THE ART AND CRAFT OF COMPILING NATIONAL ACCOUNTS STATISTICS AND THEIR IMPLICATIONS FOR RELIABILITY

BY FRITS BOS*

CPB Netherlands Bureau for Economic Policy Analysis

This paper provides a systematic overview of the compilation and reliability of national accounts statistics. It illustrates the various issues with a wide range of examples and stories from national accounts compilation practice. National accounts statistics are estimates of a universal accounting model (SNA93). The operational versions of the model decide what is actually estimated. They are estimated by expanding and transforming the available data with accounting identities, assumptions, and plausibility checks. The estimates reflect personal knowledge and skills, resources, and policy. For a specific type of use, the universal and operational national accounting concepts are usually not perfect. The quantitative importance of such conceptual "measurement errors" is often overlooked but can be substantial. For assessing the reliability of national accounts statistics, sampling theory is not very important. The major methods are consistency checks, sensitivity analysis, and analysis using a description of the data sources, operational model, and compilation methods.

1. INTRODUCTION

National accounts statistics have something miraculous, like the miracle of estimates based on a sample. The miracle of sample-estimates is that a sample suffices to make estimates about the whole population. The miracle of national accounts statistics is that all over the world, very incomplete, imperfect, and partly outdated datasets are transformed into complete, consistent, and up-to-date standardized pictures of national economies. This miracle of the national accounts is often misunderstood. For most fellow-statisticians and users of the national accounts, compiling the national accounts is a black box. Some label the compilation process even as obscure and non-scientific ("a voodoo ritual"). Data users often take the availability of national accounts statistics for granted, while stressing some specific limitations in terms of reliability and comparability. However, they often do not realize that limitations in terms of reliability and comparability are partly the direct consequence of the miracle of the national accounts.

In this paper, the art and craft of compiling national accounts statistics and their implications for the reliability of national accounts statistics are discussed.[1]

Sections 2 and 3 provide a look behind the scenes. Section 2 is devoted to the role of the national operational models, i.e. the national operational versions of the universal accounting model System of National Accounts 1993 (SNA93, see UN, 1993; a new set of guidelines is currently being prepared). The national

*Correspondence to: Frits Bos, CPB Netherlands Bureau for Economic Policy Analysis, PO Box 80510, 2508 GM, The Hague, the Netherlands (f.bos@cpb.nl).

[1]For a general overview on the concepts, history, and uses of national accounting, see Bos (2009).

operational model decides what is actually to be estimated. Furthermore, differences between national operational models affect the international comparability of national accounts statistics. Insufficient knowledge of the underlying operational models can lead to misuse and misinterpretation of national accounts statistics.

In Europe, national accounts statistics play a major role in economic and fiscal policy and are also used to assess the contributions for the EU Member States.[2] In order to improve comparability, European guidelines on national accounting (ESA95; see Eurostat, 1995) have been issued. These guidelines have a legal basis and are more prescriptive and detailed. For some specific issues, e.g. measuring the services of owner-occupied dwellings, separate official regulations have also been issued. The European guidelines on national accounting can be regarded as a somewhat more operational version of the universal guidelines. However, they are still so general that they can never serve as the operational model for a specific country during a specific period.

Section 3 explains the compilation process. It describes the roles played by data, the three compilation tools (accounting identities, assumptions, and plausibility checks), and environmental factors (personal skills and knowledge, resources and policy). In this way, the logic underlying the compilation process is revealed.

The reliability of national accounts statistics is a major outcome of the compilation process. It is the topic of Section 4. For data users, it is important to know how reliable official national accounts statistics are in absolute terms and in comparison to other data sources. Lack of knowledge of the reliability of national accounts statistics can also lead to misinterpretation (see Bos, 2007b; Bos, 2009, chapter 8).

The international organizations (OECD, the UN, IMF, and Eurostat) expend considerable effort collecting the national accounts statistics of Member States and adjusting them to assure as much comparability as possible. They also provide handbooks and training courses on compiling national accounts statistics, and organize support for some specific countries. These efforts are very important for improving the comparability and reliability of national accounts statistics. However, in this paper, these activities of the international organizations are not further discussed.

The literature on national accounts compilation methods and the reliability of national accounts statistics is very limited.[3] The value added of this paper in view of the current literature is therefore manifold:
- It provides a systematic account of the various steps in compiling national accounts statistics.

---

[2]On national accounting in Europe, see Bos (2009, pp. 54–8).

[3]On compilation methods, see in particular Séruzier (1996; on compiling national accounts in general), Bos and Gorter (1993; a concise but partly outdated overview on the Dutch estimates of Gross National Product), Statistics Netherlands (2008; the most recent, much more elaborate, description of the Dutch estimates of Gross National Income), Takema (2000; on compiling input–output tables), den Butter (2007; on the construction and use of indicators from the national accounts for policy analysis), UN (2000; on the link between business accounts and national accounts), and Magnus *et al.* (2000; on Bayesian methods). On the reliability of national accounts statistics, see Novak (1975), Weale (1988), Wroe *et al.* (1998), Young (1987), and Vanoli (2005, chapter 5). More examples clarifying compilation practice can be found in Bos (2007a; 2009, chapter 7).

- It clarifies the role of accounting identities, assumptions, and plausibility checks.
- It discusses the role of environmental factors like skills and knowledge, resources, and policy.
- It puts the issue of the reliability of national accounts statistics in a general context and stresses the importance of sensitivity analysis and conceptual issues.

2. The Operational Model

The concepts prescribed by the universal national accounts guidelines cannot be estimated directly. They should first be translated into operational concepts for a specific country at a specific moment in time. The specification of the operational model involves four steps (see Table 1).

The first step in the specification of the operational model is the specification of the scope and detail of the national accounts statistic. The universal model is very encompassing. The specification of the scope therefore generally involves the selection of parts of the universal model, e.g. a simple set of sector accounts excluding balance sheets and other changes in assets accounts. The universal model contains also a considerable amount of detail, e.g. at very low levels of aggregation in the classification by industry and the classification of flows and stocks. Due to trade-offs, scope and detail generally also depend on the frequency and timeliness of the national accounts statistics, e.g. a trade-off between detail and timeliness. The desired detail can also depend on national demands for specific detail, the available detail in national data sources, or requirements of reliability, e.g. price measurement is influenced by the level of disaggregation used. Differences in level of disaggregation can therefore also influence the reliability and comparability of national accounts statistics.

The second step in the specification of the operational model is the interpretation of the universal concepts. This step is in fact an intermediary step, i.e. intermediary in the further specification and application of the universal concepts. Interpretation of the universal concepts is not always simple and straightforward. Concepts can be formulated in vague or even misleading terms; literal reading of the definitions of these concepts can then lead one astray. Furthermore, related concepts can also be defined in different and sometimes contradictory terminology. Finally, the definitions of some concepts may even conflict with the general accounting principles. An example can illustrate such problems of interpretation.

TABLE 1

From Universal Concepts to Operational Concepts

Four steps for the translation of universal concepts into operational concepts:
1. Specification of the desired scope and detail
2. Interpretation of the universal concepts
3. Specification of the universal concepts
4. Application to the national economy for a specific period

Do the services of owner-occupied dwellings also include imputing services and income for free-standing garages and holiday homes abroad? According to the general accounting logic, such imputations should indeed be made. However, the current universal guidelines (SNA93) are not explicit about this and create thus differences in interpretation. The current European guidelines (ESA95) contain therefore some explicit statements that such imputations should indeed be made.

The third step in the specification of the operational model is the further specification of the universal concepts at a national level. This step is required for universal concepts defined in too general terms. The example of volumes and prices may illustrate this. The *volumes and prices of the supply and use of goods and services* are defined in a not very specific way, e.g.:

- The volume of banking and insurance services is not at all defined.
- The volume of education services may be measured in terms of outputs (e.g. the number of pupils), but also in terms of inputs (e.g. the number of teachers).
- The quality change of computers and other ICT capital goods may be measured by a hedonic method but also by alternative methods. As a consequence, some countries (e.g. the U.S.) apply the hedonic method, while for other countries the price indices of ICT capital goods are only partly or not at all based on hedonic methods. In order to control for such differences, Collecchia and Schreyer (2002) make estimates of harmonized price indices for nine OECD countries on the basis of the U.S. ICT price indices.
- Fisher indices are preferred, but Laspeyres and Paasche indices are acceptable (in combination with chain linking).
- No clear guidelines are given about how to incorporate new products into price indices.

