

## Bayesian integration of large SNA data frameworks with an application to Guatemala

Jan W. van Tongeren<sup>a</sup> and Jan R. Magnus<sup>b,\*</sup>

<sup>a</sup>Former Head of the National Accounts Branch, United Nations Statistics Division, New York, USA

<sup>b</sup>Department of Econometrics and Operations Research, Tilburg University, Tilburg, The Netherlands

We present a Bayesian estimation method applied to an extended set of national accounts data and estimates of approximately 2500 variables. The method is based on conventional national accounts frameworks as compiled by countries in Central America, in particular Guatemala, and on concepts that are defined in the international standards of the System of National Accounts. Identities between the variables are exactly satisfied by the estimates. The method uses ratios between the variables as Bayesian conditions, and introduces prior reliabilities of values of basic data and ratios as criteria to adjust these values in order to satisfy the conditions. The paper not only presents estimates and precisions, but also discusses alternative conditions and reliabilities, in order to test the impact of framework assumptions and carry out sensitivity analyses. These tests involve, among others, the impact on Bayesian estimates of limited annual availability of data, of very low reliabilities (close to non-availability) of price indices, of limited availability of important administrative and survey data, and also the impact of aggregation of the basic data. We introduce the concept of ‘tentative’ estimates that are close to conventional national accounts estimates, in order to establish a close link between the Bayesian estimation approach and conventional national accounting.

Keywords: Macro accounts, system of national accounts, data frameworks, reliability, bayesian estimation, sensitivity analysis, aggregation

### 1. Introduction

This paper describes the use of a Bayesian estimation approach in the compilation of national accounts. The application is based on a project carried out in Central America and the results are presented for one country, namely Guatemala. The compilation involves approximately 2500 variables, which is close to what is conventionally involved in an extended compilation based on the international standards of the System of National Accounts (SNA). The basis of the SNA compilation was mainly the 1993 version of the SNA; the 2008 version was not fully implemented in the countries that participated in the project; see [19,20].

The compilation of SNA data with the help of a Bayesian estimation approach builds on what was developed in [4,9,10]. The paper presents, for the first time,

---

\*Corresponding author: Jan R. Magnus, Department of Econometrics and OR, Tilburg University, Tilburg, The Netherlands. E-mail: magnus@uvt.nl.

a real-life application to a large and realistic data set. Moving from a small data set (40 variables in our 2000 paper [10]) to a large data set (2500 variables) is a big step. The experiences acquired in the application of the Bayesian approach to several countries in Central America led to many improvements in the method and the software. In particular, extensive use is now made of 'sparse' matrix theory in the SNAER (System of National Accounts Estimation and Reconciliation) software in order to increase the accuracy and speed of the estimation process.

The Bayesian approach is applied to 'frameworks' of data and estimates, which can be used both in compilation and analysis. The frameworks are matrices or groups of matrices, in which two types of relations are defined between the variables: ratios and identities. The identities are based on SNA definitions and SNA balances; the ratios between variables are similar to the compilation ratios used by national accountants and also reflect simple ratio analyses carried out by analysts using the national accounts and other data. The frameworks make explicit the prior reliabilities of available data and ratio values that reflect the degree of willingness to change values, based on implicitly perceived trust in these values.

The BSNA (Bayesian SNA) framework used in the present study was developed for six Central-American countries, as part of a project sponsored by the Netherlands Organization for International Cooperation in Higher Education (NUFFIC) in cooperation with the Instituut voor Ontwikkelingsvraagstukken at Tilburg University, The Netherlands, and the Consejo Monetario Centroamericano (CMCA). The countries consist of the six member states of CMCA: Costa Rica, El Salvador, Guatemala, Honduras, Nicaragua, and the Dominican Republic. The data in this paper refer to the Guatemala BSNA for 2005 as the benchmark year, and 2006 as the current year.

The objective of the Bayesian method is the same as in conventional national accounting, where it is called 'reconciliation' or 'integration' of data. But Bayesian estimation or integration differs from the conventional national accounting approaches in several respects. First, all conditions of conventional estimates are formalized: identities are explicitly included and ratios are introduced as priors; second, reliabilities of data and ratio values are reflected in well-defined prior variation coefficients; third, the system is simultaneous rather than sequential; and finally, updating the system when new information becomes available is easy and fast, and does not require changes in the compilation method.

In the estimation process priors and data are combined to a posterior distribution. The mean of the posterior distribution is then taken as the estimate and the variance in the posterior distribution as a measure of precision. In this way, Bayesian estimates of all variables of the framework are derived, also for the variables for which no basic data are available and also for the ratios. In addition, standard deviations of all estimates and ratios defined in the framework are obtained.

