Journal of Official Statistics, Vol. 20, No. 4, 2004, pp. 607–614

Discussion

Peter van de Ven[1] *and George van Leeuwen*[2]

1. Introduction

First estimates on quarterly growth of Gross National Product (GNP) or Gross Domestic Product (GDP) are published shortly after the end of the reporting period, sometimes, as for the United States, even within one month after the end of the relevant quarter. These first estimates are subsequently revised a number of times as more basic data become available. In their study, Patterson and Heravi apply the co-integration framework of Johansen (1995) and recent extensions of this framework proposed by e.g., Vahid and Engle (1997) to the different vintages of first and revised data of U.S. GDP. On the basis of their analysis, they conclude that the different vintages have a single trend. The hypothesis of short-run common cyclical behaviour, however, has to be rejected. The latter result may primarily be caused by measurement errors, given the "noisier" character of the first estimates.

Although the analysis of Patterson and Heravi provides a very interesting example of how to use new time series tools for the assessment of the data measurement process (DMP) of GDP, as a statistician one stays more or less empty-handed when looking at the practical implications of their results and conclusions. In the end, the conclusions are not particularly surprising for an insider when it comes to the statistical process of compiling GDP growth rates. In our opinion, national accountants would have been better served if Patterson and Heravi had extended their analysis to the relationship between subsequent estimates of GDP/GNP on the one hand and subsequent estimates of underlying variables on the other. In order to provide a frame of reference for the discussion of their article, we will first draw attention to recent developments in relation to the pressure to publish timely and reliable data, the process of compiling GDP-estimates, the analysis of revisions and the need to use quality frameworks in order to improve transparency and reproducibility. Thereafter, we will discuss the article in more detail. In closing, we will make some suggestions for future research that uses a different multivariate setting.

2. The Need for More Timely and More Reliable Data

In general, one can see a growing demand for (even more) timely estimates of macro-economic aggregates from people involved in economic policy, the media and the general

[1] Head of the Department for National Accounts, Statistics Netherlands, Voorburg, The Netherlands. Email: pven@cbs.nl
[2] Senior Researcher, Department of Methods and Informatics, Statistics Netherlands, Voorburg, The Netherlands. Email: glwn@cbs.nl
Acknowledgments: The views expressed in this comment are those of the authors and do not necessarily reflect the policies of Statictics Netherlands.

public. Within the European Union, for example, a political request for timely data on the EU level has resulted in the setting-up of a priority list of so-called Principal European Economic Indicators (PEEIs) for which timeliness has to be improved. One of the PEEIs is the release of quarterly volume growth of GDP within 45 days after the end of the relevant quarter.

Meeting the demand for timely data is a major challenge for statistical offices all around the world. In most cases, improving timeliness means compiling estimates with less data or data of lower quality and/or with larger reliability intervals. How can one improve timeliness without losing too much reliability? How can one balance the need for fast estimates with the need for reliable statistical data? Often, users of statistical data are simply not aware of the intrinsic uncertainties surrounding e.g., first estimates of GDP growth. And it is quite hard to explain these uncertainties. This makes statistical offices somewhat cautious, as a significant revision of first estimates may result in a major loss in credibility in respect of official statistics. This is especially true in the case of important aggregates such as economic growth. One only has to look at newspaper comments on some major revisions of U.S. GDP growth. The trade-off between timeliness and reliability of estimates has therefore been an important point of discussion when setting dates for the timeliness of the above mentioned PEEIs. A benchmark in this discussion was the timeliness of macro-economic aggregates in the U.S. In the end, it has been decided not to pursue the timeliness of 25–30 days as for the first estimates of U.S. GDP.

The response of most statistical offices to the above challenges is a continuous search for possibilities of improving the reliability of the first estimates. Furthermore, there is a growing awareness of the need to improve transparency, not only to get a better insight into the compilation process and to make the compilation more objective and reproducible, but also to provide for a better and more transparent communication towards the users of statistical data. In the following section, this line of research will be briefly discussed, with a special focus on The Netherlands.

3. The Compilation of Fast Estimates and the Search for Improvement

Before turning to the search to improve GDP estimates, it may be useful to shed some light on the practice of GDP compilation. GDP can be calculated in three ways:

i) Production approach: the sum of value added of all industries;
ii) Expenditure approach: the sum of final consumption, capital formation and exports minus imports;
iii) Income approach: the sum of remuneration of employees, operating surplus of enterprises and taxes less subsidies on production.

There is an accounting identity between the three approaches. In theory, they should give the same result. In practice, however, the original source data for each of the above-mentioned items need to be balanced in the framework of national accounts.