Operational definitions of prices, volumes, economic growth, productivity, and real income therefore differ substantially all over the world. The direct consequence is that the national accounts statistics on these major variables are not very comparable. This has induced the European Union to start a program for drastically improving the comparability of the European economic growth figures, and issue separate guidelines on measuring price and volume changes (see Eurostat, 2001). Furthermore, without any information on the operational definitions employed, data users may have a wrong impression about what is actually being measured. An example on measuring government output may well illustrate this.

Since the 1950s, the volume of government output was measured in the Dutch national accounts by its inputs and by definition no productivity growth occurred. However, in 1990 a comparison by the CPB of economic growth in Germany and the Netherlands revealed that Dutch growth lagged mainly behind due to this assumption of zero-productivity growth. After comparing in a rough way alternative assumptions (see Kazemier, 1991), a new convention was adopted: productivity growth for the government was equal to the (three years moving average of the) incidental wage increase; this yielded a productivity increase of about 0.7 percent per full-time equivalent worker, and raised the annual Dutch growth rate by 0.1 percent. During the last decade, incidental wage increases were much lower, resulting in assumed productivity increases of 0.3 percent. This productivity con-

vention is hardly known to data users, not very transparent, and therefore often misunderstood. For example, politicians stated that statistics proved that labor productivity increases of civil servants were much lower than in the market sector, suggesting that they work less effectively or hard! In times of budget cuts it was argued that by cutting employment about 0.7/0.3 percent annually, civil servants could continue to do the same tasks, but with less people.

When data are collected from businesses, this must usually be done in a manner that is as convenient as possible for those businesses. The data that businesses have readily available may not correspond precisely to what national accountants would like to obtain. Nevertheless, statisticians must often accepts what they can get and then make whatever adjustments they can to align the data with the (universal) national accounting concepts.

The fourth step is the application of the universal model to the national economy and its institutional arrangements for a specific period. For example, for compiling Dutch national accounts figures on 2009, it should be investigated and decided, for example:

- Which establishments, corporations, and institutions exist in the Netherlands in 2009 and what is their industry and sector.
- What revenues from the Dutch central and local government are taxes and what are sales by the general government.
- What payments by employers are social insurance contributions and what are wages.

The universal model is also flexible in view of specific economic circumstances, e.g. in case of hyperinflation, inflation accounting is recommended.

The application of the universal model is not straightforward. It requires knowledge and interpretation of the guidelines, knowledge and interpretation of the specific situation to be described, a judgment on what is feasible and efficient from a data compilation point of view, and a judgment on what is relevant for data users.

Application of the universal model may in practice involve some clear and deliberate *deviations from the universal model*. A common deviation pertains to the statistical unit employed. In the universal model, industries are defined in terms of establishments and not in terms of institutional units. However, in most countries, even in Europe, data on industries are actually based on institutional units. This is even explicit in the European regulation on manufacturing statistics: this regulation refers only to institutional units. The consequence of this common deviation is that the output and production processes of industries measured by national accounts statistics is much more heterogeneous and incomparable than suggested by the universal model.

For example, the output of shoe producers recorded in the national accounts statistics as part of the industry manufacturing may also include retail activities organized and administered completely separately from the shoe producing activities. Similarly, the output, wages, and employment of public administration in the national accounts statistics may also include those of units of the state government, provinces, and municipalities with respect to garbage disposal, cultural services (e.g. museums and libraries), social services, healthcare services (e.g. preventive health care such as vaccinations and health surveys), and manufacturing by workplaces for disabled people.

In the Netherlands, the revision of the national accounts in 1998 (see Buiten *et al.*, 1999) amounted to an increase of 10 percent of the value added of the sector general government. This increase was mainly due to a stricter adherence to the European rules for classifying units by institutional sector; it was only to a very limited extent caused by changes in the European rules.

Application of the universal model may also be *affected by non-statistical uses of the operational model*. For example, in Germany the classification of individual companies by industry in the statistical business register is also used to decide which collective wage agreement is relevant for a company. Therefore, there may be a tendency to incorporate changes only with some delays or not at all. Similarly, in some countries the industry defense should for political reasons reflect the responsibilities of the Ministry of Defense. When the Ministry of Defense owns, supervises, and exploits machine producing plants, these are therefore included in the national accounts statistics on the industry defense.

These four steps demonstrate that the operational model can have a considerable impact on what is actually measured by official national accounts statistics. Differences in operational models can also substantially affect the international comparability of official national accounts statistics.

When compiling national accounts statistics for the first time or when implementing new universal concepts, defining the operational model involves a considerable amount of work. When compiling national accounts figures annually, annual updating of the operational model is required. For example, companies can merge, change their major product (e.g. from mining to chemical products or from forestry to environmental protection), or go bankrupt, and the government can deregulate or privatize its tasks, start special employment projects, or reorganize social security. Without updating and an excellent communication of the updates, the national accounts figures will not be able to describe these developments.

3. The Compilation Process

3.1. *Introduction*

The major characteristics of the data, the compilation process, and environmental factors can be summarized as follows.

The *available data are very heterogeneous* in all respects, e.g. scope, concepts, detail, reliability, time of availability, and frequency. The available data will always be incomplete in terms of scope and detail. As a consequence, many estimates cannot directly be based on the available data. For reliable estimates a good frame of reference, e.g. a business register, is essential. The frame of reference helps in completing and combining surveys and administrative data, in updating former estimates, and in making estimates for parts of the economy on which hardly any information is available.

The compilation process is based on three estimation tools:
- Accounting identities.
- Assumptions for completing.
- Plausibility checks.

Compiling national accounts statistics amounts to exploiting as best as possible the many *accounting identities*. Examples are: supply is equal to demand (both in current and constant prices); the three basic ways to estimate GDP; taxes paid should be equal to taxes received; and the changes in stocks are equal to the sum of the flows. Accounting identities are friends and foes of national accounts statistics. They ensure consistency, can act as a plausibility check, and allow residual estimates. However, they can also force the modification of best estimates for the sake of consistency. Furthermore, all residual estimates are likely to be very unreliable, as they serve as the garbage bag for errors in all the other estimates.

Assumptions are essential in combining and completing the basic set of data. Many types of assumptions are used, e.g.:

- Grossing up of a survey on the basis of a frame of reference.
- The use of strong institutional, technical, or economic relationships.
- The use of fixed ratios, transition schemes, and lifetimes.
- The use of specific conventions, e.g. the productivity increase of government output is 0.5 percent per year.
- Assumptions are based on fragmentary qualitative information, expert opinions, historical trends and ratios, analogies, and anecdotes.

The more encompassing, up-to-date, detailed, reliable, and conceptually close the basic dataset, the smaller the role played by assumptions can be. Plausible assumptions can remedy to a substantial extent the absence of data and are to be preferred to implausible data. However, when for substantial parts of the national economy no plausible data or assumptions are available, national accounts statistics transform into guesswork.

Plausibility checks are very important for the reliability of national accounts statistics. Three types of plausibility checks can be distinguished:

- Comparison of different data/estimates.
- Investigation of all "strange" developments and ratios (numerically, conceptually, institutionally, economically) by looking for a plausible explanation.
- Investigation of the data on the presence of expected developments; in case of absence look for a plausible explanation.

Plausibility checks can weed out erratic developments in data sources (e.g. due to conceptual changes), can help in detecting all sorts of compilation errors, and are important in making estimates during all the various stages of the compilation process. What is regarded as plausible is ultimately decided by the compilers' skills in inventing plausibility checks, by the compilers' skills in finding plausible answers, and by the compilers' personal knowledge and model of the national economy.

The estimation process is influenced by three types of *environmental factors*:

- *Skills and tacit knowledge*, e.g. personal skills and knowledge for combining data and making plausible assumptions.
- *Resources*, e.g. resources for compiling good price statistics, for maintaining a reliable business register, or for compiling national accounts statistics.
- *Policy*, e.g. preference for prudence and stability, priorities for some parts of the national accounts, and independence from political pressure (see Section 3.3).

TABLE 2

SIX STEPS FOR COMPILING NATIONAL ACCOUNTS STATISTICS

Six steps for compiling national accounts statistics:
1. Specify the operational model
2. Collect data
3. Translate the data into the concepts of the operational model
4. Make an incomplete set of first estimates
5. Add supplementary estimates
6. Balance parts and the total

3.2. *The Compilation Process in Six Steps*

Compiling national accounts statistics can be summarized in six steps (see Table 2).

The first step, i.e. the specification of the operational model, determines what is actually to be estimated. It has already been discussed in Section 2. In this subsection therefore the other five steps are discussed.

