How to judge the reliability of provisional National Accounts

Brugt Kazemier, Henk Nijmeijer and Remko Hijman

Discussion paper (201110)

Statistics Netherlands

The Hague/Heerlen, 2011

Explanation of symbols

.	= data not available
*	= provisional figure
**	= revised provisional figure
x	= publication prohibited (confidential figure)
–	= nil or less than half of unit concerned
–	= (between two figures) inclusive
0 (0,0)	= less than half of unit concerned
blank	= not applicable
2010–2011	= 2010 to 2011 inclusive
2010/2011	= average of 2010 up to and including 2011
2010/'11	= crop year, financial year, school year etc. beginning in 2010 and ending in 2011
2008/'09–2010/'11	= crop year, financial year, etc. 2008/'09 to 2010/'11 inclusive

Due to rounding, some totals may not correspond with the sum of the separate figures.

Publisher
Statistics Netherlands
Henri Faasdreef 312
2492 JP The Hague

Prepress
Statistics Netherlands - Grafimedia

Cover
TelDesign, Rotterdam

Information
Telephone +31 88 570 70 70
Telefax +31 70 337 59 94
Via contact form: www.cbs.nl/information

Where to order
E-mail: verkoop@cbs.nl
Telefax +31 45 570 62 68

Internet
www.cbs.nl

ISSN: 1572-0314

60083201110 X-10

How to judge the reliability of provisional National Accounts[1]

Brugt Kazemier, Henk Nijmeijer and Remko Hijman[2]

Paper prepared for the European Conference on Quality and Methodology in Official Statistics, Mainz (Germany), 24-26 May 2004

Abstract

It is not only important to produce good quality statistics, but also that the users of these statistics believe that they are of good quality. Therefore, it is necessary that the subsequent provisional estimates of the national accounts show a similar picture of economic performance. I.e. the subsequent estimates have to be sufficiently reliable.

In this paper, it is analysed to what extent this requirement is met, taking into account that this requirement seriously conflicts with timeliness. A computer program was developed to enable a quick reliability-check to a large number of economic variables, derived from the national accounts. To conduct this test, a minimum reliability standard is defined, with respect to the difference between an estimate and the final estimates.

In many cases these checks will signal one-time events, but sometimes they will signal latent problems in either the basic statistics, underlying the national accounts, or the compilation (process) of the national accounts itself. Assuming that these problems will be solved, these checks may lead to more reliable provisional estimates, and in some cases perhaps more accurate final accounts.

Keywords: quality, national accounts, provisional estimates

[1] This paper is a revised and extended version of the paper by Brugt Kazemier and Robert van Rooijen titled "Assessment of the reliability of the Dutch provisional National Accounts" of may 2002 which was presented at the 27th General Conference of the International Association for Research in Income and Wealth in Stockholm, Sweden, 18-24 August, 2002.

[2] Statistics Netherlands, National Accounts Department, PO Box 24500, 2490 HA The Hague, The Netherlands; b.kazemier@cbs.nl.

1. Introduction

"*The provisional figures of Statistics Netherlands are systematically biased***"**[3]**.** This was the streamer of an article in the Staatscourant (a Dutch newspaper) of March 8th 1991. It was based on a research of the Netherlands Bureau for Economic Policy Analysis (CPB) (Van Vlimmeren et al., 1991). The main findings were that, for the period 1970-1984, the provisional estimates of the volume growth of Gross National Income and the volume growth of Private Consumption were, on average, underestimated by a ½ respectively 1 percentage point. Although things have improved after the general revision of the Dutch National Accounts in 1987, this article was an incentive for Statistics Netherlands to intensify its research on the reliability of its provisional estimates.

Quality defined

There are many definitions of the term "quality". The shortest definition is "quality is fit for use". This is clearly a definition from a users point of view. The quality of a product is directly related to the extent of which a product does what its owner expects it to do. According to this definition the technically superior Philips Video 2000 system, for example, was a bad quality product, as hardly anybody was prepared to buy it.

At the other end of the spectrum one can think of the definition "quality is fit for purpose". Here the quality of a product is determined by its technical state. A product is of good quality when it does where it is made for. According to this definition the above-mentioned Philips Video 2000 system was a high quality product, despite its bad sales.

The above definitions of quality do not suffice when it comes into implementation. Then we have to define the different attributes that make up for the quality of a product. In the context of statistics, the most frequently mentioned attributes are "*accuracy*" and "*reliability*". See for example Australian Bureau of Statistics (1983), Young (1987) or Bloem and Khawaja (2001). "Accuracy" refers to the extent to which the final estimates of statistics describe reality. "Reliability" is the extent to which provisional estimates predict the final estimates. Therefore provisional estimates are of good quality if they predict the final estimates well, even if the final estimates are wrong.

During the last decade, the interest in the quality of statistics has increased rapidly. This is mainly due to the increasing importance of statistics in national and international policy, especially within the European Community. Another reason is the increasing emphasis of the (inter)national statistical offices to better serve the needs of their customers. This leads to the identification of a number of additional quality attributes. Eurostat (Franchet, 1998) identifies *"relevance"* of statistical concepts, *"accuracy"* of estimates, *"timeliness"* and *"punctuality"*, *"accessibility"* and *"clarity"* of the information, *"comparability*" of statistics, *"coherence*" and *"completeness"*.

Statistics Netherlands (De Vries and Van Brakel, 1998) add the *"cost-effectiveness"* of the production of statistics and the *"survey-burden"* for data providers. These attributes do not directly apply to the statistics themselves, but to the compilation process of these statistics. They give an indication of the efficiency of the production process. One can say, they are indicators of "value for money".

---

[3] "In voorlopige cijfers CBS is sprake van systematische onderschatting", Staatscourant, 8 March 1991.

Most of the quality attributes, mentioned above, are affected by, or even relate to the production process. This is especially the case for timeliness, punctuality, completeness, cost-effectiveness and survey-burden. Moreover, if statistical co-ordination of classifications and definitions is also considered as part of the production process, comparability and coherence fall into this category as well. This offers good holds for quality control.

Quality control

Quality control consists of at least four parts. First, for all final products, one must define minimum standards for each of the identified quality attributes. A prerequisite is that all these standards can be quantified and measured. In the case of statistics, one must define minimum levels for accuracy, reliability, consistency, timeliness et cetera. To assure that the statistics fit for use, setting these standards is preferably done in dialogue with the users of the statistics. Second, one must determine the different stages in the production process: for example data collection, data editing, the different stages in the compilation of statistics, tabulation and publication. These stages require a full inventory once in a while, and regular incremental updates in between.

Third, one needs a system of assessing the production process, in terms of efficiency, effectiveness, and the extent to which the different stages contribute to or derogate the quality of the final product. This can be done best by deducing quality requirements for the outputs of each production process, given the quality requirements of the final products at the very end of the production chain. Statistical auditing, as introduced at Statistics Netherlands (De Vries and Van Brakel, 1998) fits here.

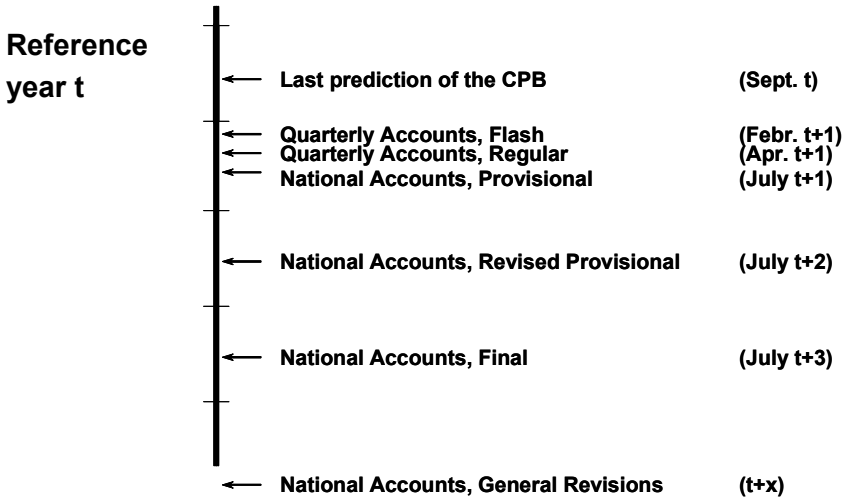
Finally, one needs a system to assess the quality of the final products. For an assembly line this can be easily achieved by physically testing a random sample of the final output. If that sample fails the quality-tests, the whole lot is rejected and replaced by the next lot. However, such a quality assessment is not appropriate for statistics. There is no replacement for statistics that do not meet the pre-set quality standards. The only solution is to do the compilation again, which is not a guarantee for improvement. Moreover, certain quality attributes can only be assessed after the statistics have already been published, for example reliability, accuracy, accessibility and clarity.