When compiling first estimates, only partial source information is available. Depending on the available information, the process of compilation may differ from country to country. Some countries may predominantly use the production approach, others the expenditure approach. Some, like The Netherlands, combine these two approaches. There

may also be a different attitude towards the use of more statistical techniques or towards the use of econometric "now-casting" techniques.

In the U.S., first estimates of GDP are mainly based on expenditure and employment data. Estimates from the expenditure side and the ones from the income side are not reconciled with each other. In The Netherlands, a different method of compilation is used. Right from the start, for the first so-called flash estimate (45 days after the end of the relevant quarter), supply (output and imports) and use (intermediate consumption, final consumption, capital formation and exports) of approximately 100 different products are balanced in a supply and use table. In doing so, output and intermediate consumption are allocated to approximately 100 industries. Of course, at the time of the first GDP estimate, full source information is not available or only partially available (e.g., one or two months of the relevant quarter). The methodology for the subsequent estimates is basically the same. "Only" more source data is available and more detail is added. Furthermore, data on income is added into the process of confronting and reconciling data from different sources. Apart from benchmark revisions (once every 5–10 years), a full 'revision cycle' covers a time span of about two and a half years, in which GDP growth data is revised four times.

In order to improve the estimates of GDP, the first step in the research is, of course, a detailed descriptive analysis of the revisions. What is the magnitude of the revisions? Is there an upward or downward bias in them? The analysis, however, should not only relate to the revisions of the GDP growth itself. Also a decomposition analysis has to be made. In the case of the Dutch method of estimation, data sources link up with value added (or output minus intermediate consumption) per industry and categories of expenditures. Questions relate to which components of GDP and which sources actually cause the major proportion of revisions.

For The Netherlands, it is found that for the period 1991–2000, the total of revisions between the first estimate and the final estimate is 0.35% upwards on average.[3] The most significant average revision takes place when compiling the very detailed 'final' estimates. It should be noted that revisions as a result of major benchmark revisions (on average once every 5–10 years) are not taken into account here. Estimates before and after a benchmark revision should not, in our opinion, be compared to each other, as they may result from changes in recording (e.g., recording of software as capital formation instead of intermediate consumption) and/or changes in methodology (e.g., using chained indices instead of indices with a fixed weighting scheme; a method that, by the way, has been applied in The Netherlands since the beginning of the eighties). At a more detailed level, it shows that, on the production side, especially commercial services and trade have contributed significantly to the revisions of GDP growth. On the expenditure side, the contribution of final consumption of households was the largest.

The second step in the research is to look for possibilities of improving the source data. Several ways are pursued here. They range from a more detailed analysis of major causes

[3] See Brugt Kazemier, Henk Nijmeijer and Remko Hijman, How to Judge the Reliability of Provisional National Accounts (paper prepared for the European Conference on Quality and Methodology in Official Statistics, Mainz, Germany, 24 – 26 May 2004). Furthermore, a report containing a more detailed analysis of revisions of Dutch GDP for the year 2000 has been made public. In future, this will be done on an annual basis for subsequent periods.

of revisions in source data to improving timeliness of source data, e.g., through the application of new "now-casting" techniques, and trying to find new indicators for industries and/or expenditures for which source data are missing. Regarding the latter two possibilities, it can be noted that, in the case of The Netherlands, the first flash estimate is based on direct source data for approximately 35% of GDP.

In general terms, one can state that, as compared to the U.S. methodology, the Dutch method makes more use of available source data in the attempt to reconcile supply and use of goods and services. On the other hand the U.S. method is more transparent and better reproducible than the one applied in The Netherlands. To make possible a better evaluation of Dutch revisions and to increase transparency and reproducibility of estimates, a more systematic compilation of so-called "process tables" will be pursued in The Netherlands. In these tables, the whole estimation process from basic source data to the end-result will be recorded systematically, not only for quarterly estimates but also for annual data. In doing so, source data will be put into different categories such as directly available information, models and extrapolations. In addition, changes to source data will be distinguished into e.g., changes for discontinuities in the source data, changes in concepts, and changes due to reconciliation of the source data, i.e., making supply and use consistent with each other. We expect that this will also prove to be a powerful tool to meet the increasing demand from our main users (and suppliers of source data as well) for transparency and accountability.

From the discussion above, we conclude that it would be fruitful to analyse the revisions of GDP growth in coherence with revisions of the composite elements of GDP, be it either value added by industry or a breakdown by expenditure categories. Such an endeavour could reveal which of the component estimates should be improved in order to obtain more timely and more adequate data on short-term developments and swings in the business cycle. After all, this is the primary concern and the most important goal of the flash estimates of GDP growth. One can argue that the most recent releases of (provisional) GDP data are more important than "old" data, in the sense that they add the "real" news to the information sets of economic agents and policy makers. And if the research by Patterson and Heravi shows that the vintages of GDP estimates "can differ in the short-run picture," we have to acknowledge that we do a poor job. Below, we will persue their results in more detail and we return to this question as well.

