecise it is plainly not sensible to adopt policies based on thresholds.

12. Industrial production and the index of manufacturing

12.1 An analysis of revisions to the Index of (Industrial) Production [IoP] is important for two reasons. First, they are an important variable *per se* representing over 25% of GDP(O), used widely and indeed mentioned in the list of key revisions. Secondly, as an input into GDP(0). If there is no bias in IoP then biases in GDP must be due to the forecast of the third month or other components of GDP.

12.2 Users are encouraged to use the 3-month on previous 3 months growth rate to avoid excessive volatility. We follow this advice.

12.3 We examine revisions between the first estimate of the IoP and two final vintages. This lets us isolate causes of the revisions. We also look at revisions to the Index of Manufacturing (IoM). The IoM is the largest component of the IoP itself amounting to 22% of GDP(O).

12.4 This first estimate of the IoP is used in the m2 estimate of GDP growth. If there is no bias in the IoP we can deduce that any biases in GDP must be due to the forecast of IoP for the third month of the quarter or other components of GDP. Indeed the biases in GDP occur after the m2 release so should not be expected to be associated with the unpublished m1 IoP component of GDP growth.

12.5 The first IoP estimate is published just over a month after the end of the reference period. At this stage about 80% of returns (to the Monthly Production Inquiry [MPI]) have been processed.

12.6 As the ONS make clear in their press releases that the first estimate is liable to revision as it is based on incomplete information. The final estimates are based on over a 90% response to the MPI. Other factors causing revisions arise from seasonal adjustment, changes at the Blue Book, and methodological changes.

12.7 Recently, in Autumn 2003, important methodological changes have been made to the IoP. These involved a move to annual chain-linking plus a series of other improvements; see Fletcher and Williams (2002, 2003) for details.

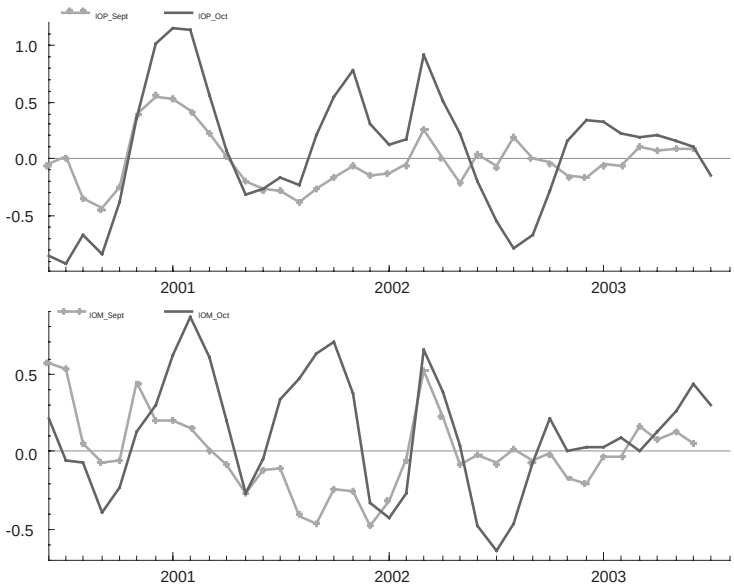
12.8 These changes were clearly indicated in advance by the ONS. For example, the ONS on 12th September 2003 released a historic dataset using the new methodology to help users understand the effect of the changes, and articles were published in *Economic Trends*.

12 Industrial production and the index of manufacturing

12.9 A long run data base of revisions to IoP is not available for analysis. Using the available data base we find that the methodological changes introduced in October 2003 had a large effect on revisions to the IoP and IoM. This effect is easily isolated by comparing the revisions to the IoP and IoM between the first estimate and the estimates computed in September 2003 (prior to the methodological changes) and the estimates of October 2003 (computed after the changes) for the period June 2000-July 2003.

12.10 Figure 12.1 shows the impact the methodological change had. Also see the effect of the Jubilee in June 2002, and the ensuing improvements in the seasonal adjustment procedures. Note that this effect is attenuated by our focus on the 3-month on previous 3 months growth rate.

Figure 12.1: Revisions to the IoP and IoM. Comparing the estimates in September with those in October after the introduction of a series of methodological changes



12.11 Table 12.1 confirms this view. The mean, variance and RMSE of the revisions to the three-month on three-month growth rates for IoP and IoM have all increased when the first estimates are compared to the estimates using the new methodology.