The next paragraphs describe the system Statistics Netherlands is implementing to assess the reliability of the national accounts estimates, and present the first results.

2. Reliability versus timeliness

Two of the most important figures in (inter)national economic policy are Gross Domestic Income (GDP) and Gross National Income (GNI). These figures act as the main indicators for economic performance and play a crucial role in the European Economic and Monetary Union (EMU). Furthermore, already 43 percent (in 2002) of the total budget of the European Community is based on the level of GNI. Therefore, many users of these statistics welcome an estimate of the size of at least one of these indicators as early as possible. Figure 1 shows a time-axis with the many subsequent (provisional) estimates of GNI in the Netherlands, starting with the last prediction of this variable by the Netherlands Bureau for Economic Policy Analysis (CPB), before Statistics Netherlands take over.

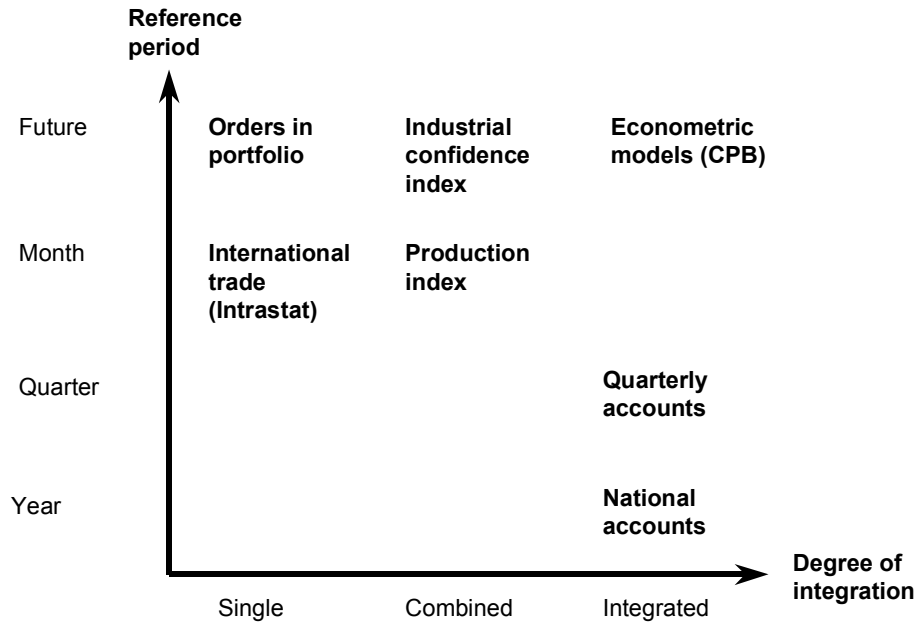
Figure 1.The subsequent estimates of GNI in the Netherlands



As can be seen from figure 1, Statistics Netherlands' first official estimate of GNI is published in the second month after the reference year. Then, within a period of five months, this figure is revised two times, each being considered more reliable then the previous one. But, the figure published in July of year t+1 is still not the final estimate. It takes at least two additional years to arrive at sufficiently accurate figures to 'seal the books'. There is a trade-off between timeliness and reliability.

The main reason that it takes about two and a half years to finalise the national accounts is that it is a highly integrated set of statistics, subjected to a complex set of definitions and constraints. These definitions and constraints are known as the System of National Accounts (SNA, 1993) and the European System of Accounts (ESA, 1995). The national accounts can only be finalised after the last bit of information needed to compile these statistics sufficiently accurate, becomes available. Figure 2 visualises timeliness and reliability in relation with the reference period of statistics and the degree of integration (see also Algera, 1995). Timeliness goes from top-left to bottom-right. Reliability goes the other way round.

Figure 2. The relation between the reference period of statistics and the degree of integration



The final national accounts are compiled using almost all data available, and are subjected to the most well defined and consistent set of definitions and constraints. Therefore, these statistics are considered to be the most accurate description of reality. To be consistent over time in its description of reality, all other statistics that describe (some of) the same phenomena should be as close as possible to the final national accounts estimates. This leads to criteria that played an important role in the introduction of the flash-quarterly accounts (Reininga et al., 1992; Reininga and Kazemier, 1998).

3. The accuracy of the final national accounts

The final national accounts estimates should describe the economy as accurately as possible. However, there is no way to measure the distance between these figures and reality. Most of the final figures can not be validated. Moreover, some can't even be directly observed.

Statistical indicators, like variance and skewness, can be of help. Assuming that, on average, a statistic describes reality rather well, variance and skewness are indicators of the likelihood, the size and the sign of a deviation. Stone et al. (1942) developed a method to arrive at national accounts estimates with minimal variance. This method has been extended and refined by a number of authors, amongst others by Byron (1978) and Weale (1988), Solomou and Weale (1993), Sefton and Weale (1995) and Magnus et al. (2000). A recent application of this method (Wroe et al., 1998) shows that combining two ore more estimation-approaches (production-method, expenditure method, and income-method) can increase the accuracy of the estimates of GNI a lot. The same is true when GNI in current prices and GNI in constant prices is compiled simultaneously. On the other hand, putting more detail into the compilation process adds only rather little. According to this study the standard error of the final estimate of Dutch GNI is about 0.15 percent.

Variance, however, tells only part of the story. Very often non-sampling errors, which are not covered by these indicators, are more significant. This is best illustrated by the results of occasional general revisions, which very often lead to substantial upward adjustments to GNI. Up till now, a lot of research has been done to arrive at a quantitative quality-score of more or less integrated statistics, which takes non-sampling errors into account. See for example Renssen et al. (1998) for an application on business surveys, and Calzaroni and Puggioni (1995) for an application on national accounts statistics.

The crucial assumption is that on average the final national accounts describe reality very well, that it is not biased, and that the impact of all kinds of (non-)sampling errors can be quantified. This assumption needs to be proved, or at least made plausible. For example, it is not necessarily true that all definitions are correctly implemented or that all relevant income generating transactions are fully covered in the national accounts. In this respect one may think of the inclusion of hidden and other unobserved activities, which are difficult or even impossible to observe (OECD, 2002). To be able to judge the accuracy of statistics, especially rather complex statistics such as the national accounts, it is necessary to judge its completeness or exhaustiveness. For that a minimum requirement is a transparent and well-documented compilation process, possibly supplemented by a system of internal or external auditing.

4. The reliability of the provisional national accounts

4.1 Quality control

In contrast to the final national accounts, the provisional accounts can be validated. Ideally, the latter equal the former. This enables the implementation of a system of quality control or, to be more precise, a system of reliability control:

- Determine the variables that are relevant for the assessment of the reliability of the provisional estimates. In general these are the value (growth) and volume (growth) of the macro-indicators that can be derived from the national accounts, for example GNI, compensation of employees, operation surplus and mixed income, household and government final consumption, exports and imports of goods and services, disposable income and net savings. One may also think of all kind of ratio's that play a role in (inter)national politics, like government deficit and government debt as percentage of GNI, or total investment as percentage of GNI or as percentage of operation surplus and mixed income.
- The next step is to decide upon a minimum reliability-standard, i.e. a maximum distance between the assessed estimate and the reference estimate, in general the final one. Standards can be either relative or absolute. An obvious minimum standard is that each provisional estimate should be better than the previous one. In terms of figure 1, the provisional July t+1 national accounts should perform better than the April t+1 quarterly accounts, and all estimates of Statistics Netherlands should be better than the CPB's estimates.
- Then, the assessment takes place, preferably as soon as possible after the compilation of the final estimates is completed. Variables that do not meet the pre-set standards need further examination. Notice that not the final estimates are assessed, but the earlier published provisional estimates. So there is always a (sometimes substantial) time-lag between the production of provisional accounts and its assessment.

The reason for not meeting the standards can be twofold. The first and most likely reason is that the statistics used to compile the provisional estimates show a pattern which later turns out to be misleading or even wrong. If that happens too often, it may be an indication that the reliability of these statistics does not meet the required quality or that its quality deteriorates. For example, this has happened with the international trade statistics (Intrastat), directly after the deregulation of the internal EC trade in January 1993. Communicating these findings to the people who are in charge of those statistics, is the appropriate action.