4. The Distinction Between the Short and Long Run

Whereas the preceding discussion stressed the importance of adequate data on the most recent state of the economy, the main focus of the article of Patterson and Heravi is on the long-run properties of the different vintages of GDP and their relation to data revisions. We first comment on their choosing to emphasise the role of revisions for the assessment of the long-run component of GDP and the research strategy derived from this choice.

The main concern of the article is that revisions of GDP should not distort the most important features of GDP as regarded within a long-term perspective. Indeed, it would be rather discomforting if the different vintages of GDP should drift apart as a consequence of data revisions. Thus, preferably, the long-term trend of GDP should be maintained in

different vintages. Otherwise one has reason to doubt the data measurement process (DMP) for this core variable of National Accounts.

A distinction between the short- and long-run comovements in different measurements of the same variable (GDP) also refers to the use of times series and their components of interest. In particular, the higher frequency component (the short-run trend-cycle) of GDP is of utmost importance for monitoring the present states of economies. By contrast, lower frequency components may be of more interest for research aimed at understanding the fundamental forces that are driving economic growth (as is the case in growth studies). The estimation of these two components (more precisely: the trend and cycle extraction from a time series) has formed a long tradition of empirical research. Recent examples of this strand of research stress the view that the different components underlying a single time series may be driven by different (stochastic) forces.[4]

However, the article is not about signal extraction, but about the statistical testing of the properties of data revisions. From the viewpoint of the main concern of the authors this is an understandable research strategy. As the different realisations of GDP refer to the same (generic) variable, one may expect them to be driven by similar stochastic trends. This explains the decision to adopt the co-integration framework of Johansen (1995) to investigate the issues at stake. Another reason for discarding the signal-extraction problem is that quarterly GDP time series also consist of a seasonal component. Accounting also for the role of seasonal patterns in different vintages of quarterly GDP would complicate the analysis severely (see e.g., Hecq 1998).

Therefore, by placing the long-run properties of GDP at the centre of interest, the focus of the article and the research strategy chosen can be well understood. However, in spite of this, there remains the question whether the long-term properties of the various releases of the same GDP variable should be the main focus of analysis in view of other uses of GDP data (see the discussion in Section 2). Thus, one could question the authors' focus on the implications of data revisions for the long-run properties of GDP as the main interest is the most recent growth rates of GDP.

5. What Can Be Learned from the Empirical Analysis?

The most impressive part of the article concerns the use of the cointegration framework in order to reveal some important properties of the data revision process. There are several hypotheses involved in the testing of the role of revisions that show up in a sequence of GDP vintages. The first group of statistical tests concerns the presence of unit roots in different vintages. Not very surprisingly, it is found that all vintages are driven by stochastic trends. Subsequently, the article determines the number of common stochastic trends (or equivalently, the rank of the cointegration space). The results show convincing evidence for the existence of a single common stochastic trend in GDP, irrespective of the number of vintages used. This is an important result in view of a possible distortion of the long-run properties of the GDP time series as a consequence of data revisions. It implies

[4] The now frequently applied view of univariate times series modelling is that times series consist of different components and that ARIMA models are reduced-form representations of structural unobserved components models (see e.g., Harvey and Koopman 2000).

that revisions of the most recently released quarterly data do not alter the long-run properties of GDP time series as a whole.

The second stage in the testing sequence is designed to discriminate between two competing views on data revisions: the Efficient Forecast Hypothesis (EFH) and the Measurement Error Hypothesis (MEH). The article clearly explains how different hypothesis tests can be carried out in a cointegration framework in order to shed light on this important issue. The authors also show how these competing views on the nature of the DMP for GDP can be investigated for each pair of vintages separately. Thus one can test the first release for GDP of a specific quarter against the (conditional[5]) final estimate for the same quarter. Furthermore, one can also test "intervening" vintages against each other or against the first or (conditional) final estimate in order to get a better understanding of the revision process in view of these two competing hypothesis. Much of the article is devoted to the testing of these competing views on revisions, and the results for U.S. GDP are convincingly in favour of the MEH hypothesis.

Moreover, the empirical application also elaborates on the distinction between sequential revisions and cumulative revisions. The evidence presented clearly points to the importance of the final stage of the revision process. For U.S. GDP the corresponding co-integration parameter is close to -0.93, which implies that the difference between the preliminary estimates and the (conditional) final estimates amounts to approximately 7% of the final estimate.