12.12 The increase in the average revision associated with the introduction *inter alia* of annual chain-linking, was consistent with the ONS's prediction. Fletcher and Williams (2003) predicted this. When rising volumes are associated with rising prices or vice-versa annual chain-linking may have a positive effect on growth. This appears to have happened.

12.13 Despite the increase in the average revision, there is, however, no statistical evidence for bias even when considering the new (October) estimates.

Table 12.1: Revisions to IoP and IoM growth estimates between the first estimate and the September 2003 and October 2003 estimates for the period June 2000-July 2003. (Percentage Points; 3 months on previous 3 months growth rate). t-test (p-value) is the p-value for the statistical test of whether the bias is significantly different from zero using a Newey-West type robust estimator

	mean	variance	t-test (p-value)	RMSE
IOP 3 month Sept	-0.022	0.057	0.420	0.236
IOP 3 month Oct	0.065	0.304	0.411	0.548
IOM 3 month Sept	-0.006	0.065	0.468	0.251
IOM 3 month Oct	0.114	0.137	0.223	0.382

References

Akritidis, L (2003a), “Revisions to quarterly GDP growth”, *Economic Trends*, 594, 94-101.

Akritidis, L (2003b), “Revisions to quarterly GDP growth and expenditure components”, *Economic Trends*, No 601

http://www.statistics.gov.uk/articles/economic_trends/ETDecAkritidis.pdf

Bates, J.M. and C.W.J. Granger, (1969), “The combination of forecasts”, *Operational Research Quarterly*, 20, 451-468.

Box, G.E.P. and G.M. Jenkins (1976), *Time Series Analysis: Forecasting and Control*, revised edition, San Francisco: Holden Day.

Cahill, M.B. (2003), “Teaching Chain-Weight Real GDP Measures,” *Journal of Economic Education*, 34(3), 224-234.
(http://www.indiana.edu/%7Eeconed/pdffiles/summer03/cahill.pdf)

Camba-Mendez, G., G. Kapetanios, R.J. Smith and M.R.Weale (2001), “An Automatic Leading Indicator of Economic Activity: Forecasting GDP Growth for European Countries”, *Econometrics Journal*, 4 (1), S56-S90.

Castle, J. and C. Ellis (2002), “Building a real-time database for GDP(E)”, *Bank of England Quarterly Bulletin*, February, 12-19.

Clements, M.P. and D.F. Hendry (1999), *Forecasting Non-Stationary Economic Time Series*, The MIT Press, London.

Cook, L. (2004), “Revisions to economic statistics: their role in measuring economic progress”, *Economic Trends*, 603, 36-43.

Dicks, G. and R. Walker (2003), “What do you think of that Blue Book then?”, The Royal Bank of Scotland, Financial Markets UK economy.

Diebold, F. X. and R. S. Mariano (1995), “Comparing Predictive Accuracy”, *Journal of Business and Economic Statistics*, 13, 253-263.

Doornik, J.A. and M. Ooms (2001), “A package for estimating, forecasting and simulating Arfirma models: Arfirma package 1.01 for Ox”, Working Paper, Nuffield College, Oxford.

Dynan, K.E. and D.W. Elmendorf (2001), “Do provisional estimates of output miss economic turning points?”, Working Paper, Federal Reserve Board, Washington D.C..

Eggington, D, A. Pick and S. Vahey (2002), '"Keep it real!": a real-time UK macro data set', *Economics Letters*, 77, 15-20.

Faust, J., J. Rogers, J. and J. Wright, J. (2000), "News and Noise in G-7 GDP Announcements", *Journal of Money, Credit, and Banking*, forthcoming.

Fixler, D.J. and B.T. Grimm (2002), "Reliability of GDP and related NIPA estimates", *Survey of Current Business*.

Fletcher, D. and M. Williams (2002), "Index of Production redevelopment", *Economic Trends*, October 2002.

http://www.statistics.gov.uk/articles/economic_trends/ET587_fletcher.pdf

Fletcher, D. and M. Williams (2003), "The impact of redeveloping the Index of Production", *Economic Trends*, vol No. 595.

http://www.statistics.gov.uk/articles/economic_trends/FletcherWilliamsJune03.pdf

Garratt, A. and S. Vahey (2003), "UK real-time macro data characteristics", Working paper, University of Leicester and Cambridge.

Granger, C.W.J. and R. Ramanathan (1984), "Improved methods of combining forecasts", *Journal of Forecasting*, 3, 197-204.

Hamilton, J.D. (1994), *Time Series Analysis*, Princeton University Press.

Harvey, A.C. (1989), *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press.