The other reason can be the compilation of the provisional estimates, itself. The treatment or choice of the basic statistics to arrive at provisional national accounts can be wrong. This is the most likely reason when provisional estimates show a systematical bias. Now, improving the compilation process is the appropriate action, for example changing the process planning or changing to alternative basic statistics as input for the process.

4.2 Setting reliability standards

In setting reliability standards several aspects must be taken into account. Very important is history. History may show whether there is a systematic bias or not. Further, history is needed to determine trend-breaks. An obvious reason for a trend-break is a significant improvement of the compilation process achieved at a general overall revision of the whole system. Another reason is a sudden change of the quality of the used basic statistics, like the one in the international trade statistics, which, as already is explained above, was caused by an external factor that fell outside the competence of the national statistical offices. Finally, history can be used to arrive at reasonable reliability standards. When history shows that on average the distance between a provisional estimate and the corresponding final one is, for example, 10 percent, then a minimum standard of 1 percent is not reasonable, at least not in the short run. While setting standards, the chance that these standards are met must be taken into account.

Another aspect that determines the required minimum reliability standard is “proud” and “competition”. An example of this is the already mentioned requirement that all provisional national accounts, compiled by Statistics Netherlands, perform better than the predictions of the Netherlands Bureau for Economic Policy Analysis (CPB). But it can also be that it is decided to be more ambitious than necessary, to create some margin in case time gets worse.

Last but not least, the opinion of the main users must be taken into account. Important users of the Dutch national accounts data are the central government (Ministry of Finance including the Central Bank, Ministry of Economic Affairs, Ministry of Social Affairs and Employment), the ECB, the European Union because of the fourth resource, the European Economic and Monetary Union (EMU) because of the Stability Pact, other international bodies, the CPB, universities and other scientific institutions.

4.3 Reliability tests

From the discussions above it can be derived that at least three statistical tests are needed. Firstly, trend-breaks must be identified, both in the level and in the variance of the variables assessed. In the software, developed in the framework of a national accounts reliability-control system, this has been implemented in such a way that the most likely places for trend-breaks are identified and reported. A window of fixed width, for example eight years, is spread out over the time-series under consideration. Then it is assumed that exactly in the centre, a trend-break has occurred. This is tested using the traditional t-test for the equality of two sample means, respectively an F-test for the equality for two sample-variances. Both tests can be found in almost all handbooks on statistics. If this test leads to not-rejecting the assumption this possible candidate for a trend-break is put into portfolio. Then the window is moved on one observation and the whole procedure starts over again. After having scanned the whole time-series, the test-statistics of all possible candidates are compared with each other. Only those that are more likely than their neighbour candidates are reported.

Next, the significance of the average difference between the provisional and reference estimate is tested. This is simply done by testing whether the average distance between both estimates significantly differs from zero or not. If it does, the sign of the difference determines whether there is a systematic downward or upward bias.

Finally, it must be possible to judge whether a pre-set reliability standard is reasonable or not. Calculating the chance that the absolute distance between the provisional and the reference estimates exceeds the standard can do this. Assuming that the series of differences between the two time-series follow a Normal distribution, as is also assumed for the tests mentioned above, this can easily be calculated.

4.4 Setting reliability-standards in practice

In practice, there are two ways, to arrive at reasonable reliability standards. The first one is to go through all variables, selected for assessment, and decide explicitly on the standards, keeping an eye on the chance that these standards will not be met. This can easily be done, if the number of variables is relatively small. However, as the number of variables is large, it is virtually impossible to do it this way, keeping the set of standards in balance.

The other way is to use a rule of thumb for the majority of the variables, and only decide explicitly on a very small number. Such a rule of thumb is implemented in the software mentioned above, by means of a small number of general rules:

- The first thing one needs to decide upon is the chance that a standard is not met: a kind of hit-rate. This hit-rate should not be too high. If, in the ultimate case, a variable fails its reliability-test each year, then it is very likely that each year the investigations into the reasons of not meeting the standards yield no results. Further, if the standards for all variables are set this high, the additional investigations more or less duplicate the final reliability-checks, which are part of the regular compilation process. A reasonable hit-rate is 0.2, which means that on average, a variable does not meet the pre-set standards once every five years. Even this once-every-five-years hit rate will put a considerable additional burden onto the whole compilation process, especially if the number of variables assessed is quite high.
- Next, one has to decide upon the length of the period that is used to calculate the default-standard. A long period yields a more accurate estimate of the hit-rate. On the other hand, too long a period means that all (gradually) introduced improvements of the production process are not taken into account. An appropriate period is 5 to 10 years. Of course, the results of the year assessed are not included in the period. Another way to determine the length of the period is by restricting it to the period since the last trend-break. In this way, occasionally introduced major improvements, for example at general overall revisions, or major changes in the quality of the basic data, are taken into account.

After having calculated the standards in an automatic way one can go through the list of resulting standards and only decide whether these standards are acceptable or not. For example, because of their customers Statistics Netherlands will never accept a default standard for the growth of Gross National Income, that exceeds 0.5 percent. If the default is higher, it will be overridden.

Besides history based rules, one might think of rules based on the relative importance of the variable, here defined as its contribution to GDP. The general form of such a rule is $S_i = f(R_i) * S_{GDP}$: the reliability standard S_i for variable i is a function (f) of the share (R_i) of variable i in total GDP times the pre-set reliability standard (S_{GDP}) for GDP. Three such rules will be used for comparison in chapter 5:

[A] $$S_i = \frac{1}{R_i} * S_{GDP}$$

[B] $$S_i = \sqrt{\frac{1}{R_i}} * S_{GDP}$$

[C] $$S_i = Ln\left(1 + \frac{1}{R_i}\right) * S_{GDP}$$

4.5 Implementing standards in the production process

To allow for a systematic and regular reliability-check of the provisional national accounts, a computer program was developed. With this program it is possible to inspect time-series individually by means of the statistical indicators, described above, as well as visually by means of plots of the time-series.

Furthermore, the program can be used as a management information tool, for instance in the preparation of the publication of the estimates, to anticipate at least part of the questions users will pose.

Monitoring the revisions is not a goal in itself. First of all, the staff members who work on these figures must be aware of the findings. In case of violations of the standard there should be a follow-up. Research should be done to find out which underlying estimates are responsible for the revisions and how future estimates can be improved. Secondly, there should be some kind of communication to the users of the provisional National Accounts figures. The users have the right to know how reliable the provisional figures have been in the (near) past so they can take this information into account. The revisions observed however cannot be interpreted as an indication or a kind of guarantee for future figures.

5. A case: the preliminary Dutch national accounts

5.1 The reliability of successive estimates

The main reason for publishing the very early quarterly accounts based estimates is that the main users of the statistics ask for them. The main incentive to revise them so often is that within a period of about six months reliability can significantly be improved. This is checked by investigating the performance of the provisional estimates of the volume growth-rates of the main macro-economic aggregates: GDP, imports and exports of goods and services, household consumption, government consumption and gross fixed capital formation of government and enterprises. Table 1 summarises the revisions of the corresponding statistical indicators. Most indicators show that the successive estimates improve.

Table 1.Revisions between provisional and final estimates, 1991-2000

	Average revision				Average absolute revision			
	Flash (February t+1)	Quarterly Accounts 4th quarter (April t+1)	National Accounts provisional (July t+1)	National Accounts provisional (July t+2)	Flash (February t+1)	Quarterly Accounts 4th quarter (April t+1)	National Accounts provisional (July t+1)	National Accounts provisional (July t+2)
Gross Domestic Product (mp)	0.35 [1]	0.31 [1]	0.24 [1]	0.19 [2]	0.45	0.41	0.30	0.25
Household final consumption		0.27	0.21	0.26 [2]		0.43	0.33	0.28
Government final consumption		0.44	0.21	0.08		0.94	0.61	0.48
Gross fixed capital formation of enterprises		-0.16	-0.72	-0.13		1.04	1.26	0.65
Gross fixed capital formation of government		0.83	1.25	0.69		1.99	2.53	2.47
Imports of goods and services		0.57	-0.05	-0.21		1.31	0.99	0.51
Exports of goods and services		1.04 [1]	0.56 [2]	-0.21 [2]		1.08	0.74	0.31

1) Significantly different from zero at a 99% level
2) Significantly different from zero at a 95% level

It can be seen from the table that in general the provisional estimates underestimate economic performance. Most revisions are upwards. All average revisions of the GDP and the export of goods and services significantly differ from zero, as well as the July t+2 average revision of household consumption.