The plan of this paper is as follows. In Section 2 we describe some characteristics of the economy of Guatemala, and discuss data availability and how conventional national accounts are compiled. These economic characteristics and the data compilation approaches are reflected in the design (or 'architecture'; see [7,8,24]) of the

BSNA data framework, presented in Section 3. Particular attention is paid to the use of ISIC and CPC classifications in the BSNA framework and to the SNA transactions that are incorporated in the framework and the types of analyses that they support. Section 4 provides details on the Bayesian inputs that are used in the compilation: ratios, identities, and reliabilities of basic data and ratio values. In Section 5 we summarize the Bayesian estimation method and how this method is reflected in two software programmes, SNAER and INTERFACE, that are used to arrive at Bayesian estimates. Section 6 presents the results of a number of tests of the framework, including tests that measure the impact on the Bayesian estimates if fewer data are annually available, and tests of different prior reliabilities assigned to basic data and ratio values. This section also describes the use of so-called 'tentative' estimates that are close to conventional national accounts estimates and that result in improved Bayesian estimates. Section 7 describes the results of a number of sensitivity tests. These tests quantify the impact on the Bayesian estimates of different scopes of the framework, of different availability of basic survey and administrative data, and of aggregation of basic data. They also quantify the impact of these alternative options on the posterior reliability of the Bayesian estimates. Section 8 summarizes our experiences with the Bayesian estimation method, and provides some suggestions for further work.

## **2. Characteristics of Guatemala**

Guatemala is a small country in Central America, about 5% of the size of Mexico. It is bordered by Mexico to the north and west, the Pacific Ocean to the southwest, Belize to the northeast, the Caribbean to the east, and Honduras and El Salvador to the southeast. With a population of close to 14 million, it is the most populous of the Central-American countries. Its per capita GDP (\$2115 in 2005) is about one quarter of Mexico's. Coffee, sugar, and bananas are the main products. The Central American Free Trade Agreement between Guatemala and the USA, signed in 2006, has increased investment in the export sector. Guatemala has an important 'free trade' zone, in which semi-manufactured products are temporarily imported, processed, and then exported again. Guatemala's economy is dominated by the private sector, which generates about 85% of GDP. Most manufacturing is light assembly and food processing, geared to the domestic, US, and Central-American markets. In addition to the main exports (coffee, sugar, bananas), the exports of textiles, apparel, and non-traditional agricultural products (mainly winter vegetables, fruit, and cut flowers) have boomed in the past few years, as of course has tourism. The revenues of migrant workers, employed mainly in the USA, are an important element of household disposable income. The distribution of income remains highly unequal; in fact, the most unequal in the region. The government sector is small with its business activities limited to public utilities, ports and airports, and several development-oriented financial institutions. The economy's structure is reflected in

the International Standard Industrial Classification of All Economic Activities (ISIC) and Central Product Classification (CPC) breakdowns of the BSNA framework, as explained in Section 3.

As in other countries in Central America, national accounts in Guatemala are compiled by the Central Bank. The main focus of the national accounts compilation is the Supply and Use Table (SUT), but it also includes the compilation of Integrated Economic Accounts (IEA). The Central Bank also compiles financial accounts and balance sheets for monetary analysis. The two compilation activities are carried out by different departments in the Central Bank and managed by staff with different specializations: national accountants versus monetary economists and modelers. Communication between the two types of specialists is often difficult.

Since an extension of BSNA to monetary accounts in the future is likely, the BSNA framework pays special attention to the development of the IEA and to sectorization therein, emphasizing in particular the three pillars of monetary-fiscal-financial analysis: the government sector, the financial corporate sector, and the rest of the world. The remaining sectors are further broken down for national accounts purposes into private and public non-financial corporations, households, and non-profit institutions. By making households a separate sector in the IEA, the framework distinguishes, within the ISIC classification by industries, between large production units that belong to the non-financial corporation sector and small production units that belong to the household sector. It is then also possible to distinguish between a measure of net product, i.e. GDP generated by the economy, and household disposable income, which is that part of GDP that reaches households and thus can be considered as an approximate measure of well-being. In addition, through concepts of household saving and net lending, we can then measure the contribution of households to the financing of investments.

### **3. The Bayesian SNA framework**

The BSNA framework is an SNA framework designed in Excel format. The Excel cells do not only represent values of variables in the framework, but also identities and ratios defined between the variables, and reliabilities attached to the values of the variables and ratios. A number of studies have recently questioned the accuracy of Excel as an analytical software program (see [11,12]). Our approach relies on the SNAER package (written in Fortran) to perform the complicated calculations, while Excel is only used for organizing and checking the data.

The degrees of freedom in designing the BSNA framework were not large, because the data to which the Bayesian method has been applied are based on what the 1993 SNA recommends in terms of transaction and transactor concepts. The framework is mainly redesigned in terms of the classifications used and the variables included, representing specific features of Guatemala's economy.

A synopsis of the framework is presented in Fig. 1. Cells with different types of information are distinguished by different shades. There are four types of information in the framework, namely:

- Variables for which basic data are available;
- Variables for which no basic data are available, but have values that are derived, using base year structures and SNA identities;
- Ratios that are defined between the variables of the framework, i.e. variables with or without basic data. Both ratio definitions and values are used; and
- Identities that are defined between the variables of the framework. These are defined in such a manner that, if satisfied, identity cells have value = 0. Both identity definitions and identity values are used.

BSNA Guatemala is represented by six alternative frameworks for 2005 and 2006, which all have the same format as Fig. 1. These are referred to as alternative options in Table 3 and in the tests and sensitivity analyses of Sections 6 and 7.