Harvey, D. I. and S. J. Leybourne and P. Newbold (1997), "Testing the Equality of Prediction Mean Square Errors", *International Journal of Forecasting*, 13, 273—281.

Hendry, D.F. and Clements, M.P., (2002), "Pooling of forecasts", *Econometrics Journal*, 5, 1-26.

Mitchell, J. (2003), "Should we surprised by the unreliability of real-time output gap estimates? Density estimates for the Eurozone", NIESR Discussion Paper 225; [http://www.niesr.ac.uk/pubs/discuss.htm]

Mitchell, J., R.J. Smith and M.R. Weale (2002a), "Quantification of qualitative firm-level survey data", *Economic Journal*, 112, 117-135.

Mitchell, J., R.J. Smith and M.R. Weale (2002b), "Aggregate versus disaggregate survey-based indicators of economic activity", Presented at *Econometric Society European Meeting*, Lausanne University (August 2001). NIESR Discussion Paper 194; [http://www.niesr.ac.uk/pubs/discuss.htm]

Mitchell, J., R.J. Smith, M.R. Weale, S. Wright and E.L. Salazar (2003), "An Indicator of Monthly GDP and an Early Estimate of Quarterly GDP Growth", National Institute of Economic and Social Research Discussion Paper 127 (Revised).

References

Morgenstern, O. (1963), *On the Accuracy of Economic Observations*, Princeton University Press, Princeton: New Jersey.

Newey, W.K. and K.D. West (1987), "A simple, positive semi-definite heteroskedasticity and autocorrelation consistent covariance matrix", *Econometrica*, 55, 703-708.

Öller, L. and Hansson, K. (2002) "Revisions of Swedish National Accounts 1980-1998 and an International Comparison", Statistics Sweden http://www.scb.se/Grupp/ekonomi/_Dokument/ReportNArev.pdf

Pesaran, M.H. (1987), *The limits to rational expectations*, Basil Blackwell.

Reed, G. (2000), "How the preliminary estimate of GDP is produced", *Economic Trends*, 556, 53-62.

Richardson, C. (2002), "Revisions to GDP: a time profile", *Economic Trends*, No 584 http://www.statistics.gov.uk/articles/economic_trends/ET_July02_Richardson.pdf

Richardson, C. (2003), *"*A Time Series Approach to Revisions*",* *Economic Trends,* No 601 http://www.statistics.gov.uk/articles/economic_trends/ETDecRichardson.pdf

Ruffles, D., G. Tily, D. Caplan and S. Tudor (2003), "VAT missing trader intra-Community fraud: the effect on Balance of Payments statistics and UK National Accounts", *Economic Trends*, 597, 58-70.

Samuelson, P.A. and S. Swamy (1974), "Invariant economic index numbers and canonical duality: survey and synthesis", *American Economic Review*, 64, 566-593.

Schwarz, G. (1978), "Estimating the Dimension of a Model", *Annals of Statistics*, 6, 461-464.

Smith, D.B. (2003), "Target practice?", *the parliamentary monitor*, November 2003.

Statistics Commission (2003a), "Forecasting in the national accounts at the Office for National Statistics", *Statistics Commission Report No. 12*, http://www.statscom.org.uk/resources/reports_docs/forecasting%20final.pdf

Statistics Commission (2003b), "The 2001 Census in Westminster: Interim Report". http://www.statscom.org.uk/resources/reports_docs/census2001.pdf

Stock, J.H. and M.W. Watson, (2003), "Combination Forecasts of Output Growth in a Seven-Country Data Set", Department of Economics, Harvard University.

Tay, A.S. and K.F. Wallis (2000), "Density forecasting: a survey", *Journal of Forecasting*, 19, 235-254.

Tuke, A. (2002), "Analysing the effects of annual chain-linking on the output measure of GDP", *Economic Trends*, No 581.

http://www.statistics.gov.uk/articles/economic_trends/analysing_effects_annual_chain-linking_on_output_measure_GDP.pdf

Tuke, A. and V. Aldin, (2004), "Reviewing the methods and approaches of the UK National Accounts", January 2004, *Economic Trends*, 47-57.

Tuke, A. and J. Beadle (2003), "The effect of annual chain-linking on Blue Book 2002 annual growth estimates", *Economic Trends*, 593, 29-40.

http://www.statistics.gov.uk/articles/economic_trends/ETApr03Tuke.pdf

Tuke, A and G. Reed (2001), "The effects of annual chain-linking on the output measure of GDP", *Economic Trends*, No 575.

http://www.statistics.gov.uk/articles/economic_trends/chainlinking.pdf

Turnbull, A. and M. King (1999), "Review of the Revisions to the Average Earnings Index". Report submitted to the Chancellor of the Exchequer.