5.2 Setting standards

In section 4 it was mentioned that, in first instance, applying a rule of thumb is the best way of setting reliability standards in practice. This is especially the case when the number of variables assessed is quite large. The rule, implemented in the software tool is one that is based on the history of the time-series. Table 2 shows what three variants of this history-driven rule would do for the 2000 quarterly national accounts estimates. The three variants are:

1. The hit-rate (chance that an estimate does not meet the norm) is 0.25, i.e. on average once every four years. The length of the period used to calculate the default standard is 5 years. Trend-breaks are not taken into account.
2. The hit-rate is 0.2, i.e. on average once every five years. The length of the period used to calculate the default standard is 5 years. Trend-breaks are not taken into account.
3. The hit-rate is 0.2. The length of the period used to calculate the default standard is the whole period since the most recent trend-break, with a maximum of 10 years. Trend-breaks are calculated, using a time-window of 8 years (see section 4), with a minimum distance between two successive trend-breaks of 3 years, and a minimum significance of 95 percent.

Table 2. Alternative default reliability standards

	History-based standards			Contribution-to-GDP-based standards[1]					
	1. Hit-rate: 0.25; Period: 1995-1999	2. Hit-rate: 0.20; Period: 1995-1999	3. Hit-rate: 0.20; Period: 1990-1999, and since the last trend-break	A. Si = (1/Ri) * Sgdp		B. Si = V(1/Ri) * Sgdp		C. Si = Ln(1+1/Ri) * Sgdp	
Gross Domestic Product (mp)	0.46	0.50	0.61	0.50		0.50		0.50	
Household final consumption	0.70	0.78	0.73	1.00	(0.10)	0.71	(0.25)	0.55	(0.37)
Government final consumption	1.00	1.12	1.42	2.20	(0.01)	1.05	(0.23)	0.84	(0.34)
Gross fixed capital formation of enterprises	1.62	1.79	1.51	2.63	(0.05)	1.15	(0.43)	0.92	(0.53)
Gross fixed capital formation of government	3.90	4.35	3.93	16.13	(0.00)	2.84	(0.40)	1.75	(0.61)
Imports of goods and services	2.19	2.43	2.02	0.80	(0.68)	0.63	(0.74)	0.48	(0.80)
Exports of goods and services	2.01	2.20	1.73	0.74	(0.71)	0.61	(0.76)	0.45	(0.82)

1) Corresponding hit-rate between brackets

For reasons given earlier, a standard based on a hit-rate of 0.2 or less is to be preferred. This leaves rules 2 and 3 that only differ in the length of the reference period. As can be seen from the table rule 2, which is based on the shorter period, yields stricter standards for GDP and government final consumption. In all other cases the longer period leads to stricter standards. This shows that for these variables the reliability of the quarterly estimates have decreased over time. Nevertheless, as GDP is considered as the main indicator of economic performance, rule 2 is preferred. Coincidentally, the default standard for GDP equals Statistics Netherlands own standard of 0.5 percent.

For comparison three alternative rules are included in table 2, based on the final figures for 2000. The reliability standard for GDP (S_{GDP}) is fixed on 0.50 percent

point. The corresponding hit-rates, based on the observations in the period 1995-1999, which is the same period that is used for rule 2, is added between brackets.

Although rules A, B and C may yield more balanced standards with respect to the relative importance of a variable, the results in terms of hit-rate are not balanced at all. Moreover, this kind of rules can not be applied to variables that (in terms of table 1) are not expressed in current or constant prices, for example labour force or government deficit ratio, and which units of measurement are not growth rates, but shares. For this reason this kind of rules are rejected.

5.3 Revisions of provisional national accounts

In this section the reliability of the provisional annual national accounts estimates are analysed. To allow for a comparison over time, the figures for 1970-2000 are presented. In this period, the national accounts have been significantly revised and re-based three times: in 1977, 1987 and the most recent ESA '95 revision, see CBS (1985, 1992, 1999). The main reason for such general overall revisions is that the figures in the National Accounts have to meet at least two requirements:

1. They must equal the actual level of the phenomena described.

2. They must be comparable over time.

These two requirements can not always be met simultaneously. When both requirements conflict, the requirement of comparability over time gets priority and the requirement of actuality is slightly violated. General revisions are meant to correct for deviations from actuality. Further, general revisions are used to introduce changes in definitions and major improvements in the compilation process.

These revisions, however, affect the analysis of the reliability of the provisional estimates. When a revision takes place with base-year T, the results become available in T+4 (the results of the general revision with base-year 1977 became available in 1981, those of 1987 in 1991 and those of 1995 in 1999). For the years in between, the provisional estimates of T+1 and T+2 and the second provisional estimate of T+1 are pre-revision based. In the analyses below, they are compared with the post-revision final estimates. Differences between these two estimates can be real, but can also be caused by the general revision.

5.3.1 Gross Domestic Product

The revisions between the provisional national accounts estimates (July t+1 and July t+2) and the final estimates do not affect public opinion on the quality of the estimates as much as the revisions of the first estimates do. Nevertheless, they are still important as the analysis of these differences may, at the end, lead to improvements of the quality of both the provisional and the final estimates. Figure 3 presents a graph of the revisions between the provisional and the final national accounts estimates of GDP-growth.

Figure 3. Revisions between provisional and final national accounts of GDP volume-growth (%), 1970-2000

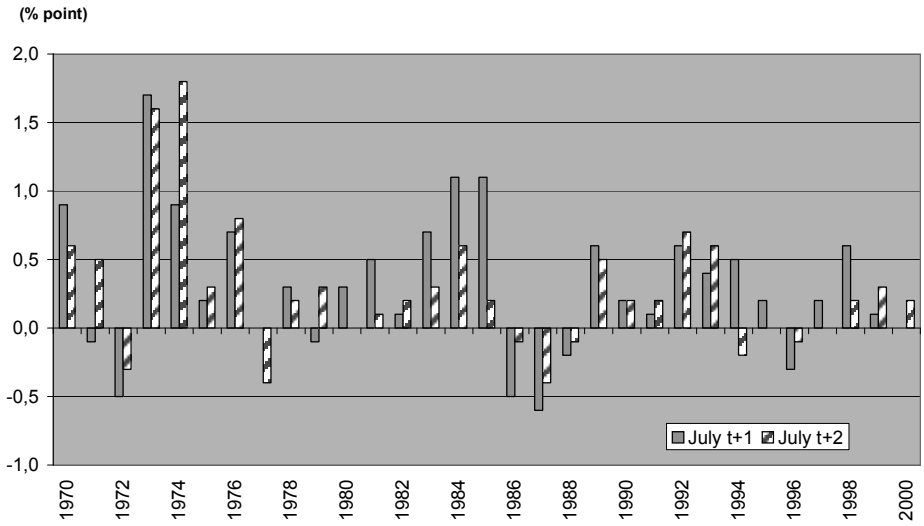


Figure 3 clearly shows that the pre-set reliability standard of 0.5 for GDP is quite strict. Since 1970, that standard was not met 11 times (35 percent) for the July t+1 estimates and 7 times (23 percent) for the July t+2 estimates. The most severe violations of this standard are concentrated in the seventies and the mid-eighties. Further it can be seen that most violations were underestimates.

The average revisions for the period 1970-2000 as a whole are 0.31 for the July t+1 estimates and 0.20 for the July t+2 estimates. Both significantly differ from zero, which implies that the provisional estimates significantly tended to underestimate economic growth. This confirms the complaints of the Netherlands Bureau for Economic Policy Analysis (CPB) formulated in 1991 (see introduction) and recently repeated in "Centraal Economisch Plan" (CPB, 1999).

