Journal of Official Statistics, Vol. 20, No. 4, 2004, pp. 615–621

# Discussion

*D. M. Egginton*[1]

## 1. Introduction

Those not closely involved with the production and use of official statistics are often surprised to discover that published economic data are frequently revised and that these revisions are often significant far into the past. For economists, familiarity with this problem has brought an unwillingness to acknowledge its potential scale. If data were indeed unreliable, in the sense that different vintages of data were statistically unrecognisable from each other, then empirical studies would only be valid for that particular data vintage. Faced with such a problem most, but not all, economists have taken the line of least resistance. They have tended to assume that data revisions are not a problem and, even if they are, the problem is located in the most recent data. Consequently, dropping these recent observations from statistical analysis can circumvent this. This raises the question of how many observations should be dropped from the end of the sample. Researchers who adopt this approach often implicitly determine this number of observations by reference to their need to carry out forecasting tests to establish the robustness of their model. However, whilst the choice of, say, 8 quarters may make their study comparable with others in the literature, such a choice may not eliminate all of the data that will be subject to revisions. Indeed, the existence of data at different stages in the revision process means that the procedure of reserving data for forecasting tests is questionable because the model is estimated on different data from the data the forecasting test is applied to.

## 2. Terminology

Examination of data revisions is, therefore, an important problem and Kerry Patterson and Saeed Heravi in their article in this journal squarely face up to the problem of whether revisions led to statistical differences between different versions of the data, in this instance U.S. GNP. It should be noted that Patterson and Heravi (2004) (henceforth PH) do not examine different data vintages where vintage refers to the date the data were published – the real time data literature. Rather they compare a series of preliminary data with the "final" data, second release data with the "final" data and third release data with the "final." With the exception of the "final" data their data is a mixture of real time vintages, "multi-vintages," and the title of the PH article refers to this older use of the term vintages.

[1] Daiwa Institute of Research, 5 King William Street, London, UK. Email: donald.egginton@dir.co.uk
**Acknowledgment:** The views expressed here are the author's alone and should not necessarily be attributed to the Daiwa Institute of Research.

Whilst on the question of terminology, another point should be raised. This is PH's use of the term "final." The Bureau of Economic Analysis (BEA) statisticians designate one of their early National Accounts releases as the "final" release and this may confuse the unwary. It might, therefore, have been preferable to describe the "final" figures as the "benchmark." PH correctly note that the benchmark vintage is no such thing. It is usually just the latest available to the researcher and, therefore, has no more claim to be special than any other vintage of data. The release schedule of the BEA is discussed further below.

## 3. Advantages of the "Multi-vintage" Approach

The "multi-vintage" approach of PH cleverly avoids one potential problem of examining data revisions by comparing vintages. With there being as many vintages as there are time periods in the data, this can result in a possible $n(n-1)/2$ comparisons ($n$ is the number of vintages), which can easily result in the message being lost in the detail or the concern that the choice of another pair of vintages to compare may lead to different results. Croushore and Stark (1999) (henceforth CS), for example, simply chose vintages that were the last prior to the comprehensive revisions to the national accounts to compare with their benchmark vintage. This raises the suspicion that this choice will maximise the difference between the series, whereas choices of vintages at less than the five-year revision cycle would produce smaller differences, unless the comparisons are across years that contain comprehensive revisions. PH's approach side steps these choices and is preferable to CS's methodology.

PH set out five questions in their article and these and PH's answers are set out below.

(1) Do the different series share the same long-run movement? Answer yes.
(2) Do different series share the same short-run movements? Answer no.
(3) Can we use just one series of data? Answer yes.
(4) Do data revisions arise from measurement errors or efficient forecasts? Answer measurement errors.
(5) Is there a unit root in U.S. GNP? Answer yes.

Question 1, whether or not the data share a single stochastic trend (a common trend), is the key question of the article. By asking the question in this form PH acknowledge that the GNP data are nonstationary and that this has to be taken into account in choosing the analytical technique. PH choose the Johansen technique (Johansen 1988) to estimate cointegrating VAR models. This choice may be contrasted with that of CS who, after taking logs (as do PH) to induce approximate linearity, difference the data to give approximate growth rates from which they calculate descriptive statistics. This induces stationarity but the cost of differencing is that information about the long-run relationship is lost.