Weale, M.R. (2003), "Box 1. Implications of the classification of Network Rail to the private sector", *National Institute Economic Review*, 181, 53.

Youll, R. (2002), "Quality of the estimates of earnings growth from the Average Earnings Index", *Labour Market trends*, Technical Report, 207-213.

Annex A: Was an inadequate model used to forecast construction output growth?

A.1 We use simulated out-of-sample experiments, designed to mimic the real-time application of the models, to examine the performance of the DTI's m1 and m2 estimates against plausible and widely used alternative models, including the Activity Balances model from 1999q1.[34] This is possible despite the fact that the Activity Balances model has only been in existence from 2002. The application involves recursive application of the models in an out-of-sample period. This strategy is used widely in the forecasting and econometric communities as a means of evaluating alternative forecasting models.

The following sub-section outlines the theoretical rationale for our experiments.

Annex A.1 Evaluating alternative forecasting models on the basis of their out-of-sample performance

A.2 It is always possible to explain past growth, be it of construction or GDP as a whole, reasonably well, using a relatively small number of carefully chosen variables with carefully chosen lags. But there is no reason to expect that such equations will necessarily be good forecasting tools. With a reasonably large number of potential variables, a relatively small number of observations, the ever-present risk of structural change and without the constraints of any structural framework, the aim of the researcher must be to describe and evaluate a forecasting strategy rather than simply to find an equation which happens to fit the data.

A.3 Nevertheless some practitioners in their search to find the best forecasting equation have justified their chosen equation using arguments based on their equation offering a good in-sample fit, as evidenced by a high R^2 or low standard error of regression. However, little or no discussion is given as to why the preferred equation was chosen instead of an alternative.

[34] The DTI's forecasts from 1999q1-2003q3 were kindly made available to us. Before 2002, the m1 estimate is based on the Orders model. From late 2002 the DTI have also forecasted using the Activity Balances model; however, the m1 estimate remains largely determined by the Orders model.

A.4 It is crucial to finding the best forecasting equation that the practitioner also looks at out-of-sample performance. This is central to evaluating a forecasting strategy, rather than simply finding an equation which happens to fit the data well. It should be added that one does run into a philosophical question; if out-of-sample performance is used as a basis for model selection it becomes a criterion and it is not clear whether it can still be a test.

A.5 We can simulate the out-of-sample performance of alternative models by splitting the sample period into estimation and out-of-sample periods. The models can then be estimated recursively throughout the out-of-sample period and a sequence of one-step ahead forecasts obtained.

A.6 Forecasts are evaluated, as is traditional, using RMSE.[35] The smaller the RMSE the better the forecast. When the RMSE is less than the standard deviation of the growth rate the model helps explain the variation of growth about its mean. But to test a methodology, rather than just a specific equation, we should test the performance of our forecasts against those obtained by other (non-structural) forecasting methods. Given our concern with short-term forecasts we need only consider single-equation models.[36]

Annex A.2 Alternative forecasting models

A.7 As alternatives to the DTI's m1 and m2 estimates we consider the following models for (aggregate) construction:

1. The Activity Balances model. In contrast to the evaluation in Experian's report, the models are evaluated based on recursive estimation through the out-of-sample period. This allows the parameter values to be updated. Three variants of the model are considered. The first includes no dummy variables, the second the dummy variables for 1998 only and the third both sets of dummy variables, the dummy variables for 2000 being allowed to enter from 2001.

2. ARIMA models.[37] We consider three types of ARIMA model estimated in the level of construction output, all of which include an intercept:

    A. ARIMA(0,2,2). Forecasts produced from this model relate to the forecasts produced from Holt's method; see Harvey (1989) and Box and Jenkins (1976, pp. 105-8, 146-9).

[35] Recently, forecasts have begun to be evaluated more widely not just on the basis of their point forecasts but according to the degree of predicted uncertainty. Measures of uncertainty surrounding a "central tendency" (the point forecast) can enhance the usefulness of the forecast. So-called "density" forecasts are being used increasingly in economics and finance since they provide commentators with a full impression of the uncertainty associated with the forecasts; see Tay and Wallis (2000) for a review and Mitchell (2003) for an application measuring the 'business cycle' in real-time.

[36] When forecasting horizon t+h, h≥1, at time t the indicators can no longer be assumed exogenous as they need to be forecast. In such a case multivariate models, such as Vector Auto-Regressive models, constructed from the set of indicator variables provide a useful forecasting tool; see Camba-Mendez, Kapetanios, Smith and Weale (2001).