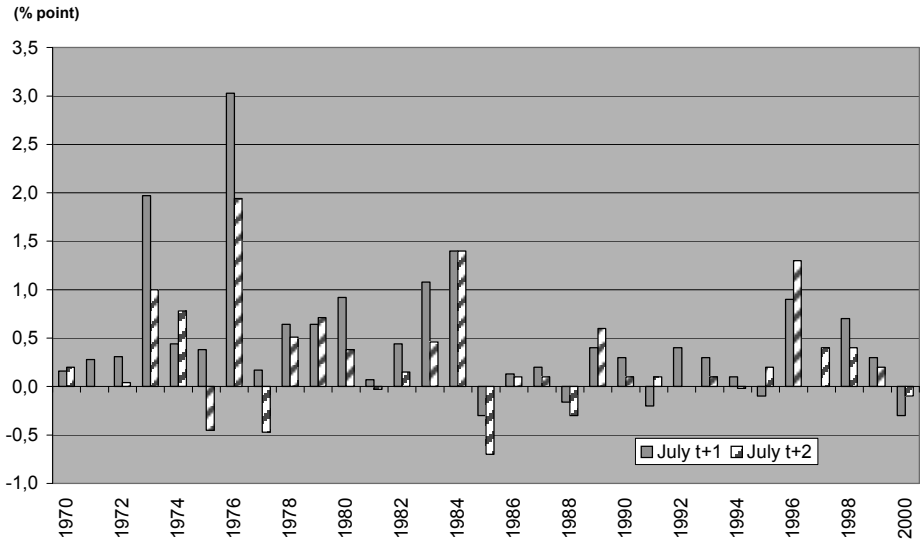
A significant trend-break is found round 1986, which is most likely caused by the general revision of 1987. Due to the fact that the work on the 1987 general revision was already in progress at the time the 1986 final estimates were compiled, the 1986 final estimates were of less quality than usual. If one compares the post-revision period 1988-2000 with the pre-revision period 1970-1987 one may conclude that the reliability of the provisional estimates has improved. The average revisions have changed from 0.37 (July t+1) respectively 0.35 (July t+2) in the earlier period to 0.23 and 0.19 respectively afterwards. However, these differences still significantly differ from zero. The average absolute difference fell from 0.57 and 0.48 respectively in 1970-1987 to 0.31 and 0.25 in 1988-2000.

On average the second provisional estimates are better than the first ones, but still in 12 cases out of 31 it is the other way round.

5.3.2 Household final consumption

Until the general revision in 1987, household consumption was treated as a balancing item. Together with changes in stocks it was calculated as the difference between total production and imports on one hand, and exports, intermediate consumption, government consumption and investment on the other. The estimates, both provisional and final, of household consumption were not the most reliable and accurate ones. This is clearly shown in figure 4.

Figure 4. Revisions between provisional and final national accounts of volume changes (%) in household final consumption, 1970-2000



Household consumption growth was significantly underestimated. The average absolute distance between the provisional estimates and the final one was 0.70 (July t+1) respectively 0.52 (July t+2). Severe underestimates did occur in 1973, 1976 and 1984. Since the final estimates of 1987 (and the provisional estimates of 1990) household consumption is explicitly estimated. The main sources for the estimates are the household budget surveys and the retail trade statistics (CBS, 1992; Buiten, 1993 and 1994; Bos and Gorter, 1993). This change significantly improved the quality of the figures.

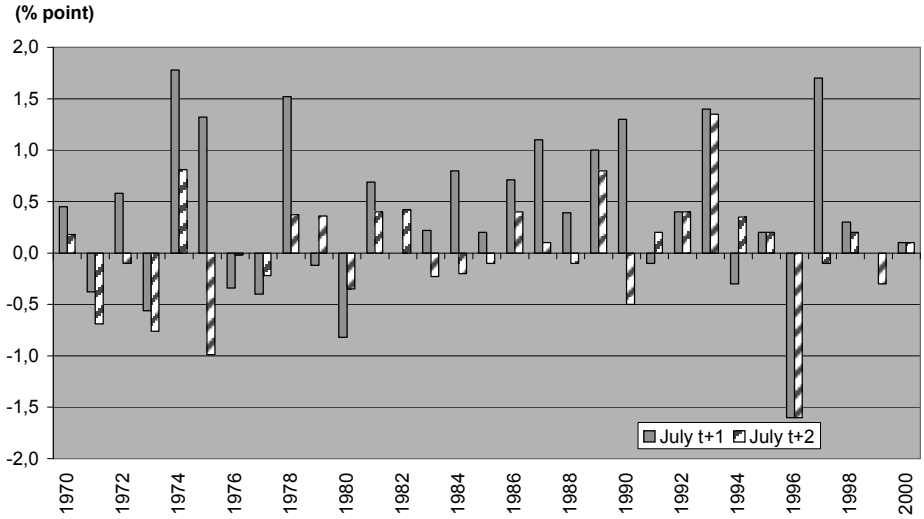
The big revisions in 1996 are a result of the general revisions, ESA'95. The July t+1 and t+2 estimates were made before the general revision, the final estimate afterwards. During this general revision, a large shift between government and household consumption was made resulting in large revisions.

5.3.3 Government final consumption

The first provisional estimates of government final consumption are systematically biased, see figure 5. This is mainly due to the government final accounts not being available that soon. The main source for the provisional estimates is the government budget. The fact that the budget tends to underestimate realisation significantly is not used in the production process of the statistics and is therefore reflected in the estimates.

The big revisions in 1996 are a result of the 1995 general revision, see the remarks under household final consumption.

Figure 5. Revisions between provisional and final national accounts of volume changes (%) in government final consumption, 1970-2000



5.3.4 Gross fixed capital formation

Figure 6. Revisions between provisional and final national accounts of volume changes (%) in government gross fixed capital formation, 1970-2000

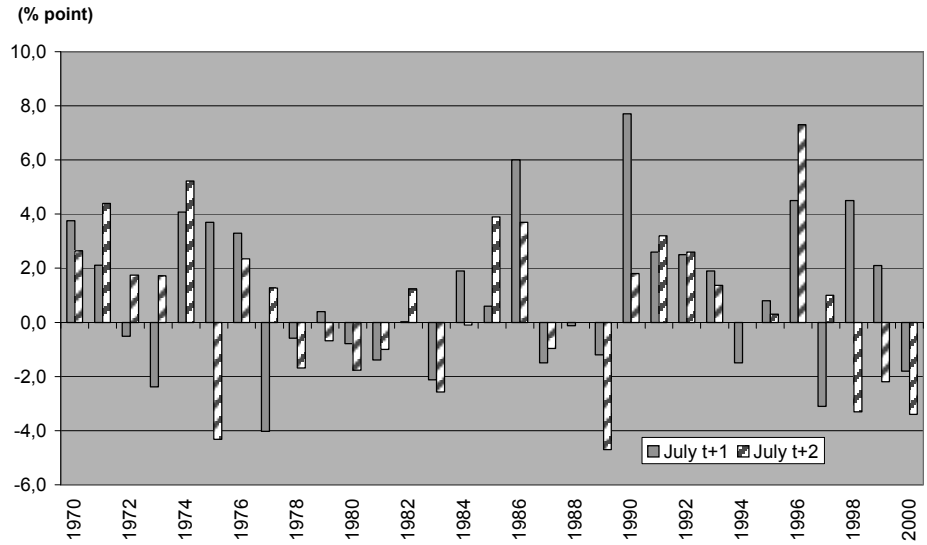
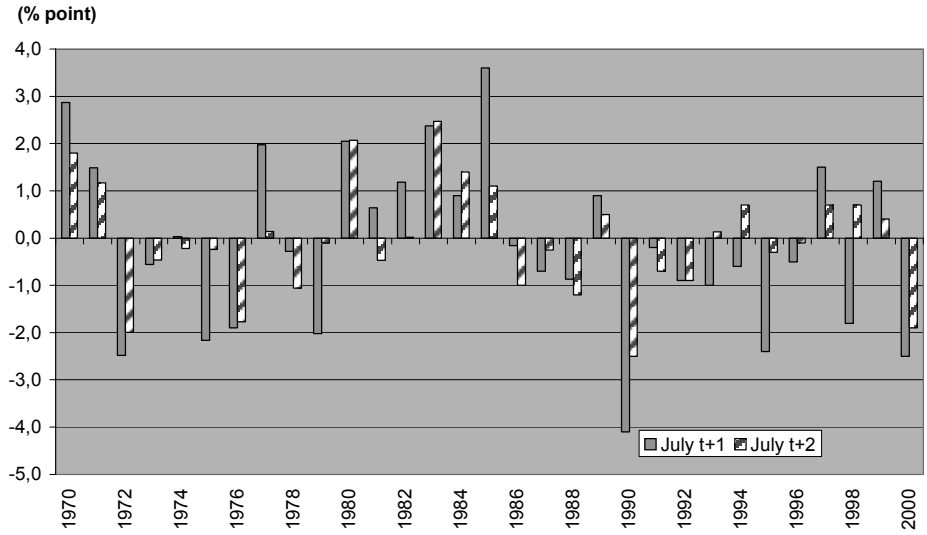


Figure 7. Revisions between provisional and final national accounts of volume changes (%) in gross fixed capital formation of enterprises, 1970-2000



The final estimates of the volume growth of government gross fixed capital formation are very difficult to predict, as can be seen in figure 6. The revisions between the provisional and the final estimates range from –4.7 to +7.7 percentage-points. The reason is twofold. Firstly, similar to the provisional estimates of government final consumption, the final government accounts are not yet available when the provisional national accounts are compiled. And with respect to gross capital formation, the discrepancies between intentions and realisations are even bigger. It very often happens that investments are actually realised one or more years later than initially intended or even postponed. On the other hand, investments often turn out to be more expensive than originally accounted. The second reason is that total government gross capital formation is a relatively small amount. Therefore,

even modest adjustments made to the originally published figures can be relatively large. The biggest revision of provisional figures occurred in 1990 and in 1996. These provisional estimates wouldn't have passed the reliability check. The revision in 1996 is a result of the 1995 general revision. The 'hit' in 1990 has its counterpart in the estimates of the volume growth-rate of gross fixed capital consumption of enterprises, see figure 7. In 1990 there was an allocation problem with respect to the gross fixed capital formation.