Step 2: Collection of Data

The second step is the collection of data and qualitative information. The major data sources are usually specific statistics, e.g. on the sales and production costs of producers, on capital formation, on employment, on wages and salaries, on household expenditure, on consumer prices, producer prices, and interest rates, on imports and exports, or on revenues and expenditure by government bodies.

However, raw administrative data (i.e. those that are not translated into a specific statistic) can also be very important for compiling the national accounts. This can apply to, for example, VAT records, the business accounts of some large companies, annual reports by supervisory bodies on banking and private insurance, or the annual or quarterly accounts of the central government and social insurance bodies. Furthermore, mainly qualitative information can also be important. For example, articles in newspapers or specialized magazines may provide qualitative information on developments (e.g. on sales of furniture or software) or specific events (e.g. a large direct investment project or a reorganization of a social insurance). This information can be used to complete other data, to check the plausibility of other data, or to decide on the best way of bookkeeping for specific events and developments.

The inputs collected can play a direct or indirect role in compiling the national accounts. Direct inputs are specific statistics or administrative data sources used directly in estimating national accounts statistics. Qualitative information can provide a major quality check on other data and assumptions used during the compilation process. Specific statistics and administrative data sources can also serve as indirect inputs. For example, in compiling national accounts figures for 2008, specific statistics can be used that refer to earlier years or to 2008. Similarly, tax data can be used not only to estimate national accounts data on taxes, but also to estimate the tax base or to check the plausibility of such an estimate.

The collection of inputs is not a passive process but requires a lot of structural and ad hoc work. This work may consist of, for example:

- Negotiations and agreements on data delivery: which data will be delivered, which detail and frequency, when, in what format, how reliable, etc.
- The active monitoring and checking of the data delivery process: do the data really arrive at the time and as complete and detailed as agreed upon or expected.
- The storage of the data in the automated systems for compiling the national accounts: this can be, for example, typing in information into spreadsheets or databases, selecting only the relevant parts for national accounts purposes, or translating data to the type of software or layout used by the national accountants.
- The search for other relevant quantitative and qualitative information, e.g. by reading specialized journals, newspaper articles, and annual reports, or by explicitly asking corporations, institutions, and experts.

Step 3: Translate the Data into the Concepts of the Operational Model

The third step is to translate the input data into the mould of the operational model. In this respect it is meaningful to distinguish two types of input data: economic base statistics, and other, generally administrative, data.

Economic base statistics generally translate administrative concepts into national accounts concepts or proxies of them. Cases in point are production statistics and government finance statistics:

- A production statistic (the results of statistical inquiry of producers) is often based on a transition scheme deriving national accounts concepts of production from the questions asked of business in the questionnaire. For example, the value of production according to the national accounts can be defined as the aggregate of sales of own-production, net change in stock of own-production, own-account capital formation, trade margin on goods and services produced by other units, sales of other goods and services (e.g. rental), and other revenues excluding royalties on subsoil assets but including revenues for overhead activities.
- Government finance statistics on the state government can be compiled by translation of administrative records of the state government into the national accounts concepts of government revenue and expenditure and their components. For a proper translation it is generally required to exploit additional explanations about the exact content of each budgetary chapter and article.

As a consequence, only minor adjustments of such input data are often required to comply fully with the operational model. A major difference between economic base statistics and the national accounts is often the product breakdown. For economic statistics with very detailed product breakdowns, like Foreign Trade statistics, translation towards the national accounts classification involves the use of a transition scheme at a rather aggregate level. However, for other economic statistics, e.g. household budget surveys or a production statistic with respect to intermediate consumption, the product breakdown can be very

limited or rather different. In such instances, product groups in the economic statistics are to be split and rearranged in order to derive the product breakdown of the national accounts.

In case of the direct use of *administrative records*, the translation from administrative concepts to the operational model is performed by the national accountants themselves. This applies to, for example, VAT data, personal income tax data, business accounts of some very large companies, annual accounts of the largest municipalities, supervisory reports by the Central Bank on the banking sector, or social security institutions on wages, social security contributions, and social security benefits. However, it should be noted that the data needs of the national accounts may have been taken explicitly into account in some administrative records, e.g. when reporting for the government budget is based on national accounting concepts. It may also happen that survey divisions blend administrative data with survey data to obtain better and/or timelier survey estimates. Sometimes business surveys are designed to focus on the medium and large businesses. Administrative records and modeling are then used to take the small businesses into account. The statistics thereby produced are then used by the national accountants.

Frames of reference

The backbones for compiling national accounts statistics are the frames of reference. Frames of references provide a listing of all the existing enterprises, institutions, persons, and households, and their major economic characteristics, e.g. size in terms of number of employees or sales and type of product sold. Frames of reference can be, for example, a business register based on Chamber of Commerce registrations, a VAT register, a population census, or a count of employment. In order to obtain a complete and up-to-date frame of reference, it is generally necessary to combine the data from various data sources.

Frames of reference serve two major roles. Firstly, they are used to draw samples for surveys and to gross up surveys or administrative data for missing units, e.g. relatively small units or units exempted from registration. Secondly, they are used to combine different data sources and to transform the many partial estimates into a complete estimate about all producers, employees, or consumers in the national economy.

In compilation practice, different registers may be used for different groups of producers, e.g. a VAT register, a register with government units, a register with banks and insurance companies, and a register for healthcare institutions. This will generally imply omissions and double-counting in estimating the national economy. Furthermore, the use of registers that are not up-to-date and contain unreliable information (e.g. with respect to the size of the different companies or about the number of households) will cause unreliable estimates about the size and developments in the national economy.

The ultimate vision for national accountants is to develop detailed and comprehensive micro databases for all the individual governments, businesses, other institutions, and households. These data are constrained to be consistent with the national accounts aggregates. Within these databases, coherency would also be

enforced between, for example, an individual enterprise or legal entity and the various establishments that comprise it. Survey statisticians would then benefit from feedback from national accountants, just as the latter benefit from the products of the former.

Step 4: An Incomplete Set of First Estimates

The fourth step is to make a set of first estimates. Before actually starting to estimate, various plausibility checks on the various potential items of input data are made. Examples of such plausibility checks are:

- Check on the plausibility over time of one variable in one data source. For example, detecting erratic developments in output, such as increases of 20 percent and decreases of 30 percent in the next year without any clear motivation. Another example is detecting that a major part of the increases in VAT receipts was not due to increased sales but was caused by accelerated collection.
- Check on the plausibility of the level and development of ratios between different variables in one data source, e.g. between the volume of labor and the output.
- Check on the plausibility of values and volumes in one data source, e.g. increases of output of 20 percent accompanied by decreases of employment of 5 percent are generally not very plausible.
- Check on the plausibility by comparing different data sources, e.g. detailed production statistics are compared with more general information on the developments in one industry, with export statistics on the major product of this industry or with information on wages or employment of that industry.

For many different reasons, the first estimates are usually *based on several data sources*, for example:

- One data source is used for determining last year's level, while the change in level is derived from another, less reliable, data source. The reason for this mixed estimate is that the most reliable data source is not available for the current year.
- One data source is used for estimating the level of a variable, while another is used for estimating the composition (e.g. the breakdown by product group). The reason for this mixed estimate is that the first data source does not contain (sufficiently reliable) information on the composition.
- One data source is used for estimating the level of a variable, but corrections for conceptual differences (e.g. income in kind or underreporting of sales, income, and employment) are derived from other data sources. The reason for this mixed estimate is that the first data source does not fully comply with the concepts of the operational model.
- One data source is used for estimating the level of a variable, but corrections for missing units or units not to be included are derived from other data sources. This grossing up is usually done on the basis of a business register, a VAT register, a register on the number and composition of

households, or a population census. The reason for this mixed estimate is that the first data source does not comply with the classification of units in the operational model.

- One data source is used for estimating the volume of a variable, while another data source is used for estimating the price. An example is the estimation of agricultural output as harvest estimates in tons per product multiplied by auction prices for agricultural products. Another case in point is services of owner-occupied dwellings: by convention they are to be estimated as the volume of dwellings times the market rent for similar dwellings.
- One data source is used for estimating the value of a variable, while another data source is used for estimating the price or volume.

The first estimate can involve various types of *assumptions*, for example:

- The data in the sample (e.g. a survey of households or establishments in construction) or administrative data sources are sufficiently representative for those not included.
- Unchanged composition of a total, e.g. of the commodity breakdown of intermediate consumption by industry or of the breakdown of car registration taxes paid by consumers and by producers (by industry/institutional sector).
- Similar development of a total, e.g. assuming that the prices changes observed for some products are relevant for others, or that the average price change observed is a solid approximation of the average price change for the parts not observed.
- Constancy of a ratio, e.g. between sales and the number of employees of an establishment, between taxes on products and the sales of these products, or between income transfers by the government and the compensation of employees and purchases of goods and services financed by these transfers (e.g. income transfers to public schools).
- The change in the volume of government output is equal to the change in the volume of the various costs of production plus a fixed productivity increase of 1 percent.