**2005GV** is the framework for the benchmark year 2005. It includes a complete data set with all values of variables treated as basic data (colored green (G) in the Excel frameworks). The values (V) of all framework cells are known. The benchmark framework does not only include basic data, but also estimates that are made by national accountants on the basis of similar assumptions (ratios and identities) as are used in the BSNA compilation. All cells in the benchmark framework are treated as basic data for BSNA purposes.

**2005YF** simulates for 2005 limited data availability in annual estimates. Only those values of variables are treated as basic data that are normally available annually, when using administrative data sources and surveys. The values of the remaining variables are treated as not available (framework cells are colored yellow (Y), meaning that those values are only dealt with to a limited extent in the Bayesian estimation). The latter variables have so-called tentative values, as explained in Section 6.3. They are derived with the help of a subset of formulas (F) of ratios and identities and 2005 ratio values, and are therefore the same as the corresponding values in the 2005GV framework.

**2006GV** includes a nearly complete data set for 2006, which is based on ‘estimates’ of the conventional SNA compilation. Nearly all framework cells are treated as basic data, except some that are not covered in the SNA conventional compilation. Values for these variables are derived indirectly, using 2005 data structures, and are not treated as basic data. This applies in particular to the breakdown for intermediate and final uses between local output and imported products.

**2006YF** is similar to the 2005YF framework, except that 2006 and not 2005 data are used as basic data. The values of the remaining variables without basic data are derived from the 2006 basic data with the help of the same ratios and identities that were used in the 2005YF framework. Identities not used in the latter derivation may not be equal to zero, and ratios that are not used may have values that are different from the 2005 values.

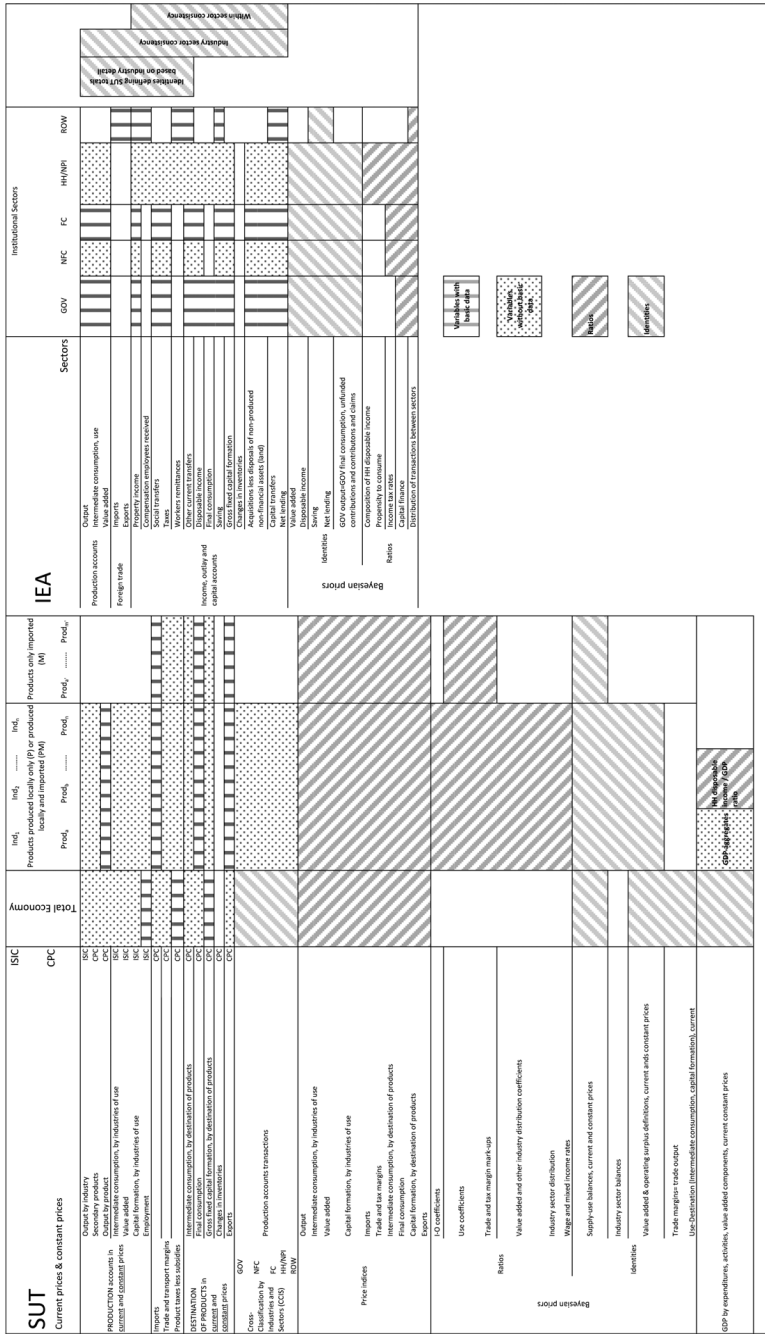


Fig. 1. BSCN framework, synoptic presentation.

**2006GF** uses the same formulas as in the 2006YF framework, but instead of treating the derived values as variables without basic data, they are treated as basic data (colored green), and are therefore also assigned prior reliabilities.

**2006YV(2005)** refers to the framework with basic data for 2006. If no basic data are available in 2006, we use 2005 data.