5.3.5 Imports and exports of goods and services

Figure 8. Revisions between provisional and final national accounts of volume changes (%) in the imports of goods and services, 1970-2000

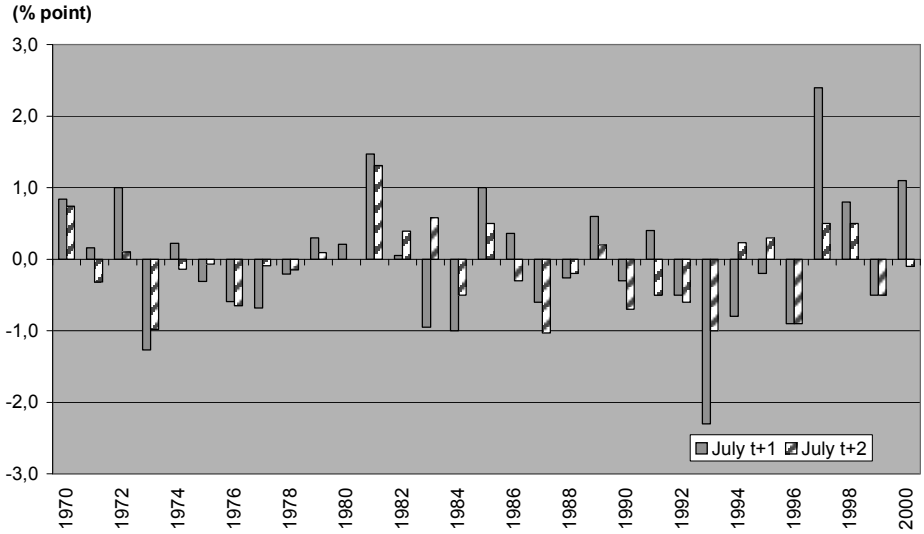
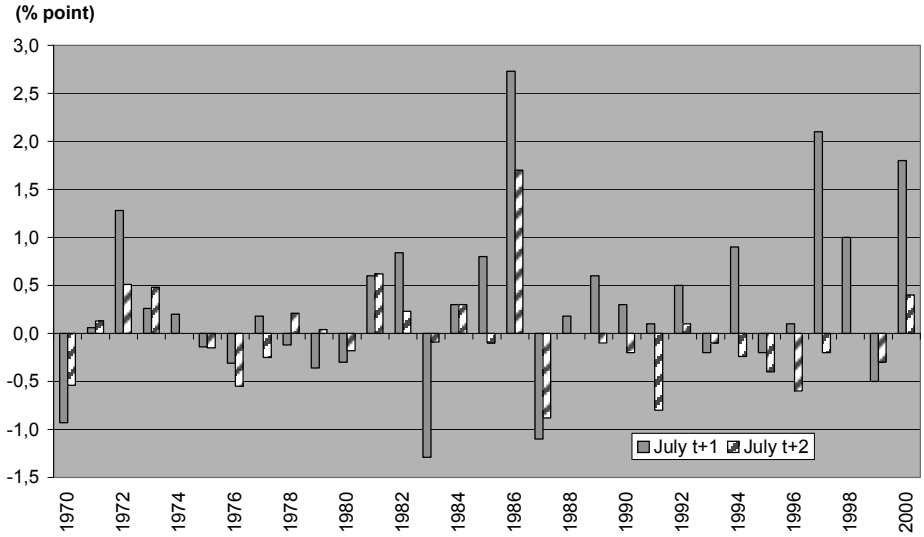


Figure 9. Revisions between provisional and final national accounts of volume changes (%) in the exports of goods and services, 1970-2000



Figures 8 and 9 show the differences between the provisional and the final estimates of imports and exports volume growth. Up till 1992 the international trade statistics were considered to be very reliable. As a consequence they were hardly changed during the compilation of the national accounts. Differences between provisional and final estimates were mostly due to differences in the completeness of the international trade statistics. As a result the provisional figures on imports and exports were very reliable. Consequently, the default reliability standards were rather strict and didn't change very much over time. For imports of goods and services the average standard in this period was approximately 0.90 for the first and 0.75 for the second provisional estimates, for exports 1.00 and 0.56 respectively. There would have been 13 'hits' (14 percent), all series taken together. The most

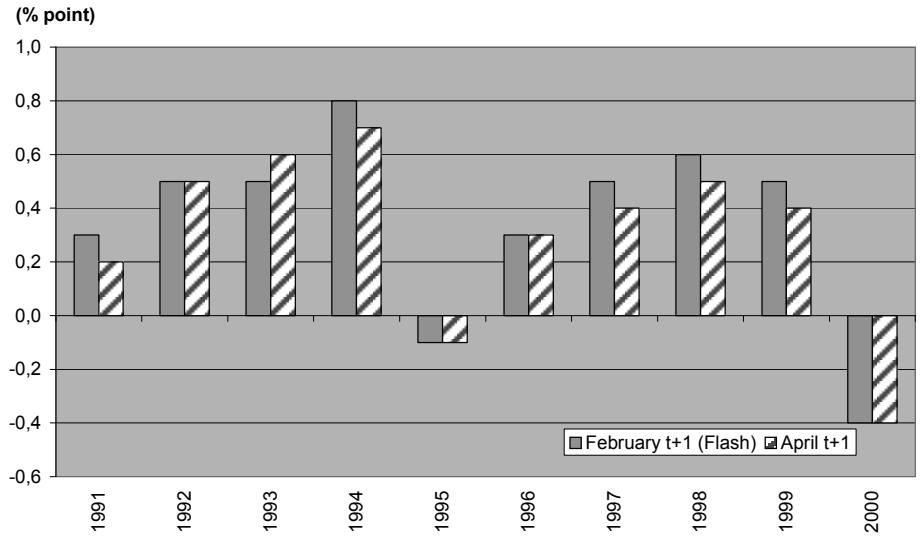
severe were in 1981 for both estimates and in 1987 only for the second provisional estimate of imports. For exports the most severe hits were in 1986 for both.

Since 1993 things have changed. Because of the liberalisation of the intra-EC trade, the custom forms, which were the bases for the international trade statistics, had to be replaced by a normal questionnaire. This has made it especially more difficult to estimate the size of total imports: in contrast to exports, the number of companies that imports a sometimes very small amount of goods and services is rather high. This is a common problem for all statistical institutes within the European Community. For the EC as a whole, total exports exceed total imports by a few percent, where it should be equal. In compiling quarterly and national accounts, it took a year to solve the problem.

5.4 Revisions of the flash and quarterly accounts of the Gross Domestic Product

Since 1991 Statistics Netherlands compiles so called flash estimates on GDP. These flash estimates are published in February t+1. The revisions between these flash estimates and the final estimate are shown in figure 10, together with the corresponding revisions of the quarterly accounts estimates.