[37] Estimation and forecasting of the ARIMA models is carried out using the *Ox* programming language, and the Arfirma package; see Doornik and Ooms (2001).

B. ARIMA(0,1,1). Forecasts produced from this model relate to those produced from exponentially weighted moving average models; again see Harvey (1989) and Box and Jenkins (1976, pp. 105-8, 146-9).

C. ARIMA(BIC,1,BIC). Lag selection is determined by the Bayesian information criterion (BIC); see Schwarz (1978).[38] At each recursive sample the BIC is used to determine the number of lags in the ARMA model estimated in the first differences of construction. This approach allows for the fact that it is possible that the best fitting model at the start of the period is not the best fitting model in, say, the middle of the period. To correct for this effect, e.g. following Camba-Mendez, Kapetanios, Smith and Weale (2001), we recursively find the best fitting model at each period in the forecasting evaluation period using the BIC.

D. The delta-squared model of Hendry (DAR) ($d^2y_t=0$, where d is the first-difference operator). The delta-squared model implies $dy_t=dy_{t-1}+e_t$, where e_t is a zero mean error process. Consequently $E(dy_t)=dy_{t-1}$. It is important to consider such benchmark models since it is well known that parsimonious non-structural models, such as models in first differences, often deliver better short-term forecasts than structural models; see Clements and Hendry (1999). This is because these models can guard against unforeseen events such as structural breaks.

A.8 We also considered forecasting (aggregate) construction growth via aggregation of disaggregate forecasts. Each sector (the six new work sectors and four R&M sectors: public and private housing and public and private other) is modelled separately using an ARIMA(BIC,1,BIC) model. Sequences of one-step ahead forecasts are then obtained based on recursive estimation and forecasting of the models in the out-of-sample period. The disaggregate forecasts are then aggregated by weighting them according to the relative size of the sector in levels.

Annex A.3 Data vintage: forecasting the m3 estimate versus forecasting the "final" estimate

A.9 There are minor difference between the m3 estimates as published at the time (in "real-time") and the revised estimates of construction output growth using the "final" data vintage as available now, in February 2004.[39] This reflects revisions that have occurred post m3.

[38] The BIC is a criterion that seeks to maximize the fit of the model, as measured by say the sum of squared residuals, subject to a penalty for the inclusion of extra explanatory variables. It is used widely to select among nested econometric models.

[39] Values are never truly final because data revisions are a continuous process.

A.10 Our data set consists of real-time estimates of m3 construction output growth from 1999q1 to 2003q3. Comparing these growth estimates with the revised estimates, based on the 2003q3 vintage, we see a correlation coefficient of 0.98. There are some differences in scale if not in shape at the beginning of the sample period. As expected, the two series are increasingly similar the closer we get to the end of the sample since revisions post m3 have yet to be implemented.

A.11 We therefore ignore these slight discrepancies. It should be acknowledged, however, that increasingly within the academic community the effects of data revisions on forecasting are being considered; e.g. see Eggington, Pick and Vahey (2002).

Annex A.4 Forecast performance: evaluating the alternative models

A.12 Three out-of-sample periods are considered. The first is from 1999q1-2002q4 and is chosen to mimic the situation the DTI were in back in 2002q4, in advance of making at both m1 and m2, what later proved to be, their poor forecasts for 2003q1. Such an experiment could have been performed prior to the DTI making their forecast for 2003q1; it could have been used to help indicate the reliability of the m1 estimator. The second sample period is 1999q1-2003q1; this is designed to represent the situation after, what proved to be the "surprising", negative growth rate for 2003q1 but prior to making the forecast for 2003q2. The third period is the full out-of-sample period, 1999q1-2003q3.

Annex A.4.1 Evaluation over the period 1999q1-2002q4

A.13 Table A1 shows the RMSE of the recursively computed forecasts for the various forecasting models against both the final and real-time (m3) outturns. Against the 'final' estimates (not of course available in 2002q4), the DTI's m1 estimator does deliver a lower RMSE than the ARIMA models, the disaggregate forecast and the benchmark DAR model, but the Activity Balance models perform better. However, against the m3 outturn the m1 estimator does not perform as well as either the ARMA models estimated in the growth rate of construction or the Activity Balance models. This certainly indicates that the DTI should have been alerted to the potential inadequacy of the Orders model.