Former estimates of the national accounts can also serve *as a data source*. This applies when the level of last year is extrapolated or when some breakdowns of last year are incorporated in the first estimates. The former estimate can reflect all elements of the previous compilation process, i.e. the inputs, the successive estimates, and the balancing.

Step 5: Secondary Estimates Resulting in a Complete Set of Estimates

The fifth step is to add secondary estimates. These secondary estimates are made starting from the set of first estimates.

These secondary estimates are often the direct consequence of applying national accounts identities. Examples are:

- In a commodity-flow approach, final consumption expenditure by households can be estimated as a residual item.

- Final consumption expenditure by households can also be estimated as the sum of the first estimates of final consumption per product group.
- Value added can be estimated as the residual of output minus intermediate consumption minus consumption of fixed capital.
- Taxes paid by the various sectors can be estimated starting from the taxes received by the government.
- Government output is defined as the sum of the various production costs. As a consequence, government output can only be estimated starting from the estimates of these costs.

However, there are also some other types of secondary estimates, for example:

- Splitting of the first estimates. For example, the first estimate of car registration taxes is to be split into those paid by producers (i.e. taxes on production) and those paid by consumers (i.e. current taxes on income, wealth, etc). Similarly, income in kind can be first estimated on the basis of income tax data and should then be allocated to the industries involved. Another case in point is sales via retail trade: these should be split into those bought by producers, those bought by domestic households as consumers, and those bought by tourists.
- Exploiting institutional relationships, e.g. the data on excise duties on beer are used to estimate the sales of beer, the investment grants on infrastructure are used to estimate capital formation on infrastructure, or the income transfers by the state government mainly financing other government units are used to estimate the size of the expenditure by the latter.
- Exploiting economic relationships, e.g. the first estimate of output in some industries is used to estimate their capital formation or their employment, or the first estimate of loans by banks to non-financial corporations is also used to estimate the corresponding interest payments. Some of these estimates may also be based on regression analysis.
- Exploiting technical relationships, e.g. knowledge about chemical processes can be used to produce a plausible estimate of the composition of the inputs and outputs of the chemical industry.

The fifth step also includes tertiary and even much more indirect estimates. For example, suppose capital formation for the national economy is a secondary estimate based on the supply of capital goods and exports of capital goods. Capital formation by industry is then a tertiary estimate. Estimating capital stock by industry is then the next estimate. Starting from this estimate, consumption of fixed capital by industry can then be estimated. This is then a quintal estimate.

After the fifth step, a set of estimates complete in detail and scope should have been obtained. Our examples have illustrated that the *sequence of estimation is generally not a straightforward bottom-up approach*. It is much more likely to be a very complicated mix of bottom-up, top-down, left-right (e.g. first supply and then demand or first taxes received then taxes paid), right-left, from values to prices and volumes, from prices and volumes to values, from flows to changes in stocks, and from changes in stocks to flows, etc. Only in this way can the limited set of available data be exploited in the most efficient way.

Assumptions have a leading role in completing the estimates. They are necessary to fill all major and minor gaps and imperfections in the basic dataset. This

implies that the more encompassing, up-to-date, detailed, reliable, and conceptually close the basic dataset, the smaller the role played by assumptions can be. Gaps and imperfections can first be mended by plausible assumptions (e.g. by exploiting economic and institutional relationships and by using a reliable and up-to-date frame of reference). All remaining gaps are then to be completed with more bold assumptions. For very small parts of the economy (e.g. construction companies representing about 10 percent of the total number of employees in the industry construction), this can be quite harmless. However, when bold assumptions about substantial parts of the national economy are necessary, national accounts statistics transform into guesswork.

Step 6: Balancing Plausibility and Consistency

The fifth step results in a complete set of estimates. However, this set is generally not consistent and not entirely plausible. The sixth step is therefore to balance parts and the total.[4]

Balancing is an overall plausibility check on the economic coherence of the estimates. What is actually checked on plausibility depends on the available accounting framework and the compilers' skills in inventing and performing plausibility checks in combination with the available time. A more encompassing and detailed framework gives more opportunities for plausibility checks. For example, a supply and use framework allows many plausibility checks, in particular when there is a simultaneous balancing in current and constant prices.

Balancing can also amount to enforcing consistency, i.e. preferring one estimate to another without having a clear opinion about which estimate is the best. Such enforced decisions will be more frequent and painful in case of rather well and complete sets of basic data in combination with a well developed compilation process involving numerous plausibility checks. Such enforced decisions hardly occur in case of a small set of reliable basic data and a very elementary compilation process. In the latter case, the focus will be fully on providing a complete set of estimates, i.e. Step 5 is then the last step.

Balancing can involve intensive interactive balancing at a disaggregate level, between different levels of aggregation, and between various parts of the national accounts. Examples of such balancing can be:

- Balancing the supply and use of specific product groups, e.g. achieving consistency and plausibility for the supply and use of textile or oil.
- Balancing the resources and uses of very specific transactions, e.g. the payment of public transport subsidies and the receipts of public transport subsidies, or the payment of income transfers by the state government to municipalities and the receipts of municipalities of income transfers by the state government.
- Balancing the overall supply and use of goods and services while investigating the plausibility of the development of the supply and use of specific product groups, e.g. those with respect to construction.

[4]On balancing in a supply and use framework, see e.g. Bos and Gorter (1993, pp. 120–30) and Statistics Netherlands (2008, pp. 281–98).

- Balancing simultaneously the production approach and the income approach for estimating GDP, e.g. with respect to compensation of employees by industry and operating surplus.

However, in practice it may also amount to a quite mechanical procedure. For example, GDP may be taken by convention as the average of the estimate of GDP via the production approach and that via the expenditure approach.

Balancing as a plausibility check amounts to *finding a plausible story behind "strange" developments*, i.e. what explains this development and which data sources and assumptions underlie these estimates.

External experts can play an important role in finding plausible answers. For example, they can indicate that specific events have occurred, such as a strike or explosion in some large plant. They could also signal major changes in the concepts or compilation methods underlying the various data sources, such as a change in the concept of wages in social security records or delays in recording changes in the number of unemployment benefits in one large administration.

What is regarded as plausible depends also on the compilers' personal knowledge and model of the national economy. Two anecdotes from Dutch compilation practice may illustrate this.

One compiler was an airplane watcher in his private time; he loved to look at planes coming and going at Schiphol and other Dutch airports. At some stage in the balancing process, the estimated use of capital goods substantially exceeded the supply of capital goods. He detected that some specific airplanes had been recorded as capital formation in capital formation statistics. However, on the basis of his personal knowledge as an airplane watcher, he knew that these planes had not yet arrived in the Netherlands. He investigated the Foreign Trade Statistics, which indeed did not include the import of these airplanes. This solved the inconsistency between supply and use: capital formation statistics included these airplanes when paid, while foreign trade statistics included these when delivered. The estimates of capital formation were adjusted, as the time of recording in the foreign trade statistics complied with the recording principles of the national accounts.

The second anecdote refers to the so-called method "Nooteboom," a former chief in the Dutch national accounts department. He was said to have a quite personal way of checking the plausibility of the estimates by his employees. When these estimates deviated substantially from his personal impression of the developments in Leidschendam, i.e. his place of residence, they were considered to be suspicious and had to be justified by his employees. As a consequence, the estimate of the development Dutch economy was partly benchmarked and calibrated on the basis of the developments in one Dutch town. Such a small "sample" is of course not very representative and implies the use of a very biased plausibility check.

Balancing is in practice not only important for solving complicated estimation problems. An important function of balancing is also to weed out various elementary human errors. Examples are a typing error drastically increasing output in a specific industry, a communication problem resulting in twice grossing up of a survey, and a drastic decrease of social benefits due to the thoughtless processing of social security data in which definitions or presentation have been changed.

Consistency checks are an elementary type of plausibility check: if estimates are inconsistent, there should be an error in the estimates. Inconsistency can

therefore help in tracing errors in the estimates. However, it can happen that despite a thorough reinvestigation no plausible explanation for this inconsistency is found. This can then only be solved in two ways.

The first solution is to modify the best estimates of some variables or part of the breakdown of variable (e.g. in some of the product groups of final consumption expenditure by households). In general this will be done where it does the least harm, e.g. where the estimates have the greatest margin of uncertainty or in which data users are least interested.