Figure 10. Revisions between the flash and quarterly national accounts of GDP volume-growth (%) and final national accounts of GDP volume-growth (%), 1991-2000

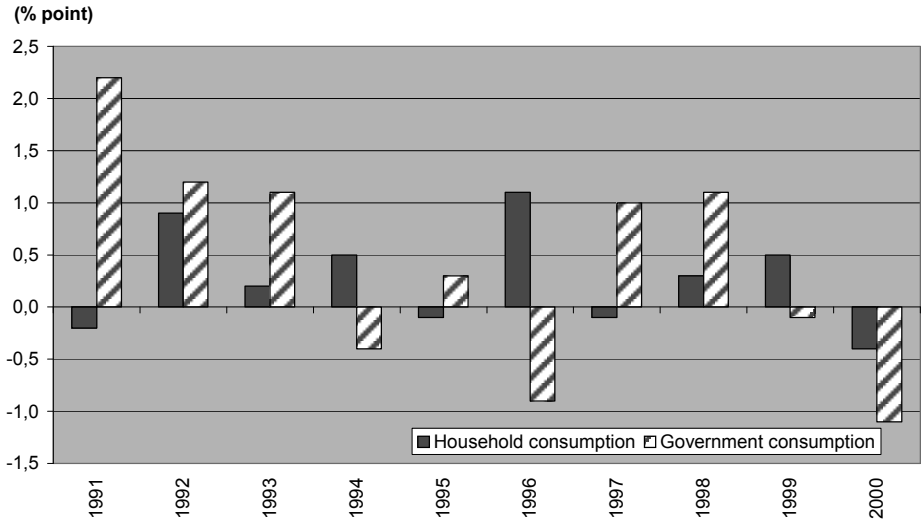


As mentioned in table 1 the average revisions of the flash and the quarterly accounts in the period 1991-2000 are 0.35 and 0.31 percent point respectively, and differ significantly from zero at a 99% level. Only in 1995 and in 2000 the first estimates proved to be too high.

The data for the underlying variables are only available since 1995. These time series are too short for a serious analysis. Therefore the analysis given below is done on quarterly accounts. In general, revisions between the flash and the quarterly estimates are rather small, so the analysis on the quarterly accounts to a large extent also holds for the flash estimates.

5.4.1 Household and government final consumption

Figure 11. Revisions between quarterly and final national accounts of volume changes (%) in household and government final consumption, 1991-2000



The revision of the quarterly estimates (April t+1) of household and government consumption are shown in figure 11. The average revision of the quarterly estimate of the final household consumption between 1991 and 2000 is 0.27 percent point. Although substantial, it does not significantly differ from zero at a 95 % level. An average revision of 0.30 percent point would be. The default standard, calculated for the year 2000 using rule 2 in section 5.3 is 0.78 percent point. This relatively large standard is a result of the major revision in 1996. This revision was mainly caused by the general major ESA'95 revision, which was not yet implemented in 1997 when the quarterly account estimates for 1996 were published. It was however implemented when the final estimate was compiled. See also the remarks under household final consumption in chapter 5.3. The 1997 quarterly accounts based estimate, however, was almost exact, while also here the major ESA'95 revision comes in between.

The average revision of the quarterly estimates of final government consumption between 1991 and 2000 is 0.44. As a result of a high standard deviation this does not significantly differ from zero at a 95 % level. The average absolute revision is much higher at 0.94 percent point. The default standard, calculated for the year 2000 using rule 2 in section 5.3 is 1.12 percent point. This standard was not met in the first two years of the period analysed. The large revision in 2000 (-1.1 percent point) is just inside the margin.

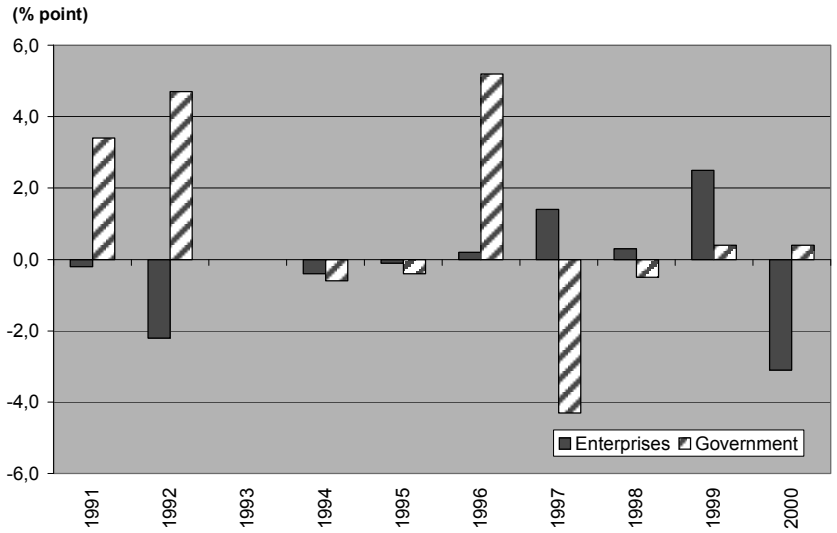
5.4.2 Gross fixed capital formation of enterprises and government

The average revision of the quarterly estimate of the gross fixed capital formation of enterprises between 1991 and 2000 is only -0.16. It is the only main variable that is, on average, over-estimated in the quarterly accounts. This small average revision, however, is misleading. The average absolute revision is much higher at 1.04 percent point. The default standard using rule 2 of section 5.3 is 1.79 percent point. This standard was not met in 1992 and in the last two years of the time series. In 2000, a large part of the revision was a result of an overestimation of investments in electronics.

The average revision of the quarterly estimates of the government gross fixed capital formation between 1991 and 2000 is 0.83 percent point. The default standard, calculated for the year 2000 using rule 2 is as high as 4.35 percent point. In 1998, 1999 and 2000 the revisions between the quarterly and the final estimates of government gross fixed capital were well inside this margin.

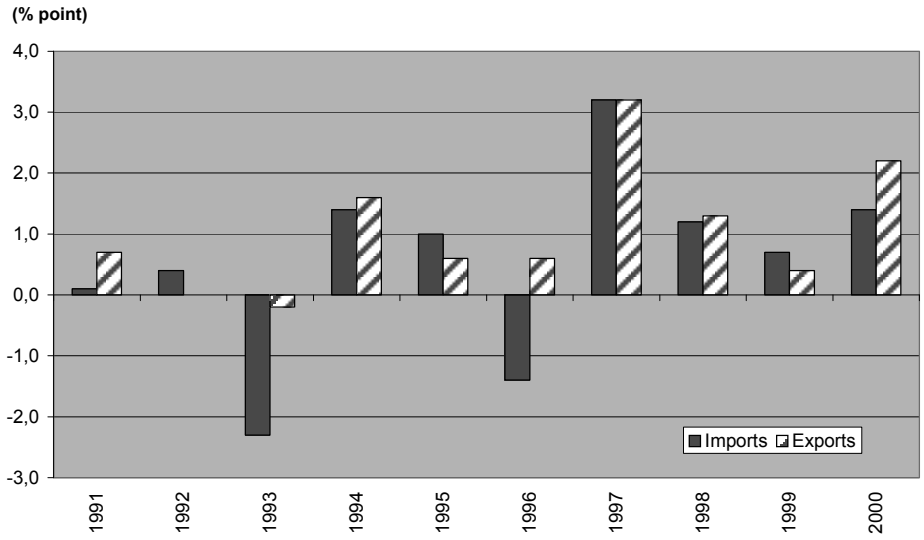
The revisions are shown in figure 12.

Figure 12. Revisions between quarterly and final national accounts of volume changes (%) in gross fixed capital formation of enterprises and government, 1991-2000



5.4.3 Imports and exports of goods and services

Figure 13. Revisions between quarterly and final national accounts of volume changes (%) in the imports and exports of goods and services, 1991-2000



In figure 13 the revisions of the quarterly estimates (April t+1) of the imports and exports of goods and services are shown. The average revision of the quarterly estimates of the imports of goods and services between 1991 and 2000 is 0.57 percent point. This does not significantly differ from zero at a 95 % level. The standard default, based on rule 2, is 2.19 percent point.

The average revision of the quarterly estimates of the exports of goods and services between 1991 and 2000 is 1.04 percent point. Hereby the exports of goods and services is significantly underestimated at a 99 % level. Since 1994, the exports are underestimated in the quarterly accounts without exception. As a result, the default standard, based on rule 2, is large at 2.20. The large underestimation in the year 2000 equals this standard default.

6. Conclusions

Quality is not a one-dimensional notion. Many different and sometimes conflicting aspects should be taken into account. With respect to statistics, quality has, at least, the following attributes: accuracy, reliability, relevance, timeliness, punctuality, accessibility, clarity, comparability, coherence and completeness. And as government generally pays the national statistical offices, the costs in terms of resources and survey-burden must also be taken into account. This paper deals with aspects of the quality of the national accounts.