A.14 This impression is confirmed when considering the results from some forecast combination and encompassing tests. These tests examine whether the forecasts obtained using the alternative models add information [see Granger and Ramanathan, 1984] vis à vis the DTI's m1 estimates.

A.15 To this end, we ran Ordinary Least Squares regressions of the form:

$$x_t = c + W_1 x^{m1}_t + W_2 x^{a}_t$$

where x_t is actual construction output growth (i.e. the final or real-time m3 estimates), c is an intercept, x^{m1}_t is the DTI's m1 (or m2) growth estimate and x^{a}_t is growth

implied by the alternative model. The estimated coefficients in these regressions, W_1 and W_2, show how the information from the m1 (or m2) estimate and the alternative model should be combined to provide the "best" (in a mean squared error sense) estimates of construction output growth: W_1 and W_2 tell us the weights that should be attached to the m1 and alternative models. We also test the following hypotheses: Ha: W_1=1,W_2=0 and Hb: W_1=0,W_2=1. The first hypothesis, Ha, tests whether the m1 estimates encompasses the alternative model, and the second hypothesis, Hb, tests if the alternative model encompasses the m1 estimates.

A.16 The results of the forecast encompassing tests are shown in Table A2.[40] The results over the sample period 1999q1-2002q4 suggest that the ARIMA and Activity Balance forecasts do add information relative to that contained in the m1 estimator. However, neither forecast encompasses the other. The "best" forecast is a linear combination, with weights as given in Table A2, of the competing forecasts.

A.17 The fact that the alternative forecasts add information relative to the DTI's m1 estimator implies that the model used by the DTI to supply its m1 estimates is inadequate.

A.18 None of the (individual) models predicted accurately the outturn for 2003q1 growth, namely growth of -2.6%. But, none of the alternative models were as off-target as the DTI's m1 estimate of 3.1%. The other models all predicted growth in the range 0.6%-1%. For the ARIMA models, in particular, this is unsurprising as it is well known that ARIMA models tend to produce "quite smooth" forecasts. Indeed all of the alternative models, except ARIMA(0,2,2), predicted lower growth in 2003q1 than 2002q4. This stands in contrast to the DTI's m1 estimate that rose from the forecast of 0.9% for 2002q4 to a forecast of 3.1% for 2003q1. Moreover, the combined forecasts, that weight the m1 forecast and that from the alternative model according to the weights in Table A2, are also less than 3.1%.

A.19 This all suggests that the DTI's m1 (and m2) estimate for 2003q1 was out-of-line with what other models were suggesting. Given the fact that this out-of-sample analysis could have been undertaken by the DTI itself in advance of supplying its estimate for 2003q1 to the ONS, perhaps they should have down-weighted the forecast. This is not to say that the outturn for 2003q1 was not unexpected; it could not have been predicted successfully by the other models. But there was reason to doubt the reliability of the DTI's m1 estimator.

A.20 The DTI and ONS, apparently, were aware of these shortcomings in the m1 estimates produced by the Orders model, as evidenced by Experian being commissioned to carry out their review that led to the development of the Activity Balances model. But the forecast the DTI gave to the ONS for 2003q1, that appears to be due largely, if not entirely, to the Orders model, as indicated above, was out-of-line with forecasts from both the Activity Balances model and the ARIMA models.

[40] Results are presented only using the final series for construction growth; results are qualitatively similar using the real-time (m3) estimates.

Annex A.4.2 Evaluation over the period 1999q1-2003q1

A.21 Let us now consider the situation the forecasters were in when predicting growth for 2003q2. They would have been able to evaluate forecasts using data up to 2003q1. We see that undertaking such an exercise, i.e. conditional on the outturn for 2003q1, the DTI's m1 and m2 estimators are seen clearly to be unsatisfactory; see Table A1. Not just are they outperformed by the ARIMA, disaggregate and Activity Balances models but, against the m3 outturn, by the benchmark DAR model. This is also reflected in an increasing weight being given to the alternative model relative to the m1 forecast in the forecast combination regressions; see Table A2.

A.22 The ARIMA and Activity Balances models predicted growth for 2003q2 in the range 0.7%-1%, in contrast to the m1 prediction of -0.4%. Although still far lower than the outturn of 5.3% the forecasts from the alternative models are again better than those of the m1 estimator, and indeed in-line with the DTI's m2 estimate made a month later. As in 2003q1 perhaps a correction to the forecast given to the ONS by the DTI at m1 should have been made. It is dangerous for the DTI to focus on the forecast of one model alone.