The variables that are each time sacrificed serve as a sort of garbage bin. Examples of such garbage bins can be changes in inventories or the use of business services (in order to solve the supply and use by product), and net sales of land by municipalities (in order to reconcile a plausible estimate of net lending with plausible estimates of the major categories of income and expenditure).

When the standard garbage bins are full or when there are no clear garbage bins available, painful choices have to be made for the sake of consistency. For example, a global indicator of final consumption expenditure by households, like VAT receipts, may indicate drastic increases, while the data for none of the quantitatively important product groups in household final consumption expenditure provide a clear explanation for this. The *choice is then between a reliable global estimate without a plausible underpinning by product group, and a not reliable global estimate with plausible estimates by product group*.

The second solution is to *show the inconsistency explicitly*. This is an approach often adopted for the substantial and highly fluctuating differences between net lending estimated from the non-financial side and net lending estimated from the financial side. Suppose the estimate from the financial side is considered to be the best. Consistency could then be achieved by modifying the estimates of the non-financial transactions, but which are to be adjusted: output, compensation of employees, interest payments? Substantial and highly fluctuating differences lead to substantial and highly fluctuating adjustments. This puts severe strains on the plausibility of the estimates over time.

Explicitly showing such differences between estimates is in particular a suitable solution when the inconsistencies are very large and not stable over time and where no relatively unimportant residual item(s) can be found. Nevertheless, very large and unstable inconsistencies, and also clear trends in the size of the inconsistencies, can still raise fundamental questions about the reliability of the estimates. So, in compiling national accounts statistics, consistency and plausibility seem to be friends but can also be foes that are difficult to reconcile.

3.3. *A More Complex Picture*

These six steps give in four respects a too simple picture of compilation practice:
- There is no strict chronology of steps.
- There is no identical compilation process each year.
- There is not one compilation process.
- It ignores the role of compilation policy and political pressure.

These steps do not represent a strict chronological sequence of separate steps: *in practice all these steps are intertwined, interact strongly, can occur in somewhat different chronological sequences, and have recursive loops*. For example, while balancing attention can be paid to the operational model, the plausibility of specific input data, and the translation into national accounts concepts.

A major feature of the national accounts is that *the compilation process is not identical each year*. This reflects that the available data sources and the national economy are not stable over time. Examples of changes in data sources are the arrival of new data sources, the disappearance or temporary absence of some data sources, or changes in the scope and concepts. As a consequence of such changes, the compilation process should often be adjusted. The same may apply to changes in the national economy, e.g. a new type of subsidy, new financial instruments, drastic increased importance of specific products (ICT), or a specific big event (e.g. a flood destroying crop).

The six steps suggest that a set of national accounts statistics is the outcome of one compilation process. However, *a set of national accounts statistics is generally the outcome of many successive compilation processes over a very long period of time*. Firstly, for a base year or a base period of several years, national accounts statistics are compiled. Secondly, starting from the base year estimates, for more recent years final estimates are compiled with some years delay. Thirdly, starting from the most recent final estimates, for the most recent years provisional estimates are compiled with some months or one or two years delay. Fourthly, starting from the base year estimates, time series are made for years before the base year. Fifthly, several countries now and then revise all their estimates, e.g. after a period of ten years.

The six steps are silent about the role of *compilation policy*. However, the compilation policy influences the reliability of the national accounts statistics in several ways:

- By choosing the timeliness of publication.
- By allocating the resources.
- By choosing a strategy of continuity.
- By choosing a strategy for prudence and stability of the estimates.

The timelier are national accounts statistics, the less data sources are available. As a consequence, more timely national accounts statistics should in principle be less reliable.

By allocating the resources for estimating the different parts of the national economy, the reliability of the estimates for these parts is also influenced. For example, increasing resources for estimating the services industries and decreasing those for manufacturing and construction, will likely (ceteris paribus) change the reliability of these estimates in the same direction. Similarly, allocating more resources to more timely estimates and less resources to the later estimates will influence the reliability of the estimates. However, in this case in the long run the more timely estimate may also become less reliable, as the latter is built upon the estimates of the least timely estimates.

National accounts statistics should preferably be both *up-to-date and continuous*. The former requirement means that estimates must comply with the most recent findings. Continuity means that the data from different reference

periods must be mutually comparable. These two requirements come into conflict whenever definitions change, the availability of data sources alters (e.g. a new more reliable data source becomes available), or estimation methods improve. Meeting the requirement of up-to-date estimates of levels means a permanently changing set of national accounts time series and can be very labor-intensive. However, a never changing set implies continuous but never up-to-date estimates. Most countries adopt a compromise (a mixed strategy of continuity). In compiling the annual figures, the continuity aspect is often given priority—with the result that levels are not up-to-date in some cases. At intervals, however, the data are revised to bring the whole series into line with the updated level for a specific base year. The mixed strategy of continuity implies that the estimates do not always represent the best estimate of the level.

Prudence and stability can be other reasons for not pursuing the best estimate. When the best estimate suggests a very strong or surprising development in the national economy, a prudent estimate is a somewhat weakened, less surprising, development. *A prudent estimate can also refer to the estimation method itself, i.e. prefer a method that can be most easily explained and defended against criticism.* Examples of practical rules adhering to his principle are:

- Make prudent assumptions, e.g. take an average or take the development of last year provided it was a "normal" year.
- Prefer the data source from the most reputed institutions.
- Prefer an explicit data source to plausible assumptions.
- Do not lightly deviate from long established estimation methods.
- Do not include any estimate at all when no solid information is available (e.g. on underreporting of income).

In case of successive published estimates, stability of the estimates can be an argument to deviate from the best estimate. It boils down in fact to a bias for the first published estimate, i.e. only deviate substantially from such estimates in case of strong evidence.

National accounts statistics and official macro-economic forecasts are part of public and private decision-making processes. The quality of statistics and forecasts in such a context does not only depend on purely statistical criteria. This is illustrated by a quote from the Director of the Dutch CPB, the institute that provides the official Dutch forecasts:

> Statistical criteria for forecast quality in practice have limited relevance. Three non-statistical criteria for forecast quality are put forward: logical coherence, economic coherence and stability. . . . a forecaster must enable his client to form his opinion on the uncertainty associated with the forecast. To this end, uncertainty variants and alternative scenarios appear adequate. (Don, 2001, p. 155)

Political pressure may also influence the reliability of national accounts statistics. This occurs in particular when national accounts statistics are compiled by government departments or a dependent national statistical institute. An independent institute taking care of compiling national accounts statistics is therefore an important requirement for reliable national accounts statistics. Major administrative uses of national accounts statistics, such as in the

European Union for fiscal and monetary policy and for determining the contribution to the EU, may endanger their reliability. However, the major upward revisions of the GDP estimates of several EU Member States (see Section 4.2) demonstrate that this is certainly not always true. Furthermore, it should be realized that governments have many different interests: in general, they may indeed want to minimize their contribution to the European Union and the UN and therefore a low national income. However, they may also want to show high economic growth rates as a measure of success for their policy. Furthermore, a high national income (per capita) is sometimes even a source of national pride: after a major revision of the Italian national accounts, all the front pages of the Italian newspapers hailed the victory over the U.K., as Italian national income surpassed that of the U.K.!

4. The Reliability of National Accounts Statistics

4.1. *Introduction*

The ambition of the universal model is to provide a relevant, complete, consistent, and standard overview of national economies all over the world. This ambition has clear and inherent trade-offs with reliable and comparable estimates, because:

- For the sake of relevance various imputations are to be estimated, e.g. income in kind, consumption of fixed capital, and the services of owner-occupied dwellings. However, a reliable estimate of such imputations is inherently difficult.
- One universal model is to be estimated even though available data sources, specific circumstances, and resources for compiling economic statistics differ widely internationally and inter-temporally; this implies that the reliability will differ widely internationally and inter-temporally.
- A complete set of price and volume measures is to be estimated even though perceptions on volumes and quality may differ widely or are not well specified by economic actors. This implies that reliability and comparability is in particular a problem for price and volume measures.
- The universal model can never serve as the operational model for a specific country during a specific period. This implies that differences in operational models will always affect international comparability.

National accounts statistics are generally the only available estimates of the universal model. As a consequence, they are relatively reliable for the model as a whole and its major aggregates like GDP.

For specific parts or individual variables, e.g. household consumption expenditure, alternative data sources can be available, e.g. a household budget survey. There are several reasons why national accounts statistics are likely to be more reliable and comparable than such alternative data sources:

- National accounts statistics will in general incorporate all reliable data sources on specific parts of the national economy; they combine these data in a sophisticated bookkeeping system which allows various extra checks on the reliability of the data.