## 4. Rebasing

There is, however, a benefit of differencing as growth rates can be compared across vintages whereas the levels of real GNP suffer from periodic rebasing as the price data are changed. Unfortunately, the rebasing is often coupled with methodological changes so that a simple rescaling of the GNP data potentially confuses vintage effects with price base changes. A good example of this problem is to be found in the CS data set. Prior to 1992 the data are for GNP but from the first quarter of 1992 onwards they are for GDP. (This change is due to the need to keep the date of the vintage constant within the year.) This matches the change to GDP from GNP as the primary measure of U.S. production (see Bureau of Economic Analysis 1991). Unfortunately, the data are also rebased for the first quarter 1992 figures to 1987 prices, from 1982 prices. Thus, simple rescaling factors will mix definition and price changes.

There are a number of ways of getting around the simultaneous methodological and price change problem. One is to use nominal GNP data and then turn them into real data by deflating by a price index that is not revised over time. This is the approach used by Patterson (2002), who used the retail prices index (RPI) to deflate UK GNP. Unfortunately, it is not clear that the RPI is an appropriate deflator for GNP. Alternatively, it is possible to estimate a system of equations for different vintages and price bases by seemingly unrelated regression techniques and imposing across equation constraints to recover the price conversion factors. This is the method used by PH and follows from their earlier article Patterson and Heravi (1991). Patterson (2002), in his study of UK GNP, rejected this method because it "fundamentally altered the properties of the series" and in the current article the data from 2001 could "not be handled by standard rebasing methods." Given that this methodology is not foolproof, the article might have benefited if it had contained more details of the methodology used. This raises the question of whether the rebasing method alters PH's results significantly. To examine this I attempted to replicate PH's study using CS's data, which is available from the Federal Reserve Bank of Philadelphia's web site at www.phil.frb.org/econ/forecast/readow.html

However, the CS data needed some adjustments prior to the analysis, the main one being the change in definition from GDP to GNP in 1992 as described above. To remove this problem, I added estimates of net income from overseas at 1987 prices to the GDP estimates available in the first quarter of 1992. This changes the GDP figure back into a GNP figure and allows a corrected price factor to be calculated. As net income from overseas was positive during this period the comparison of GNP to GDP underestimates the price conversion factor by almost 1%. A figure for the final quarter of 1995 is also included. (It was omitted from CS's data because its publication was delayed beyond the mid-February cut off date for inclusion.)

Although not reported in the original article, the factors that PH use to convert the first three data series to a constant 1992 $U.S. base are given in Table 1. These can be

*Table 1. Price conversion factors used by Patterson and Heravi and Egginton*

| | PH | Egginton | % difference |
|---|---|---|---|
| 1958 prices to 1972 prices | 1.4826 | 1.4761 | 0.4 |
| 1972 prices to 1982 prices | 2.1222 | 2.1148 | 0.4 |
| 1982 prices to 1987 prices | 1.1750 | 1.1788 | −0.3 |
| 1987 prices to 1992 prices | 1.2293 | 1.2314 | −0.2 |
| 1992 prices to 1997 prices | N/A | 1.1281 | |
| 1997 prices to 2000 prices | N/A | 1.0675 | |

Source: Private communication with Kerry Patterson and own calculations using the CS data.

compared with the conversion factors that have been calculated by simply taking the ratio of GNP pre and post the price base change for the fourth estimate of GNP. This arbitrary choice was designed to keep the conversion factor close to the first three estimates of GNP, but to avoid the earlier estimates where more measurement errors may be corrected. As can be seen from Table 1, the percentage difference between these factors, despite being calculated in different ways, is not particularly large. It therefore seems likely that the methodological effects are relatively small and the use of this method does not materially alter the results found by PH. Of course, the simple agreement of these two sets of conversion factors need not lead to the same results in the Johansen analysis, and to check this we performed the analysis on the adjusted CS data.

## 5. Differences in PH and CS Data

Before replicating the Johansen analysis a further adjustment to the CS data is required. The second estimates of GDP in 1991 q4 and 1991 q3 and the third estimates of GDP in 1991 q4 are being compared with the benchmark GNP estimates. I circumvent this problem by changing these data points from GDP to GNP by adding the figures for net income from overseas at 2000 $U.S. prices from the BEA.