The coding of the framework names (GV, YF, YV, GF) is based on the color codes used in the INTERFACE programme, discussed in Section 5. Green (G) is used for variables with basic data and yellow (Y) for variables without basic data. When actual values are used this is indicated with V, and when ratio or identity formulas are used to derive the values of cells for which no data are available this is indicated with F.

All frameworks have identical formats, as presented in Fig. 1, in terms of the columns and rows for which transactor and transaction categories are presented. Each includes the two main data segments of the SNA, i.e. the Supply and Use Table (SUT) in current and constant prices and the Integrated Economic Accounts (IEA). In Fig. 1 the SUT is presented on the left-hand side and the IEA on the right-hand side. Both tables are matrices, cross-classifying ‘transactor’ categories in the columns with ‘transaction’ categories in the rows. Incorporation of IEA is emphasized in the BSNA framework, in order to facilitate a link with monetary analysis. Also included in the framework is the cross-classification between industries and sectors (CCIS), linking the SUT and IEA; it is presented below in Fig. 1 in the middle of the SUT segment.

Figure 1 only includes aggregate sectors, but in the extended BSNA framework used in the compilation, nine SNA sectors are distinguished:

- government (GOV);
- financial corporations (FC): Central Bank, deposit money banks, and other;
- non-financial corporations (NFC): public (NFCpu) and private (NFCpri);
- non-profit institutions serving households (NPI);
- households (HH); and
- rest of the world (ROW).

For the SUT the columns refer to a combination of industry (ISIC) and product (CPC) groupings. The format of the SUT differs from the 1993 SNA; see [22]. In this adapted SUT format, the supply and use product rows are transposed to column format and combined with the industry columns of the 1993 SNA-SUT format. Thus, supply and use data by CPC categories of products (imports and exports and final consumption and capital formation) and industry accounts by ISIC categories (output, intermediate consumption by use, value added, value added components, and employment) in current and constant prices are combined in the same column. This combination of CPC and ISIC categories in the columns is indicated by ISIC and/or CPC to the left of the SUT in Fig. 1. If ISIC is indicated, the row details concern industries, and if reference is made to the CPC, row details concern products. This distinction between industry and product categories is particularly relevant for output,

which is presented by industries and by products. If one industry only produces one product, industry and product categories are interchangeable, and output in terms of industries and products is identical. But this one-to-one correspondence does not hold if industries produce secondary products that are characteristic of other industries. The secondary products are presented in Fig. 1 between rows of output by industry and product, where they are classified by type of products in the rows and by industries producing those products in the columns. The totals of output by products in current and constant prices for each industry/product column are then derived by adding to industry output corresponding products produced by other industries in other columns and deducting secondary products produced by the industry in the column.

International standards on ISIC and CPC have been used in the framework. As a minimum two-digit breakdowns of ISIC and CPC have been incorporated, following common practice. A number of special features are reflected in the classification. When designing the country-specific classifications, there is close coordination between the details of ISIC industry and CPC-related product categories and also between sector and industry categories, so that each product category is assigned to a unique ISIC category and each ISIC category is assigned to a unique sector category. In the classifications thus designed, the one- to four-digit levels of the classifications may be combined, but the two-digit level is always maintained in support of international comparisons. Another feature is the introduction, within the product-related CPC classifications, of product distinctions that identify origins (output and imports) and destinations of products (intermediate consumption, final consumption, and capital formation). Introducing these product distinctions, which are based partly on the UNSD 1989 Classification by Broad Economic Categories see [21] considerably simplifies the supply-use identities and thus facilitate the reconciliation of data in which these identities play a role.

In some cases, special features of the SUT and IEA segments of the framework are introduced that differ from those of the conventional SNA. We mention four such deviations. First, within the destination of products in intermediate consumption, final consumption, and capital formation, a distinction is made between products originating from local output and imports, so that the import dependence of the economy can be measured in detail. Second, the framework includes many details on secondary output, which facilitate measuring the link between output by industry and product. Third, a broad set of price indices by product categories is included in the framework, linking the current and constant price product flows. Finally, the framework includes a link between GDP and household disposable income, as a means of refocusing aggregate analysis from sole dependence on the GDP production measure to a balanced focus which also includes an SNA measure of well-being. The latter is in line with the recommendations in [18]. Details of these special classification and SNA transaction features are discussed in [23].



## **4. Bayesian conditions in the BSNA framework**

### *4.1. Bayesian conditions*

The Bayesian conditions include four elements: basic data, ratios, identities, and the coefficients of variation of the basic data and ratio values. Ratios, ratio values, and identities are used in the benchmark compilation to ensure that the Bayesian estimates are compatible. In the annual compilation they are used to supplement limited data availability, thus facilitating the estimation of variables for which no basic data are available. The scope of ratios is the same in the benchmark and the annual compilation. The SUT ratios on the left-hand side of Fig. 1 include price indices, input-output coefficients, use coefficients, wage and mixed income rates, and value added and other industry distribution coefficients. The IEA ratios on the right-hand side include coefficients describing the composition of household disposable income, the finance of capital formation, and the distribution between the sectors of the IEA revenues and expenditures. The SUT identities on the left-hand side of Fig. 1 refer, among others, to CCIS-IEA identities for the production accounts, supply-use identities for products, identities defining value added and operating surplus, the identity between output of trade and transport, and the sum of trade and transport margins, as well as overall GDP identities. The identities in the IEA segment of the scheme include the identities defining disposable income, saving, and net lending of sectors. To the right of the IEA are included identities between totals and details of SUT rows, identities between revenues and expenditures in the IEA, and identities between variables of IEA and SUT.