A.23 If these models had been used instead of emphasis being given to the Orders model the revision to construction output growth between the first and third estimates for 2003q2 therefore would have been about 4.4% instead of 5.7% and in turn the revision to GDP growth may have been smaller; see paragraph 9.45 for further discussion. Certainly, the DTI and ONS would have been able to produce as accurate estimates of construction growth at m1 as they actually did at m2.

A.24 The DTI appear to be have been unduly influenced by the negative growth in 2003q1 when forecasting 2003q2. Not only did the alternative models, including their own Activity Balances model, indicate a rise in 2003q2 but so, albeit two months after the end of the quarter, did the early survey returns which indicated a rise of 4.5%.

Annex

Table A1: RMSE of forecasts from various models against the final and m3 DTI construction output growth estimates[41]

Forecast Model	Final	M3	Final	M3
	1999q1-2002q4	1999q1-2002q4	1999q1-2003q1	1999q1-2003q1
DTI m1	1.479	1.502	1.990	2.007
DTI m2	1.610	1.604	2.110	2.106
ARIMA(0,2,2)	1.661	1.625	1.790	1.759
ARIMA(0,1,1)	1.522	1.490	1.657	1.629
ARIMA(BIC,1,BIC)	1.523	1.491	1.658	1.631
Disaggregate ARIMA(BIC,1,BIC)	1.654	1.562	1.806	1.727
Activity Balances (no dummy)	1.409	1.377	1.617	1.591
Activity Balances (dummies in 1998)	1.390	1.358	1.610	1.584
Activity Balances (dummies in 1998 & 2000)	1.377	1.344	1.604	1.578
DAR	1.711	1.612	2.010	1.932

Table A2: Forecast combination/encompassing tests. Robust t-values are in parentheses below the coefficient estimates. For each null hypothesis, a or b, we report the p-value of failing to reject the null hypothesis.[42]

		W_1	W_2	Ha	Hb
1999q1-2002q4	m1 vs. Activity Balances (dummies in 1998 & 2000)	0.732 (0.904)	0.731 (0.862)	0.000	0.000
1999q1-2002q4	m1 vs. ARIMA	-0.245 (-0.314)	1.505 (0.366)	0.162	0.159
1999q1-2003q1	m1 vs. Activity Balances (dummies in 1998 & 2000)	-0.332 (-0.379)	1.249 (1.086)	0.212	0.842
1999q1-2003q1	m1 vs. ARIMA	-0.161 (-0.216)	-4.237 (-1.106)	0.214	0.272

[41] The DTI's m2 estimates are available one period later, from 99q2.

[42] ARIMA refers to the ARIMA (BIC,1,BIC) model.

Annex A.4.3 Evaluation over the period 1999q1-2003q3

A.25 We have seen that the inadequacy of the construction estimates produced by the Orders model should have been anticipated *ex ante* and corrections possibly made. But where does this leave us now? To address this question we finally evaluate over the full out-of-sample period.

A.26 Table A3 compares the forecasting performance of the DTI's m1 estimator against the alternative models over the period 1999q1-2003q3. Although, at least, better than the benchmark DAR model, the m1 and m2 estimators produce forecasts that are worse than all of the alternative models' forecasts. They have a RMSE of at best 2.2, compared with a RMSE of 1.8 for the best alternative model.[43] Indeed, in contrast to the m1 and m2 estimates, the forecasts from the Activity Balances models, and to a lesser extent the ARIMA models, are statistically better than those of the DAR model as judged via the corrected Diebold-Mariano test; see Harvey, Leybourne and Newbold (1997).

A.27 We also considered augmenting the Activity Balances model in two simple ways. The results are presented in Table A4. First, we added a lag of the dependent variable into the regression. For each of the three Activity Balances models (distinguished by their treatment of dummy variables) adding in this lag did not improve the RMSE. Secondly, we added a lag of the survey variable. Model selection criteria were used to choose between these alternative representations of the Activity Balances model. In fact, the BIC always chose the basic model, with no lags of either the dependent or explanatory variable. However, the model augmented with one lag of the survey variable does deliver, albeit only slightly, better forecasts; see Table A4.

A.28 Table A5 indicates that more weight should be attached to the competing forecasts, when undertaking a forecast combination exercise. The ARIMA model should receive a weight of 1.5 in contrast to a weight of -0.2 on the DTI's own m1 estimate; the Activity Balances model should receive a weight of 2.4 in contrast to a weight of -0.7 on the m1 estimate. Indeed in this latter case, the weight given to the alternative model is statistically significant in contrast to the weight on the m1 estimate that is insignificantly different from zero.