- National accounts statistics reflect generally a long and intensive experience with the merits and limitations of the various data sources available nationally. For example, they include various standard corrections for omissions and errors in these data sources. They will therefore be more reliable than individual data sources.
- Due to the universal model, national accounts statistics are relatively explicit about what they want to measure, e.g. which concept of capital formation is employed or which definition of wages is used.
- National accounts statistics are each year based on the same set of universal concepts.
- National accounts statistics of different countries are based on the same set of universal concepts.
- National accounts statistics will not reflect political interference when compiled by an independent institute. Some other data sources may be clearly subject to political interference.

However, national accounts statistics are not necessarily the most reliable estimate for specific parts or individual variables. Several reasons have been mentioned in Section 3. For example, compilation policies of continuity and prudence may require not making the best estimates as such, a data source may arrive with too long delay, or limited efforts may have been taken for making a best estimate of some variables or detail.

Differences in compilation methods may also affect the international comparability of national accounts statistics. So, for international comparisons for specific parts or variables it may be worthwhile to ignore national accounts statistics and to process data sources available in several countries (e.g. manufacturing statistics) in one standard way.

For many uses of national accounts statistics, it does not suffice to know that the figures are the most reliable estimates available. For a proper use and interpretation of national accounts statistics it is also important to know how reliable they are, for example:

- Are the EU Member States national income estimates sufficiently reliable as a tax basis for the European Union?
- Is the small drop in GDP growth a statistical artifact or a real economic development?
- Are the supply-and-use tables sufficiently reliable to draw conclusions about trends in the contracting out by specific industries or developments in the relative prices of some specific products?
- How reliable are the coefficients of an econometric model estimated on the basis of national accounts time series?

4.2. *Assessing the Reliability of National Accounts Statistics*

Following Novak (1975), reliability can be defined in terms of accuracy and consistency. Accuracy is defined as the discrepancy between the observed and the "true" values. Since the "true" values must usually be approximated, the results of accuracy tests may themselves be subject to errors and biases. Accuracy tests

cannot be applied directly to national accounts statistics, but they could be applied to the basic data, like statistical surveys and administrative data records.

Consistency is defined as the discrepancy between two or more observed values, all of which could depart considerably from the "true" value. Consistency between two observed values cannot prove accuracy. Consistency tests merely provide methods for detecting deficiencies in accuracy. Consistency tests are very suitable for the national accounts, as they can show the sensitivity of national accounts statistics for using different data sources or estimation methods.

In Table 3, six different methods are distinguished for assessing the reliability of national accounts statistics.

The first method is to look at the sampling features of national accounts statistics.

Sampling theory indicates how to estimate the accuracy of sample estimates. However, in the national accounts sample estimates play a minor role, e.g. many administrative data sources are used, many specific estimation methods are used, and large companies and institutions (e.g. the state government) have a disproportionate economic importance. Non-sampling errors are therefore dominant in national accounts statistics, e.g. false reporting or the use of an outdated business register. As a consequence, sampling theory has not much to offer in assessing the reliability of national accounts statistics.

The reliability of national accounts statistics can also be assessed by investigating the data sources, the operational model, and the compilation methods (the second method). For example, the Dutch inventory on compiling national income estimates contains a table indicating that 78 percent of the estimate from the production side is based on good quality institutional data sources available annually. These data sources can be, for example, production statistics (e.g. for manufacturing), reports from supervising bodies (e.g. on banking and insurance), government finance statistics, and annual reports (e.g. of some large companies). The reliability of these data sources can be clarified further by indicating their

TABLE 3

DIFFERENT METHODS FOR ASSESSING THE RELIABILITY OF NATIONAL ACCOUNTS STATISTICS

General Method	Specific Method
1. Investigate sampling features	
2. Investigate data sources, operational model, and compilation methods	
3. Use reliability indicators for inputs	
4. Use a priori information on outputs	
5. Consistency checks	a. With alternative data and estimates
	b. Within one set of national accounts statistics
	c. With successive regular estimates
	d. With the estimates after a major revision
6. Sensitivity analysis	a. Different data sources
	b. Different compilation methods (including different compilers)
	c. Plausible assumptions about maximum errors
	d. Different operational concepts
	e. Different universal concepts

major features, e.g. the sample size of the survey and their quantitative importance vis-à-vis the population, major conceptual differences with the national accounts, and biases in the registration. Examples of such biases are the absence of small units, inadequate coverage of new products, and false reporting in order to reduce tax liabilities and to evade administrative obligations. Furthermore, the reliability of national accounts statistics can be disclosed by making explicit the size and motivation of the major corrections on the basic data, the cross-checks made, and the major assumptions used.

This second method is also the major method used in the European Union for verifying the reliability of national income estimates. All Member States have to submit an extended inventory on the data sources and compilation methods for estimating national income (see, e.g. Statistics Netherlands, 2008). This is then combined with audit missions by Eurostat experts and discussions with the experts from the Member States (see Bos, 2009, pp. 54–5 for a more detailed discussion).

The third method is that the compiler uses all his expert knowledge on the data and compilation methods to provide indicators of reliability. This is quite an old method: at the end of the 1940s, in the Netherlands, Sweden, and Norway official national accounts statistics were accompanied by indicators of reliability. In the Netherlands, for every item of the institutional sector accounts the uncertainty margin was given. This margin ranged from 2–5 percent (reliable estimate) to more than 20 percent (crude estimate). In Sweden and Norway, a scale from 1 to 5 was used. Separate indicators were used for the level and trend of time series (Aukrust, 1994, p. 45). Also, in the U.K., in the 1960s such reliability indicators were regularly published for national income and its major components: "A" refers to a margin of error <3%, "B" to 3–10%, and "C" to >10% (for an extended discussion, see Maurice, 1968, pp. 39–53). However, in all these countries this practice was discontinued.

Stone and others (see Stone *et al.*, 1942; Stone, 1981, 1986; van der Ploeg, 1982) showed that indicators of the reliability of input data can be used for balancing the national accounts. This formal procedure can also be used for deriving estimates of the reliability of the estimates after balancing. This formal approach is hardly used for balancing in national accounts' practice (see Sections 3.2 and 3.3). A major reason is that this procedure does not well ensure the plausibility of the estimated development of the national economy and its major components over time.

A fourth method is to use a priori information on expected numerical outputs, e.g. about the size of indicator ratios like the relation between output and value added, or between value added of large and small establishments. This Bayesian method for estimating national accounts and their reliability was proposed by Magnus *et al.* (2000). It is in particular relevant in the compilation of national accounts when basic data are very scarce, e.g. for developing countries or for very timely estimates.

A fifth method is to perform all kinds of consistency checks. The simplest way to test consistency is to look at the national accounts statistics themselves, i.e. at differences between successive estimates and at differences within one set of estimates.

National accounts statistics are often published as a series of successive estimates. The first estimates are the timeliest, but are also the least reliable. Later estimates can be based on more data sources and are therefore more reliable. Systematic investigation of the differences between these estimates can also reveal some important statistical qualities of the estimates. This is illustrated by a sample of one year from Dutch national accounts statistics.

The Dutch national accounts contain four successive estimates, in chronological sequence: quarterly accounts for a limited set of variables, preliminary annual estimates, semi-final annual estimates, and final annual estimates. Table 4 compares the final estimates of 1998, with earlier estimates of 1998. The estimates pertain to the volume change, the price change, and the level of some major aggregates, like GDP and final consumption expenditure. Furthermore, some key ratios are presented, e.g. national income per capita and government net lending as percentage of GDP. The table suggest that:

- Later estimates are generally better, i.e. closer to the final estimates.
- Volume changes may be more difficult to estimate than price changes.
- Volume changes of relatively volatile variables, like capital formation and exports, are more difficult to estimate than less volatile variables, like final consumption expenditure by the government.
- A particular category of volatile variables is balancing items, such as net lending by the government, value added, and net exports. They include the estimation errors of the underlying variables, e.g. the estimation error in the

TABLE 4

A COMPARISON OF SUCCESSIVE ESTIMATES IN THE DUTCH NATIONAL ACCOUNTS

	Final 98	Final vs. Semifinal 98	Final vs. Preliminary 98	Final vs. Quarterly 98
Volumes change (%)				
GDP	4.3	0.2	0.6	0.5
Final consumption expenditure	4.4	0.3	0.5	0.5
Government	3.6	0.2	0.3	1.1
Households	4.8	0.4	0.7	0.3
Capital formation	4.2	0.1	−1.0	0.2
Imports	8.5	0.5	0.8	2.4
Exports	7.4	0.0	1.0	0.1
Employment	2.9	−0.1	0.2	
Average size of deviation		0.2	0.6	0.7
Average deviation		0.2	0.4	0.7
Price change (%)				
GDP	1.7	−0.3	−0.2	−0.3
Final consumption expenditure	1.7	−0.2	−0.2	−0.4
Government	1.6	−0.5	−0.7	−1.3
Households	1.7	−0.4	−0.1	−0.1
Capital formation	2.0	−0.1	0.3	0.4
Imports	−1.5	0.0	0.0	0.8
Exports	−1.4	−0.2	−0.2	0.2
Average size of deviation		0.2	0.2	0.5
Average deviation		−0.2	−0.2	−0.1

price of gross value added is the net result of the estimation errors in the prices of output and intermediate consumption.