In the annual scheme 2006YF there are 2719 variables, supported by 531 basic data, 1120 identities, and 2294 ratios. Thus, while only 19.5% of the variables are supported by basic data, per variable 1.45 information items are available. Such a large number of information items per variable allows checks to be made between basic data and priors, which is close to how conventional national accounts practices are carried out.

The difference with conventional accounting practices concerns the number of ratios and identities that can be taken into account, and the use of reliabilities. In conventional national accounting the number of assumptions is generally equal to the number of variables without basic data, while in the Bayesian estimation approach of SNAER any number of restrictions could be accommodated. (This feature of conventional national accounting is used in the derivation of so-called 'tentative' estimates prior to Bayesian estimation in the present method; see Section 6.3). The large number of restrictions in the Bayesian estimation approach serves as a means of checking the restrictions. This is a feature that is not explicitly available in conventional national accounting practices. Reliabilities are used implicitly when national accountants adjust data, but they are explicitly used in the Bayesian estimation, by adjusting prior values of variables and ratios more when their reliability is low (large standard deviation) and less when their reliability is high (small standard deviation).

If the number of basic data items increases, which will happen between preliminary and final estimation, the number of information items will grow and more checks will become available, while the number of ratios and identities will remain the same. In the ideal case, there will be basic data for all cells, and this is simulated by the 2005GV and 2006GV frameworks; see the first test in Section 6.2. In the case of the 2006GV framework, there are 2550 basic data, 1167 identities, and 2422 ratios for 3037 variables, which means that for each variable 2.02 information items are available for Bayesian estimation and checks. In the next sections the scope of basic data, ratios, and identities is reviewed in more detail, and also attention is paid to the reliabilities attached to those values. The focus is on the current estimates for the annual scheme 2006YF.

#### *4.2. Variables with and without basic data*

In the annual frameworks, eight types of basic data are assumed to be available.

These are presented in Table 1, which also includes annotations as to how they have been compiled, and indicates the prior reliability of each. This reliability, to be determined by the national accountant, is expressed as a variation coefficient, which is the inverse of the *t*-ratio. The table shows that the data that are considered to be most reliable (F, nearly fixed) are all data on output of goods and services based on economic surveys, the totals of exports and imports, total employment, the administrative data of the government, financial corporate, and rest of the world sectors, and also price indices. High (H) reliability is accorded to the total of household final consumption based on household surveys, as well as the detail of exports and imports based on foreign trade and Balance of Payments Statistics. Medium (M) reliability is accorded to the detail of household final consumption by products and the detail of employment by economic activities. In the benchmark and other schemes with more basic data, other prior reliabilities are assigned to other data. For example, superior (S) reliability is assigned to sector data of public non-financial corporations, and financial corporations other than Central Bank and deposit money banks, and low (L) reliability is assigned to production account items other than output, and also to trade and transport margins on products. Poor (P) reliability is not assigned to any item. All data in the benchmark scheme for the household, non-profit institution, and private non-financial corporation sectors are assigned low (L) reliability.

#### *4.3. Identities and ratios*

In addition to the basic data, the BSNA includes definitions of identities, and definitions and values of ratios.

Their location and specific functions are described in Table 2. The scope of identities and ratios does not differ between benchmark and annual schemes. Also included

Table 1  
Basic data in the BSNA framework

	Origin of data	Prior reliability
Output data	Output data in early estimates generally is available in constant prices, as derived from growth rates based on data from surveys that request data on output in physical units of selected establishments. The constant price data are then converted to current price data, following established procedures of national accountants. Both current and constant price data on output are included in the BSNA framework, even though only one of the data sets in current or constant prices is strictly needed, as the other can be easily derived with help of price indices.	<b>F</b> for total and detail by products
Import and export data	Import and export data are generally available in detail from foreign trade statistics (goods) and Balance of Payments (services).	<b>F</b> for total, <b>H</b> for detail by products
HH final consumption data	HH final consumption data are generally only available directly in years in which HH surveys are conducted. However, national accounts practices update the results of those surveys to future periods, using very detailed information on products, consumption weights (of CPI) and changes in the population between the years to arrive at revised estimates. As this detailed updating of the HH final consumption data cannot be well reconstructed in the present Bayesian estimation procedure (ratios and identities), the result of the national accounts updating is directly included in the BSNA framework as basic data.	<b>H</b> for total, <b>M</b> for detail by products
Employment data	Employment data are generally available from employment surveys, either or not included in HH surveys. National accountants often update this information to recent periods, using extrapolation based on demographic statistics.	<b>F</b> for total and <b>M</b> for detail by economic activities
GOV sector data	GOV sector data are directly based on data available in the GOV administrative records.	<b>F</b>
FC sector data	FC data are generally available from Central Bank reports (administrative data source) about the Central Bank and other banks and insurance companies. Other financial institutions, such as currency exchange houses, insurance agents and stock brokers, are generally estimated additionally by national accountants to complete the FC sector. All data including the additional estimates made by the national accountants are treated as basic data in the BSNA framework.	<b>F</b>
External sector data	External sector data are covered through periodic updates of the Balance of Payments by the Central Bank. Their conversion to SNA format is used to complete the basic data cells of this sector in the BSNA.	<b>F</b>
Price indices	Price indices are partly based on surveys of consumer prices, retail and wholesale trade surveys, producers price surveys. The remaining part of the indices is constructed by national accountants on the basis of these survey data. Both the price data from surveys and also the estimates made by national accountants are included in the BSNA framework as very reliable ratio information.	<b>F</b>