A.29 Table A5 also shows that whether compared against the final or real-time outturn the m2 estimates do not encompass the m1 ones. This is, in a sense, 'worrying' as one should hope that the m2 estimates encompass the m1 estimates as they are formed using all the information that was available at the time the m1 estimate was made, plus additional information accruing after the m1 forecast was made. This does not appear to be the case and is consistent with our finding of little revision between m1 and m2.

[43] As we discuss below, and as can be seen from Table A3, combining individual forecasts delivers yet more gains in terms of RMSE.

A.30 Table A5 also considers combining all three forecasts, namely the DTI's m1 estimator, the ARIMA model and the Activity Balances model. It shows that both the ARIMA and Activity Balance forecasts receive a larger weight (at 4.9 and 3.2, respectively) than the m1 forecasts (with a weight of -0.7). Moreover, the weight attached to the Activity Balances model is statistically significant with a t-value of 2.4. This implies that these alternative models contain more information about the outturn than the m1 estimates. There is also evidence that the forecasts from the Activity Balances model encompass those of m1: relative to the Activity Balance model the m1 estimators contain no additional information about construction output growth.

A.31 Combining the three forecasts (i.e. m1, ARIMA and Activity Balances) over the period 1999q1-2003q3 according to the estimated (optimal) weights given in Table A5, we find that the RMSE falls to 1.615 against the final estimates and 1.624 against the m3 estimates; see Table A3.[44] This indicates that combination can deliver better forecasts than any of the individual forecasts considered in Table A3. These weights, of course, are based on full-sample information that would not have been available in real-time. In principle one could recursively estimate the forecast combination regression and then update the forecast combination weights. However, since our out-of-sample period begins in 1999q1 only this is not possible here. Nevertheless, the results presented do suggest that further examination of the benefits of pooling is required. Pooling or forecast combination provides a way of insuring oneself against the possibility that one particular model has "broken down".

Table A3: RMSE against the final and m3 DTI construction output growth estimates: 1999q1-2003q3.[45] 95% critical value for DM test is 1.7

Forecast	final	m3	Diebold Mariano DM test against DAR
DTI m1	2.305	2.318	0.994
DTI m2	2.249	2.246	1.004
ARIMA(0,2,2)	2.016	1.991	1.392
ARIMA(0,1,1)	1.924	1.903	1.589
ARIMA(BIC,1,BIC)	1.925	1.903	1.584
Disaggregate			
ARIMA(BIC,1,BIC)	2.0385	1.976	1.373
Activity Balances			
(no dummy)	1.838	1.817	1.714
Activity Balances			
(dummies in 1998)	1.825	1.805	1.729
Activity Balances			
(dummies in 1998 & 2000)	1.816	1.795	1.736
Optimal combination	1.615	1.624	1.537
DAR	2.744	2.693	–

[44] Consistent with the regression used to combine the forecasts, an estimated intercept, of -3.398, is included also.

[45] The DTI's m2 estimates are available one period later, from 99q2.

Table A4: Alternative forecasting models using the Activity Balance. RMSE against the final and m3 DTI construction output growth estimates: 1999q1-2003q3[46]

Forecast	final	m3
Activity Balances		
(no dummy)	1.838	1.817
Lag dependent variable	1.905	1.892
Lag of balance variable	1.827	1.807
Activity Balances		
(dummies in 1998)	1.825	1.805
Lag dependent variable	1.936	1.909
Lag of balance variable	1.815	1.796
Activity Balances		
(dummies in 1998 & 2000)	1.816	1.795
Lag dependent variable	1.961	1.940
Lag of balance variable	1.807	1.787
DAR	2.744	2.693

Table A5: Forecast combination/encompassing tests: 1999q1-2003q3. Robust t-values are in parentheses below the coefficient estimates. For each null hypothesis, a, b or c, we report the p-value of failing to reject the null hypothesis

	W_1	W_2	Ha	Hb
m1 vs. m2	-0.783 (-0.469)	0.580 (0.448)	0.157	0.155
m1 vs. ARIMA	-0.245 (-0.314)	1.505 (0.366)	0.162	0.159
m1 vs. Activity Balances (dummies in 1998 & 2000)	-0.692 (-0.750)	2.359 (2.082)	0.006	0.087

	W_1	W_2	W_3	Ha	Hb	Hc
m1 vs. ARIMA vs. Activity Balances (dummies in 1998 & 2000)	-0.693 (-0.823)	4.861 (1.370)	3.165 (2.429)	0.001	0.000	0.046

[46] The DTI's m2 estimates are available one period later, from 99q2.