- Ratios as a percentage of GDP are not very sensitive to estimation errors.

By compiling such figures for a long range of years, important statistics about the reliability of the national accounts statistics of a country can be derived.[5] These statistics can reveal the average size of the estimation errors and the occurrence of systematic under- or overestimation in general and in specific situations, e.g. in case of substantial changes in the trade cycle. More detailed investigation can also clarify the links with the arrival of specific data sources and changes in compilation methods and policy, e.g. extra efforts to improve the reliability of the flash and preliminary estimates.

Such statistics on national accounts statistics can also be regarded as a *simulation of the effects of less data sources, less reliable data sources, and less resources and time for compiling national accounts statistics*. In this way, the relatively reliable national accounts statistics of countries like the Netherlands can also give a rough impression of the reliability of national accounts statistics of countries with much fewer data sources.

However, for two reasons the *differences between the successive estimates give a too optimistic picture*. Firstly, in compilation practice, a major role of the final estimates is also to calibrate the more timely estimates, e.g. to reveal their systematic errors. Secondly, ratios derived from the final estimates of last year are often used as inputs for the more timely estimates of the current year. As a consequence, the margin of error of the more timely estimates would have been substantially larger when the final estimates were not known and were not used in compiling the more timely estimates.

In many countries, national accounts statistics are once in a while revised. The changes in the estimates due to a revision (corrected for any changes in concepts) can also give an impression of the reliability of national accounts statistics. As part of the harmonization of European national income estimates, the estimates of several countries have been revised substantially, i.e. by 5, 10, or in one case even 20 percent. So, the margin of error of national income of the least reliable European national accounts statistics was between 5 and 20 percent. The national income estimates of many countries outside Europe are based on a similarly weak or even weaker statistical infrastructure. Similar margins of error can therefore be expected in the national income estimates of these countries.

A common feature of the adjustments in Europe was that all the major adjustments were upward adjustments. Furthermore, most of the smaller adjustments were also upwards. This suggests that in compiling national accounts, underestimation is a more serious danger than overestimation. This can reflect prudence as a compilation strategy. However, it can also reflect:

- Prudence in specific cases, e.g. in estimating parts of the economy on which data are very scarce.

---

[5]In the Netherlands, statistics on successive estimates are regularly published as part of general policy of quality control (see Kazemier *et al.*, 2004). At present, this is confined to the estimates of volume changes of domestic product and its breakdown by type of final expenditure and by industry.

- Ignorance. For example, the compilers are ignorant about the existence or scale of some economic activities and therefore do not include separate estimates on these activities.
- A lack of interest in estimating on some specific parts of the national economy, e.g. limited compilation efforts are devoted to estimating some types of services industries.

One set of national accounts statistics may also show differences between different estimates. The most common is a *difference between net lending estimated from the financial side and net lending estimated from the non-financial side*. This occurs in particular for sectors where the data sources and estimation process differ substantially for both sides.[6] This will in general not apply for financial institutions and the government, but it often will for households, non-financial corporations, and the rest of the world.

The differences between the estimates of net lending can overestimate and underestimate the measurement error in individual variables. Overestimation occurs when the measurement error is not located in one variable but is spread over many. However, underestimation is also likely, because measurement errors can partly compensate each other, e.g. an underestimation of exports can be compensated by an underestimation of imports.

Substantial changes in the statistical difference from one year to another can also indicate errors in measuring the change over time of some specific variables.

The last method is sensitivity analysis. It is a very important tool to shed light on the reliability of national accounts statistics. Two examples may illustrate this.

The first example is Broesterhuizen (1983). He investigated the reliability of Dutch GDP estimates in view of *fraud*, i.e. not reporting or underreporting of income to the tax and social security authorities or in statistical surveys. For this sensitivity analysis, GDP estimation was distinguished into six components reflecting the Dutch data and compilation process:

1. "Indirect methods," e.g. agriculture and operating of real estate are estimated by multiplying estimates of the volume of sales/output by an estimated price.
2. "Government," including also many enterprises and institutions supervised by the government, like public utilities, national railway and telephone companies, banking and insurance, and hospitals.
3. "Large firms," i.e. with more than 100 employees and included in surveys of producers.
4. "Small firms," i.e. with less than 100 employees and included in surveys of producers.
5. "Very small firms," i.e. with less than 100 employees and not included in surveys of producers.
6. "Fiscal data," i.e. specific groups of self-employed estimated on the basis of fiscal data.

For each component, plausible ranges of distortion by fraud are indicated. In this way a plausible upper bound for the distortion of GDP by fraud is derived.

---

[6]Minor differences can easily be reconciled by taking the estimate of one transaction as the residual item.

The conclusion was that for GDP levels a distortion of more than 5 percent is very unlikely, and that for GDP growth rates a bias of more than 0.5 percent is very unlikely.[7]

The second example pertains to *measurement errors in foreign trade statistics*. The existence of such errors is evident from comparing the mirror statistics of countries, even of those with a good statistical reputation (see van Bergeijk, 1995). For example, comparing the bilateral trade statistics of Germany and the Netherlands indicates implicit minimal measurement errors of about 1.5–3 percent. This can also result in quite a different view of the bilateral trade. For example, the balance of Dutch–German bilateral trade in 1987–89 turns into deficit according to German trade data, whereas the data published by the Netherlands show a persistent surplus.

What are the consequences of such measurement errors for the national accounts, e.g. the Domestic Product estimates? These consequences are not straightforward as they depend critically on the compilation methods used. An error in foreign trade statistics does influence Domestic Product when estimated from the expenditure side or when using commodity flow methods. But such an error is irrelevant for Domestic Product when estimated from the production side, i.e. as the difference between output and intermediate consumption. The latter may imply that in balancing the accounts the error in foreign trade statistics is detected and corrected for in estimating imports and exports. Unfortunately it may also imply that the estimates of variables other than imports and exports are adjusted in order to balance the accounts, e.g. final consumption expenditure by households. So, errors in foreign trade statistics may spread all over the national accounts.

National accounts statistics are also influenced by personal skills and knowledge (see Sections 3.2 and 3.3). The role of personal skills and knowledge in the practice of econometric model building was tested by experiments with one dataset and various participating teams (see Magnus and Morgan, 1999; Bos, 2009, p. 69). Experiments may also clarify their role for national accounts statistics. For example, ask different teams (from various countries/from the same national accounts department/experienced national accountants and novices) to compile national accounts statistics on the basis of a specific dataset and some general instructions.

A sensitivity analysis is a very important tool for assessing the reliability of national accounts statistics. Only in this way can the consequences of errors in data sources or the absence of data sources be properly investigated. The reason is that these consequences critically depend on the bookkeeping mechanisms in the national accounts and the compilation methods actually employed. For data users with a sound understanding of national accounts, documentation of the operational model and compilation methods may be sufficient to draw qualitative inferences about the consequences of measurement errors in national accounts data sources. However, for most data users, explicitly spelling out of

[7]A similar sensitivity study with similar conclusions has been done by Statistics Canada (see Smith, 1994).

the consequences of bookkeeping mechanisms and compilation methods and giving an estimate of the size of these consequences will be essential supplementary information.

From the point of view of a specific use of national accounts statistics, the underlying concepts could be regarded as approximations of the true concepts, i.e. of the theoretically perfect concepts for that type of use. As a consequence, conceptual imperfections could also be regarded as measurement errors. Such conceptual measurement errors could relate to the universal model and to the national operational model. Such errors can be very substantial. Four examples may illustrate this.

Firstly, if Domestic Product is taken as a proxy of welfare, the measurement error is likely to be very large, e.g. because leisure time, unpaid household services, and inequality are not taken into account.