Unfortunately, and as mentioned already, the results for the testing of common cyclical features in GDP vintages are less satisfying. After applying the test for common cycles proposed by Vahid and Engle (1993), the authors conclude that the different U.S. GDP vintages do not share common cycles.

6. Are the Results (Very) Surprising?

A first impression of comovement in the long run can be obtained even without the aid of econometric tests. Just collect all data in the same graph to verify that the complete path of all vintages is dominated by low frequency behaviour. Thus the results of the co-integration tests are not very surprising as they "reproduce" the well-known shape of the spectral density that is typical for time series of GDP. However, there is more in the article than the testing for unit roots in GDP. The "real meat" is in the characterisation of the revision process and the results of the many statistical tests point in the same direction, i.e., the DMP of U.S. GDP appears to be "well behaved" when it comes to its long-run properties.

However, things are different if we narrow the "observation window" to include only the end of the time series. Then, the failure to find common cycles in the most recent releases of GDP appears to be more cumbersome. In Section 3, several explanations have been put forward with regard to this result. The similarities between the final stages of the compilation process for U.S. and Dutch GDP may explain the relatively large contribution of "final" to the (conditional) final estimates. As mentioned before, it is only in this last

[5] 'Conditional' should be interpreted here as conditional to changes as a result of benchmark revisions.

stage of the compilation process that all balancing constraints can be taken into account. Thus the requirement to have a (conditional) final estimate of GDP (growth) that is consistent from the view of different approaches (production approach, expenditure approach, and income approach) may explain why the early releases of GDP (growth) cannot be efficient forecasts of final data.

Moreover, it also seems reasonable to conclude that the lack of short-run common features reported in the study reflects the other side of the same coin: the strong corroboration of the MEH hypothesis found in the (long-run) cointegration relations between the vintages. Stated otherwise, the negative results of the Vahid and Engle (1993) test statistics may also be linked to the strong assumption underlying the (augmented) cointegration analysis that long-run relations between variables and short-run dynamics are interrelated (see e.g., Hecq et al. 2000, for a discussion of this topic).

7. Alternative Approaches?

To conclude: Seen from the perspective of National Accountants, a cointegration analysis applied to GDP alone may not be the most appropriate route to reveal important features of the revision process. National Accounts are better helped with information that may guide them to assess the importance of early indicators for their final results. Noting that GDP is a balance of other aggregates, this requires a different multivariate setting. A natural starting point is to return to basics: use the cointegration framework as a model for the data generation process (DGP). Thus, reformulate the model in order to account for the different long-run and short-run dynamics of the underlying components. For instance, final consumption of households may be less sensitive to the phase of the business cycle than capital formation and thus may have different impact on the short-run dynamics of GDP.

Alternatively, one could generalise the modelling strategy as applied to UK Income and Consumption in Patterson (1995). This approach simultaneously accounts both for the DMP and (aspects of) DGP for main economic aggregates. It is also more naturally linked to the revision problem as revisions relative to the (conditional) final vintage can be interpreted as measurement errors. This interpretation of revisions enables a State-Space representation of the relation between the various vintages of GDP components that become available in the course of time (the measurement equations) and the state (the final vintage of GDP components). Given the result that preliminary vintages are better characterised as error-driven estimates than as efficient forecasts of the (conditional) final vintage of GDP, this approach also allows a better understanding of the revision process if "measurement errors" are interrelated.

8. References

Harvey, A. and Koopman, S.J. (2000). Signal Extraction and the Formulation of Unobserved Components Models. Econometrics Journal, 3, 84–107.

Hecq, A. (1998). Does Seasonal Adjustment Induce Common Cycles? Economic Letters, 59, 289–297.

Hecq, A., Palm, F.C., and Urbain, J.P. (2000). Permanent-Transitory Decomposition in VAR Models with Cointegration and Common Cycles. Oxford Bulletin of Economics and Statistics, 62, 511–532.

Johansen, S. (1995). Likelihood-Based Inference in Cointegrated Vector Autoregressive Models. Oxford University Press. UK: Oxford.

Kazemier, B., Nijmeijer, H., and Hijman, R. (2004). How to Judge the Reliability of Provisional National Accounts. Paper prepared for the European Conference on Quality and Methodology in Official Statistics, Mainz, Germany, 24–26 May.

Patterson, K.D. (1995). An Integrated Model of the Data Measurement and Data Generation Processes with an Application to Consumers' Expenditure. Economic Journal, 105, 54–76.

Vahid, F. and Engle, R.F. (1993). Common Trends and Common Cycles. Journal of Applied Econometrics, 8, 341–360.

Vahid, F. and Engle, R.F. (1997). Codependent Cycles, Journal of Econometrics, 80, 199–221.

Received March 2004