Table 2  
Identities and ratios (other than price indices) in the BSNA framework

Identities and ratios described	Prior reliability	Used in tentative estimates
<b>SUT-identities</b>		
Supply-use identities in current and constant prices for output and imports separately	X	
Identities between output by industries and product and secondary production		
Trade and transport identity, also in current and constant prices, between margin and output of trade and transport		
Identities of value added and operating surplus in current and constant prices	X	
Identities of HH final consumption by product between output and imports allocated to HH final consumption, and HH survey data in current prices		
Identities between totals of uses and import and output components	X	
Identity between intermediate consumption by industry of use and product in current and constant prices		
Identity between gross fixed capital formation by industry of use and product in current and constant prices		
Identities expressing totals as sum of details by CPC and ISIC categories in current constant prices	X	
GDP identities		
<b>IEA-identities</b>		
Identities between receipts and expenditures of sectors		
Identities defining in IEA: value added, operating surplus, GOV & NPI final consumption, social transfers, disposable income, saving, net lending		
Identities between industry and sector production accounts data		
<b>SUT-IEA-identities</b>		
<b>SUT-IEA-ratios</b>		
Coefficients of distribution of industry production account data to sectors	S	X
<b>SUT behaviorist or vertical ratios</b>		
Input output ratios in constant prices between (i) intermediate consumption and output, (ii) gross fixed capital formation by industries of use and output and (iii) employment and output (labour-productivity coefficient)	S	X
SUT secondary product ratios	H	X
User coefficients of output and imports	H	X
Wage and mixed income rates	M	X
Product tax rates	H	X
Trade and transport margins	H	X

Table 2, continued

Identities and ratios described	Prior reliability	Used in tentative estimates
<b>SUT</b> distributional ratios	S	
Coefficients of distribution of HH final consumption by products in SUT		
Coefficients of distribution of value added and product taxes between industries and products	S	
<b>IEA</b> behaviorist or vertical ratios	M	X
Finance of investment ratios		
Coefficients of components of HH disposable income	M	X
Propensity to consume of HH	M	X
Income tax ratios by sector	M	X
<b>IEA</b> horizontal distributional ratios		
IEA coefficients of distribution of revenues and expenditures between sectors	S for disposable income, saving and net lending L for other IEA distribution coefficients	
<b>SUT-IEA</b> aggregate ratios		
GDP growth rate		
Propensity to consume		
HH disposable income/GDP ratio		
PX/PM (Terms of trade)		

in the table are indicators of the prior reliabilities of ratio values. The highest reliability (superior, S) has been assigned to input-output coefficients, industry-sector distribution coefficients, coefficients of household final consumption distribution by products, and coefficients of distribution of value added by industry, while the lowest (L) reliability has been assigned to IEA coefficients of distribution of revenues and expenditures between sectors. No prior reliabilities have been assigned to aggregate ratios, i.e. GDP growth rate, propensity to consume, household disposable income/GDP ratio, and the terms of trade (PX/PM). The last column of the table identifies the ratios and identities that are used in the derivation of values for variables for which no basic data are available. They are referred to as ‘tentative’ estimates in Section 6.3, where they will be explained further.

The incorporation of ratios (and also identities) is much determined by the availability of data and the scope and design of the framework based thereon. This is clearly shown by comparing the design of our BSNA framework with the architecture of the US framework as presented in [7]. The latter framework includes data on the stocks of fixed assets and thus allows for the separate incorporation of rates of return and capital productivity ratios, which are not included in the present scheme. By incorporating constant price data up to saving it is possible to include not only growth rates of output (GDP), but also measures of the increase in the level of well-being.

## 5. Bayesian method and software

### 5.1. The Bayesian method

In contrast to the classical (frequentist) approach, a Bayesian does not assume ‘true’ parameters (latent variables)  $x$ . Instead, a probability distribution of the parameters is assumed, the so-called prior distribution. The data then serve to modify the prior idea of the ‘truth’ into a more complete idea: the posterior distribution. The mean of the posterior distribution can then be viewed as an ‘estimator’ of  $x$ , and the variance of the posterior distribution serves as a measure of its precision. When both the likelihood and the priors are based on the normal distribution, the posterior is normal as well, and therefore there is no mathematical difference between data and priors, although there is of course a conceptual difference. This simple observation leads to equivalences which are utilized in our software.