There is an important difference between the PH data set and that of CS. As CS's data are sampled in the middle of the quarter, their data omit the second revision to GDP. The BEA publishes GDP data on the following timescale. At the end of the month following the end of the quarter an advance estimate is released. At the end of the second month a preliminary estimate is released and at the end of the third month (i.e., at the end of the next quarter) a final estimate is released. CS data will, therefore, register the advance data but omit the preliminary estimate. Thus, our second estimate is more akin to PH's third estimate. This pattern of three revisions and CS sampling dates means that for most of the quarters the third estimate will be unrevised from the second estimate. However, each year there is an additional revision, known as the annual revision, normally undertaken in July. This means that CS data for the fourth quarter will potentially change between the second and third estimates. This pattern is not completely secure, however, because in the years that comprehensive revisions occur (i.e., significant revisions to surveys and/or methodology) the annual revisions are not made. When comprehensive revisions are made (every five years) it is the second quarter estimates that will change between our second and third estimates. This still leaves five changes between the second and third estimates unaccounted for. Thus our data set is rather different to PH's and this should be borne in mind when analysing the results below.

There is a further point about data release timings that does not affect the CS data but does affect PH's. Between 1958 and 1973 only two estimates of a quarter's GNP were made and these were called the preliminary and final (although they had similar publication dates as the advance and preliminary estimates). Thus, prior to 1974, PH's data will only change between the second and the third estimates due to the annual and comprehensive revision processes. CS's data, because they sample in the middle of the month, are unaffected by this change. As Hendry and Juselius (2000) conclude that statistical inference is sensitive to the assumption of parameter nonconstancy, which this data process engenders, the rerunning of PH's analysis, either excluding data prior to 1974 or allowing for the parameters to shift, would be useful.

## 6. Replicating the Johansen Analysis

After adjusting the CS data using the adjustment factors given in Table 1 and taking natural logs, the Johansen analysis is performed. The data extends that used by PH in that the period covered is longer (1966 Q3 to 2003 Q1) and we change all the data to a 2000 $U.S. price base. This, unlike PH where the benchmark vintage is expressed in 1996 $U.S., means we can test restrictions on the benchmark cointegrating vector, because we are not trying to simultaneously estimate the price conversion factors.

As in the case of PH, an analysis of the trace statistics for the four vintages revealed that there were three cointegrating vectors and this was imposed upon the model. The VAR model with four lags was then estimated with an unrestricted constant and a time trend restricted to the cointegrating vectors, and this forms our base model. We then imposed the restrictions (not all of them binding) that excluded the time trend completely, set some of the coefficients to zero and some to plus or minus unity. The loading factors were also restricted so that the equilibrium correction mechanism for the first GNP series only appears in the first VAR, the equilibrium correction mechanism for the second GNP series only appears in the second VAR and the equilibrium correction mechanism for the third GNP series only appears in the third VAR. Altogether there are 14 restrictions and a likelihood ratio test gives a value of 18.183, and the restrictions cannot be rejected at conventional levels of significance (probability 0.199). If the further restrictions are imposed that the constant terms only lie within the equilibrium correction mechanisms (not the VARs), then these restrictions are convincingly rejected with a likelihood ratio test of 34.137, which has a probability value of just 0.002. Thus, the equilibrium correction mechanism for the first estimate is the difference between the first and the second estimates, and for the second estimate it is the difference between the second and third estimates. For the third estimate the equilibrium correction mechanism is given by the difference between the third estimate and nearly 96% of the benchmark estimate (the parameter estimate is 0.957 with a standard error of 0.008).

The results indicate that, as in the case of PH, for the first three multi-vintage series and the benchmark series just one stochastic trend drives the $I(1)$ component in each of the series. There is evidence that only the benchmark series may be weakly exogenous, implying that this is the permanent component of GNP, hence the answer to Question 3 above is, as in PH's research, yes. Furthermore, because the revisions are stationary and the benchmark GNP series is weakly exogenous, this implies that the revisions between the earlier estimates of GNP and the benchmark estimate are measurement errors. This answers Question 4 above and the same result is found by PH.

We use these equilibrium correction mechanisms, in an otherwise differenced equation, to estimate, by FIML, how much of the unrestricted constant should be present in the cointegrating vectors. Although we could impose the loading factors from the multivariate analysis, we have chosen to reestimate them in the difference equations. We have imposed the restriction that each equilibrium correction mechanism enters only one equation at lag 1. An initial lag length of 4 was again used and statistically insignificant variables were removed. Despite there being 18 variables in each equation this elimination process resulted in very similar models of a constant, the equilibrium correction mechanism and the difference of the third series lagged by 3 quarters. Using these estimates the difference between the trends of the first and the third series is 0.9%; and allowing for the timing differences in the CS data set relative to PH's, as described above, the difference in the trends is 0.4%.