Secondly, gross figures on value added, Domestic Product, and National Income are often preferred to net figures, because estimates of capital consumption are not considered to be sufficiently reliable or comparable. However, this argument overlooks the importance of the conceptual error, i.e. by preferring gross figures, expenditure on capital formation is never regarded as an intermediate input for the production process. As a consequence, productivity and income per capita of capital intensive industries and countries are overestimated vis-à-vis more labor intensive industries and countries (for a more extensive discussion, see Bos, 1992).

Thirdly, for monitoring the solvency of European government finance, gross government debt figures are used. However, they do not give a balanced picture of the financial position of the government. Figures from the Netherlands can illustrate this (see Bos, 2008, box 4, p. 28). Since 1994, Dutch gross government debt decreased with 25 percent GDP. This substantial decrease in debt is more than compensated by a decrease in the natural gas stock and other property: net worth decreased with 16 percent GDP!

Fourthly, the successive revisions of the Dutch national accounts also illustrate the quantitative impact of some very specific conceptual changes. For most data users, these changes are not regarded as becoming closer or less close to the true concepts or most optimal concepts for their specific use. They are generally just regarded as modifications of earlier national accounts estimates, i.e. as if no conceptual changes were made. In the revision of 1999, the concepts of the new international guidelines (SNA93 and ESA95) were introduced (see Buiten et al., 1999). This implied a substantial change in basic concepts:

- The introduction of capital consumption on infrastructure increased government final consumption expenditure and Gross Domestic Product with 1.4 percent GDP.
- Reinvested earnings on direct foreign investment are now included in the primary income flows with the rest of the world: Dutch national income increased with 1.3 percent GDP.
- A major shift occurred between final consumption expenditure by households and by the government. The latter now also included 8 percent GDP of social benefits in kind via market producers, like healthcare services financed by social security and free training for students.

- Following new EU guidelines, the recording of the services of the financial intermediation services indirectly measured (FISIM) was changed. As a consequence, final consumption expenditure by households increased with 1.2 percent GDP, government final consumption expenditure with 0.2 percent GDP, and exports with 0.4 percent GDP. The total impact on GDP was therefore an increase of 1.8 percent.

A major drawback of official national accounts statistics is that usually no information is provided about their operational concepts and reliability. This is a major reason for misuse and misinterpretation of national accounts statistics. This should be remedied by:

- Production and dissemination of statistics about differences in successive estimates.
- Documentation of the operational concepts, data sources, and compilation methods.
- Sensitivity analysis about the consequences of various major types of measurement error.
- Sensitivity analysis about alternative specifications of the operational model, e.g. the consequences of introducing hedonic prices for computers.
- Sensitivity analysis about alternative specifications of the universal module, e.g. a welfare-oriented module on household income.

This approach has some similarity with the old habit of some statistical offices to publish margins of error. However, such margins of error represent a too indirect, crude, and subjective way to address the issues at stake. Furthermore, they do not explain the underlying logic and sources of errors and biases; also the—often quantitatively very important—conceptual issues are generally ignored.

References

Aukrust, O., "The Scandinavian Contribution to National Accounting," in Z. Kenessey (ed.), *The Accounts of Nations*, IOS Press, Amsterdam, 16–65, 1994.

Bos, F., "Reasons for Preferring Net to Gross Figures of Income and Product (and Vice Versa)," *Review of Income and Wealth*, 38(3), 267–80, 1992.

———, "Compiling the National Accounts Demystified," *National Accounts Occasional Paper 95*, Statistics Netherlands, Voorburg, 2007a.

———, "Use, Misuse and Proper Use of the National Accounts," *National Accounts Occasional Paper 96*, Statistics Netherlands, Voorburg, 2007b.

———, "The Dutch Fiscal Framework; History, Current Practice and Role of the CPB," *OECD Journal on Budgeting*, 2008/1, 7–48, 2008.

———, *The National Accounts as a Tool for Analysis and Policy; In View of History, Economic Theory and Data Compilation Issues*, VDM Publishers, Saarbrücken, 2009.

Bos, F. and C. N. Gorter (eds), "Compiling Dutch Gross National Product (GNP); Full Report on the Final Estimates After the Revision in 1992," *National Accounts Occasional Paper 57, extended*, Statistics Netherlands, Voorburg, 1993.

Broesterhuizen, G. A. A. M., "The Unobserved Economy and the National Accounts in the Netherlands," *National Accounts Occasional Paper 2*, Statistics Netherlands, 1983.

Buiten, G., J. van den Hof, and P. van de Ven, "Revision of the Dutch National Accounts; First Results and Backgrounds," *National Accounts Occasional Paper 91*, Statistics Netherlands, Voorburg, 1999.

Collecchia, A. and P. Schreyer, "The Contribution of Information and Communication Technologies to Economic Growth in Nine OECD Countries," *OECD Economic Studies*, 34, 153–71, 2002.

den Butter, F. A. G., "National Accounts and Indicators," in M. Boumans (ed.), *Measurement in Economics*, Elsevier, Amsterdam/Boston, 189–229, 2007.

Don, F. J. H., "Forecasting in Macroeconomics: A Practitioner's View," *De Economist*, 149, 155–75, 2001.

Eurostat, *European System of Accounts; ESA 1995*, Office for Official Publications of the European Communities, Luxembourg, 1995.

———, *Handbook on Price and Volume Measures in National Accounts*, Office for Official Publications of the European Communities, Luxembourg, 2001.

Kazemier, B., "Volume Measurement of Government in the Netherlands; Some Alternatives," *National Accounts Occasional Paper 45*, Statistics Netherlands, 1991.

Kazemier, B., H. Nijmeijer, and R. Hijman, "How to Judge the Reliability of Provisional National Accounts," Paper presented at the European conference on Quality and Methodology in Official Statistics, Mainz, May 24–26, 2004.

Magnus, J. R. and M. S. Morgan, *Methodology and Tacit Knowledge; Two Experiments in Econometrics*, John Wiley & Sons, New York, 1999.

Magnus, J. R., J. W. van Tongeren, and A. F. de Vos, "National Accounts Estimation Using Indicator Ratios," *Review of Income and Wealth*, 46(3), 329–50, 2000.

Maurice, R. (ed.), *National Accounts Statistics; Sources and Methods*, Central Statistical Office, London, 1968.

Novak, G. J., "Reliability Criteria for National Accounts," *Review of Income and Wealth*, 21(3), 323–44, 1975.

Séruzier, M., *Construire les Comptes de la Nation selon le SCN 1993*, Economica, Paris, 1996.

Smith, P., "Assessing the Size of the Underground Economy: The Statistics Canada Perspective," *Income and Expenditure Accounts Technical Series 28*, Statistics Canada, 1994.

Statistics Netherlands, *Gross National Income Inventory (ESA95), 2001, The Netherlands*, Statistics Netherlands, Voorburg, 2008.

Stone, R., "Balancing the National Accounts: The Adjustment of Initial Estimates—a Neglected Stage in Measurement," in A. Ingham and A. M. Ulph (eds), *Demand, Equilibrium and Trade*, Macmillan, London, 1981.

———, "Progress in Balancing the National Accounts," in D. Ironmonger *et al.* (eds), *National Income and Economic Progress*, Macmillan, London, 1986.

Stone, R., D. G. Champernowne, and J. E. Meade, "The Precision of National Income Estimates," *Review of Economic Studies*, 9, 111–25, 1942.

Takema, T., "Supply and Use Tables in Current and Constant Prices for the Netherlands; An Experience of Fifteen Years," *National Accounts Occasional Papers 92*, Statistics Netherlands, Voorburg, 2000.

UN, *System of National Accounts (SNA93)*, UN, IMF, World Bank, and European Commission, 1993.

———, *Links Between Business Accounting and National Accounting*, Studies in Methods, Series F, No. 76, United Nations, New York, 2000.

van Bergeijk, P. A. G., "The Accuracy of International Economic Observations," *Bulletin of Economic Research*, 47(1), 1–20, 1995.

van der Ploeg, F., "Reliability and the Adjustment of Sequences of Large Economic Accounting Matrices," *Journal of the Royal Statistical Society*, Series A, 145, 169–94, 1982.

Vanoli, A., *A History of National Accounting*, IOS Press, Amsterdam, 2005.

Weale, M. R., "The Reconciliation of Values, Volume and Prices in the National Accounts," *Journal of the Royal Statistical Society*, Series A, 151, 211–21, 1988.

Wroe, D., P. Kenny, U. Rizki, and I. Weerakkody, *Reliability and Quality Indicators for National Accounts Aggregates*, Office for National Statistics, London, 1998.

Young, A. H., "Evaluation of GNP Estimates," *Survey of Current Business*, 67(8), 18–42, 1987.