In this section we summarize the mathematics underlying our approach, and describe the two computer programs, SNAER (System of National Accounts Estimation and Reconciliation) and INTERFACE, which together provide estimates and precisions of the latent variables. Our problem is complex because we encounter matrices (data and restrictions) that are large and sparse. A matrix is ‘sparse’ when it has many structural zeros, and it is ‘large’ when we have, say,  $2^{11}$  variables and  $2^{13}$  observations, thus giving  $2^{24} \approx 16.8$  million entries in the design matrix.

5.2. *Data, priors, and linear restrictions*

In the formal statistical framework we consider a vector  $x$  consisting of  $n$  latent variables  $x_1, x_2, \dots, x_n$ . Data are available on  $p$  components (or linear combinations) of  $x$ . Let  $d_1$  denote the  $p \times 1$  data vector. Our starting point is a measurement equation,

$$d_1|x \sim N_p(D_1x, \Sigma_1), \tag{1}$$

which tells us that the conditional distribution of  $d_1$  given  $x$  is normal with a mean which is linear in  $x$  and a variance which does not depend on  $x$ . Typically, the  $p \times n$  matrix  $D_1$  is a selection matrix, say  $D_1 = (I_p, 0)$ , so that  $D_1x$  is a subvector of  $x$ , but this is not required. Neither is it required that the matrix  $D_1$  has full row-rank. Measurements are unbiased in the sense that  $E(d_1|x) = D_1x$ . The  $p \times p$  matrix  $\Sigma_1$  denotes a positive definite variance matrix, typically (but not necessarily) diagonal.

In addition to the  $p$  data, we have access to two further pieces of information: prior views concerning the latent variables or linear combinations thereof, and deterministic linear constraints. More specifically, we have  $m_1$  random priors:

$$A_1x \sim N_{m_1}(h_1, H_1) \tag{2}$$

and  $m_2$  exact restrictions (identities):

$$A_2x = h_2 \text{ (almost surely),} \tag{3}$$

in total  $m = m_1 + m_2$  pieces of prior information.

We assume that the  $m_1 \times m_1$  matrix  $H_1$  is positive definite (hence non-singular) and that the  $m_2 \times n$  matrix  $A_2$  has full row-rank  $m_2$  (so that the exact restrictions are linearly independent and thus form a consistent set of equations). We define

$$A = \begin{pmatrix} A_1 \\ A_2 \end{pmatrix}, \quad h = \begin{pmatrix} h_1 \\ h_2 \end{pmatrix}, \quad H = \begin{pmatrix} H_1 & 0 \\ 0 & 0 \end{pmatrix},$$

and assume that  $rk(A) = m$ , which implies of course that both  $A_1$  and  $A_2$  have full row-rank. The rank condition on  $A$  is not a serious restriction, because we can freely move priors to data (and vice versa). Hence the condition  $m \leq n$  is not restrictive either.

In order to identify all  $n$  variables from the information (data and priors) we need at least  $n$  pieces of information:  $m + p \geq n$ . But this is not sufficient for identification, because some of the information may be on the same variables. Necessary and sufficient for identification is the condition

$$rk \begin{pmatrix} A \\ D_1 \end{pmatrix} = n,$$

which is automatically satisfied when  $m = n$ .

5.3. Estimation: The SNAER software

There are several equivalent ways to estimate the components of  $x$  and their variances [3]. The equivalence is based on two facts. First, a Bayesian analysis with normal data and normal priors is closely linked with a quadratic minimization problem. Second, best linear unbiased estimation is closely linked to quadratic minimization (least squares). A Bayesian solution is provided in Theorem 1 of [10], but it involves Moore-Penrose inverses and is not easily computable for large sparse systems. An easier, but equivalent, solution is obtained by using the close relationship between best linear unbiased estimation and least squares [13,14]. Defining

$$D = \begin{pmatrix} D_1 \\ A_1 \end{pmatrix}, \quad d = \begin{pmatrix} d_1 \\ h_1 \end{pmatrix}, \quad H = \begin{pmatrix} \Sigma_1 & 0 \\ 0 & H_1 \end{pmatrix},$$

we obtain estimates of  $x$  by solving the constrained problem

$$\begin{aligned} &\text{minimize } (d - Dx)' \Sigma^{-1} (d - Dx) \\ &\text{subject to } \quad A_2 x = h_2. \end{aligned} \tag{4}$$

This can be simplified by writing

$$A_2 = (A_{21} : A_{22}),$$

where  $A_{21}$  is an  $m_2 \times (n - m_2)$  matrix and  $A_{22}$  is a *non-singular*  $m_2 \times m_2$  matrix. Partitioning  $x$  correspondingly, we can write the restriction as

$$A_{21}x_1 + A_{22}x_2 = h_2,$$

so that

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} I_{n-m_2} \\ -A_{22}^{-1}A_{21} \end{pmatrix} x_1 + \begin{pmatrix} 0 \\ A_{22}^{-1}h_2 \end{pmatrix} = Qx_1 + q.$$

Let

$$B = \Sigma^{-1/2}D = \begin{pmatrix} \Sigma_1^{-1/2}D_1 \\ H_1^{-1/2}A_1 \end{pmatrix}, \quad b = \Sigma^{-1/2}d = \begin{pmatrix} \Sigma_1^{-1/2}d_1 \\ H_1^{-1/2}h_1 \end{pmatrix}.$$

Then the constrained problem Eq. (4) can be written as the unconstrained problem

$$\min_{x_1} \|(b - Bq) - BQx_1\|^2,$$

which in turn can be rewritten as

$$\min_{x_1} \left( \frac{1}{2} x_1' \Gamma x_1 - c' x_1 \right), \tag{5}$$



where  $\Gamma = Q'B'BQ$  and  $c = Q'B'(b - Bq)$ . This is the format in the SNAER software, which relies on Harwell's HSL\_VF06 procedure which, in turn, is based on the Gould-Nocedal algorithm; see [5].