## 7. Short-run Movements in GNP

We can use the dynamic equations to examine PH's second question: "Do different vintages share the same short-run movements?" PH study this by using a further refinement of their multivariate approach to analyse common cycles. They find that there are no common cycles and, at the prompting of a referee, consider whether the first and second GNP series may provide misleading signals about the state of the economy. In particular, they contemplate whether or not these series may misidentify turning points in economic activity, with the presumption from their results that they do misidentify turning points. This view is in conflict with that of the BEA's statisticians. For example, Fixler and Grimm (2002) state that the current quarterly estimates of GDP successfully indicated the cyclical peak in four of the last five recessions in the period 1969–2000. They also admit that the current estimate may have missed the start of the 2001 recession. They claim that the current real GDP estimate captured three of the five cyclical troughs in this period. However, they do note that the quarterly estimates overstate the decline in real GDP in the quarter following a peak, and understate the recoveries in the quarters at, and following, the cyclical trough. Nevertheless, they conclude that real GDP is reliable and presents a useful picture of economic activity.

The relatively few turning points in the data make deciding whether or not the early GDP estimates are reliable predictors of turning points more of an art than a statistical science. Moreover, the remarks from Fixler and Grimm (2002) ignore how often the current estimates indicated a turning point that never occurred, the false positive problem. However, PH concentrate on the statistical significance of all the cyclical data and their finding that there are no common cycles does not mean that the differences in the cycles are numerically significant. As a practitioner I would regard differences of $\pm 0.1\%$ in quarter on quarter growth rates as being negligible and differences of $\pm 0.2\%$ as not particularly significant unless the growth rate changed signs between the estimates. Moreover, a turning point requires two consecutive quarters of large deviations to be regarded as significant, I would tend to assume that a single large deviation is just noise. Using these criteria there are relatively few instances when the unexplained growth from the FIML estimates diverges between the first and second estimates, only 37 (just over 25%), and there are only two pairs of quarters when these large divergences are sequential. A graph of the residuals from the first and second GDP estimates would show little systematic deviation. A comparison of the first and the third estimates residuals record 46 (31%) instances where the difference is larger than $\pm 0.2\%$, but there are just four pairs of quarters where these large differences are sequential. My view would be that the differences between the first, second and third estimates of GNP rarely point to differences in economic behaviour that are numerically significant. This does not mean that the different GNP series share the same short-term movements; hence the answer to PH's second question is still no, but claims that the series may signal different turning points seem unwarranted.

This leaves the fifth question posed by PH, "Is there a unit root in U.S. GNP?" In a sense asking these sorts of questions raises the potential for a gamut of responses. Suppose that a difference has been detected in the GNP series but each series has a unit root. Is the conclusion that unit roots in GNP is a robust hypothesis; or is it that the unit root tests are too insensitive to distinguish the different GNP series; or is it that the tests of GNP differences are too sensitive? Work in this area is in its infancy and it is in this area, rather than examining different common cycles, that PH's work could be profitably extended.

## 8. Conclusions

PH's commendable article examines an important subject in a technically advanced manner. This allows them to examine questions that other researchers have not been able to. Their central finding that early estimates of GDP can tell researchers about long-term trends but may be misleading about short-term movements is unexpected. However, the view expressed in the article (albeit at the behest of the referee) that the differences in GNP series may include different information about turning points is not supported by an examination of the data. Despite these criticisms, and a few quibbles about terminology, this is a highly useful article on data revisions.

## 9. References

Bureau of Economic Analysis (BEA) (1991). Gross Domestic Product as a Measure of U.S. Production. Survey of Current Business, August, p. 8.

Croushore, D. and Stark, T. (1999). A Real-Time Data Set for Macroeconomists: Does the Data Vintage Matter? Working paper no. 99–21, December, Federal Reserve Bank of Philadelphia.

Fixler, D.J. and Grimm, B.T. (2002). Reliability of GDP and Related NIPA Estimates. Survey of Current Business, January, 9–27.

Hendry, D.F. and Juselius, K. (2000). Explaining Cointegration Analysis: Part II. University of Copenhagen, Institute of Economics, Discussion Paper 00–20.

Johansen, S. (1988). Statistical Analysis of Cointegration Vectors. Journal of Economic Dynamics and Control, 12, 231–254.

Patterson, K.D. (2002). The Data Measurement Process for UK GNP: Stochastic Trends, Long Memory, and Unit Roots. Journal of Forecasting, 21, 245–264.

Patterson, K.D. and Heravi, S.M. (1991). Data Revisions and the Expenditure Components of GDP. The Economic Journal, 101, 887–901.

Patterson, K.D. and Heravi, S.M. (2004). Revisions to Official Data on U.S. GNP: A Multivariate Assessment of Different Vintages. Journal of Official Statistics (this volume).

Received March 2004