National accounts statistics play an important role in (inter)national policy and economic science. All users weigh the many quality attributes differently, especially the attributes accuracy, reliability and timeliness. With respect to the fourth (income) resource of the European Community, accuracy and completeness are the main quality requirements. The Netherlands Bureau for Economic Policy Analysis (CPB) prefers reliability and comparability over time, while the EMU stresses the need for timeliness. To combine the requirement of accuracy with the wish for timeliness provisional national accounts estimates are published, soon after a year has finished, followed by several revisions up to the final estimates two years and a half later.

It is not only important to publish good quality statistics. It is also essential that the users of the statistics believe that they are of good quality. Therefore, it is necessary that the subsequent provisional estimates of the national accounts show a similar picture of economic performance. The subsequent estimates have to be sufficiently reliable.

In this paper, it is analysed to what extent this requirement is met, taking into account that this requirement seriously conflicts with timeliness. A computer program was developed to enable a quick reliability-check to a large number of economic variables, derived from the national accounts. To conduct this test, a minimum reliability standard is defined, with respect to the difference between an estimate and the final estimate. In many cases these checks will signal one-time events, but sometimes they will signal latent problems in either the basic statistics, underlying the national accounts, or the compilation (process) of the national accounts itself. Assuming that these problems will be solved, these checks may lead to more reliable provisional estimates, and in some cases perhaps more accurate final accounts.

Reliability standards can be set in many ways. It is argued that the use of recent history leads to the most balanced set of standards. This makes the standards become stricter as the quality of the provisional estimates improves. When the quality of a variable suddenly changes, it allows for a short period to recover.

Compared with the average final estimates of the variables assessed some of these standards are very strict. Some of the standards do not even allow for a difference less than one quarter of the standard deviation of the final observation. Nevertheless, the number of hits on the macro-economic variables assessed is rather low. Together with the results of the study of Wroe et al. (1999) on the accuracy of the final accounts, one may conclude that the current national accounts, provisional and final, are sufficiently accurate and reliable.

The average revisions of the flash estimates of GDP for the period 1991-2000 is 0.35 percent point, which significantly differs from zero at a 99 % level. Looking at a larger time-period, between 1970-2000 the average revisions are 0.31 for the July t+1 estimates and 0.20 for the July t+2 estimates. Both still significantly differ from zero at a 99 % level, which implies that the provisional estimates tended to underestimate economic growth significantly. For the period 1970-1990, the average

revisions for the July t+1 estimates are 0.35 percent point. Therefore, bringing forward the first estimate from July t+1 to February t+1 didn't result in larger average revisions.

However, the complaints of the Netherlands Bureau for Economic Policy Analysis (CPB) formulated in 1991 (see introduction) and recently repeated in the "Centraal Economisch Plan" (CPB, 1999) are justified.

Underestimation is not a problem of Statistics Netherlands alone. International comparison of annual GDP figures (Öller and Hansson, 2002) reveals a general underestimation bias for many countries.

Starting with reference year 2000 Statistics Netherlands has incorporated the drawing up of a report on the revisions in the production process. As soon as the definitive figures of a reference year are available (internally), they can be used to test the provisional figures of that year. The figures on the underlying variables of the production approach and the expenditure approach show where the revisions of the figure on Gross Domestic Product originate from. Staff members of the national Accounts department responsible for the various estimates are asked to comment on the found revisions. The final report is discussed in the management team of the department to decide if additional actions should be taken. The users are informed by articles and papers (like this one) and the successive estimates are published orderly in StatLine, the output database of Statistics Netherlands on the Internet.

Implementing the system of reliability checks on main variables as part of the regular production process of national accounts, may lead to further improvements, also on a more detailed level.

Literature

Algera, S., 1995, *Reliability and timeliness.* Paper presented at the Main Economic Indicators 30th Anniversary Meeting, Paris, 16-17 October 1995. OECD (Paris).

Australian Bureau of Statistics, 1983, *Australian National Accounts, concepts, sources and methods.*

Bloem, A.M. and S. Khawaja, 2001, *A framework for assessing the quality of national accounts estimates.* IMF, Washington.

Bos, F. and C. Gorter, 1993, *Compiling Dutch Gross National Product (GNP); full report on the final estimates after the revision in 1992.* National Accounts Occasional Papers NA-57 Extended, CBS (Voorburg).

Buiten, G., 1993, *The relationship between final household consumption expenditure according to the National accounts and the Budget Survey for 1988 and 1989.* In Supplement Sociaal-Economische Maandstatistiek 93(5), CBS (Voorburg).

Buiten, G., 1994, *The relationship between final household consumption expenditure according to the National accounts and the Budget Survey for 1990 and 1991.* In Supplement Sociaal-Economische Maandstatistiek 94(5), CBS (Voorburg).

Byron, R.P., 1978, *The estimation of large social account matrices.* In Journal of the Royal Statistical Society, Series A 141, pp. 359-367.

Calzaroni, M. and A. Puggioni (1995), *A prelimenary approach for the analysis of the quality of national accounts estimates.* Paper presented at the International Conference on Survey Measurement and Process Quality, Bristol, April 1995. ISTAT (Rome).

CBS, 1985, *Nationale rekeningen 1969-1981 met herziene reeksen voor de jaren 1969-1976*. Staatsuitgeverij ('s-Gravenhage).

CBS, 1992, *Nationale rekeningen 1991, band II: Revisie van methoden en uitkomsten op het jaar 1987*. SDU/Uitgeverij (The Hague).

CBS, 1999, *Nationale rekeningen Revisiepublicatie; Beschrijving en uitkomsten van de ESR 1995 revisie.* CBS (Voorburg/Heerlen).

CPB, 1999, *Centraal Economisch Plan 1999*, CPB (The Hague).

ESA, 1995, *European System of Accounts; ESA 1995.* Eurostat (Luxembourg).

Franchet, Y., 1998, *Quality work and quality assurance within statistics.* Paper presented at the 84th DGINS conference, Stockholm, May 1998. Eurostat (Luxembourg).

Magnus, J.R., J.W. van Tongeren and A.F. de Vos, 2000, *National Accounts Estimation Using Indicators Analysis*. Review of Income and Wealth (46), pp. 329-350.

OECD, 2002, *Measuring the Non-Observed Economy; A Handbook*. OECD (Paris).

Öller, L-E and G-H Hansson, 2002, Revisions of Swedish National Accounts 1980-1998 and an international comparison. Statistics Sweden.

Reininga, F.K. and B. Kazemier, 1998, *Flash growth estimates using calendar information.* In Review of Income and Wealth 44(2), pp. 229-237.

Reininga, F.K., G.M. Zijlmans and R.J.A. Janssen, 1992, *Quality assessment of macro-economic figures: the Dutch quarterly flash.* National Accounts Occasional Paper NA-49, CBS (Voorburg).

Renssen, R., A. Camstra, C. Huegen, W. Hacking and R. Huigen, 1998, *A model for evaluating the quality of business surveys*. Research paper no. 9833, CBS (Voorburg).

Sefton, J. and M.R. Weale, 1995, *Reconciliation of national income and expenditure: balanced estimates for the United Kingdom, 1920-95.* Cambridge University Press (Cambridge).

SNA, 1993, *System of National Accounts 1993.* EC, IMF, OECD, UN, World Bank (Brussels/Luxembourg, New York, Paris, Washington, DC.).

Solomou, S. and M.R. Weale, 1993, *Balanced estimates of national accounts when measurement errors are autocorrelated: the UK 1920-38.* In Journal of the Royal Statistical Society, Series A 156, pp. 89-105.

Stone, R., D.G. Champernowne and J.E. Meade, 1942, *The precision of national income estimates.* In The review of economic studies IX, pp. 111-125.

Vlimmeren, J.C.G. van, F.J.H. Don and V.R. Okker, 1991, *Modelling data uncertainty due to revisions.* Central Policy Bureau, Research Memorandum no. 75, CPB (The Hague).

Vries, W. de, and R. van Brakel, 1998, *Quality systems and statistical auditing; A pragmatic approach to statistical quality management"*. Paper presented at the 84th DGINS conference, Stockholm, May 1998. CBS (Voorburg).

Weale, M.R., 1988, *The reconciliation of values, volume and prices in the national accounts.* In Journal of the Royal Statistical Society, Series A 151, pp. 211-221.

Wroe, D., P. Kenny, U. Rizki and I. Weerakkody, 1998, *Reliability and quality indicators for National Accounts Aggregates.* ONS (London).

Young, A.H., 1987, Evaluation of the GNP Estimates. In Survey of Current Business 67(8), pp. 18-42.