The Harwell routine does not provide the variance matrix, elements of which can be computed as follows. The  $j$ -th column  $v^{(j)}$  of the matrix  $V = \Gamma^{-1}$  can be found by minimizing  $\|\Gamma^{-1/2}e^{(j)} - \Gamma^{1/2}v^{(j)}\|^2$  for all  $j$ , that is, by solving the quadratic minimization problem

$$\min_{v^{(j)}} \left( \frac{1}{2} v^{(j)'} \Gamma v^{(j)} - e^{(j)'} v^{(j)} \right) \quad (j = 1, \dots, n - m_2),$$

using the Gould-Nocedal algorithm, where  $e^{(j)}$  denotes a vector all whose components are zero except the  $j$ -th component which is one. The value at the minimum equals  $-v_j^{(j)}/2$ . This method is specifically designed for sparse systems.

Thus we solve the constrained problem Eq. (4) in two steps. First we identify and invert a non-singular submatrix  $A_{22}$  from  $A_2$ ; then we solve Eq. (5). The two-step procedure has the advantage that the dimension of the system is much reduced. Moreover, the first step can be done once and then the results of the reduction may be used for many restricted least-squares problems. In particular, the  $A_2$  matrix is usually fixed because it represents the structure of the economy, while the matrices related to  $A_1$  are priors that will vary.

In our case, some of the priors are so-called indicator ratios. These priors are non-linear and hence need to be linearized. Suppose we have a prior indicator ratio  $R = y/x \sim (r, \tau^2)$ . We wish to replace this non-linear prior by its linearization  $y - rx \sim (0, \omega^2)$ . The question is how to choose  $\omega^2$ . This question is discussed in [4], Section 6, where an invariant linearization method is proposed. Invariance here means that we obtain the same prior whether we start from  $y/x$  or from  $x/y$ .

Summarizing, the information in our system of latent variables to be estimated consists of incomplete data (with precisions), priors on a subset of the variables or linear combinations thereof (with precisions), and exact linear restrictions. Our system is large and sparse. It is large, because we may have  $2^{11}$  variables and  $2^{13}$  observations, thus giving  $2^{24} \approx 16.8$  million entries in the design matrix. It is sparse, because information is often available on one variable at a time, and restrictions are often definitions involving only a small number of variables. The SNAER software, especially developed for this project and described in some detail in [4], provides estimates and variances (and desired covariances) of the complete  $x$ -vector. The method takes full account of all accounting identities, the solutions are continuous rather than discrete, and multiple priors on variables or linear combinations of variables are allowed. The posterior estimates take all prior and data input into account and come with precisions. The system is transparent, flexible, and fast.

#### 5.4. The INTERFACE

The operational aspects and the mathematics of the approach are reflected in two computer programmes: SNAER and INTERFACE. The SNAER program was

discussed in the previous subsection. It requires inputs (data, prior indicator ratios, identities, and prior reliabilities) in a specific format. The INTERFACE converts the inputs of the Excel worksheets into a format that is accepted by SNAER. In the process of this conversion, INTERFACE carries out a number of checks on the data and Bayesian inputs, before inputting this information into SNAER for further processing.

The INTERFACE is written for Excel spreadsheets. It extracts information of any framework that is defined in Excel. Thus, it reads and extracts the values of basic data, and of ratio formulas and values included in the framework files. The formulas in those cells are recorded by INTERFACE and converted to a format that can be read in SNAER. It reads and extracts separately the values of the ratio cells. If those values are not applicable, it reads and extracts instead (with preference) ratio value information that is included in Excel comments attached to the (purple-colored) cells. Furthermore, it reads and extracts the formulas of identities that are identified in the Excel framework with the help of (blue-colored) cells, and also converts those formulas to a format that can be read in SNAER. Finally, it reads and extracts information on reliabilities of basic data and values of ratios.

The reliabilities are expressed as 'fixed' (F), 'superior' (S), 'high' (H), 'medium' (M), 'low' (L), or 'poor' (P), and they are expressed in the framework as percentages of coefficients of variation. Thus, if the reliability of a basic data item or ratio value in the BSNA framework is poor, it may deviate in the final Bayesian estimates 24% from the original value. Similarly, it may deviate 12% if the reliability is 'low', 6% for medium reliability, 3% for high reliability, and 1% for superior reliability.

In the process of extracting the information for use in SNAER, the INTERFACE also carries out data checks and presents error messages if necessary. The three most important checks are the following. First, identities should include at least one variable of the framework. (This limitation to one variable is included for practical purposes, as many of the identities are defined as Excel sums, in which many cells are not variables in the system.) Second, both the numerator and the denominator of ratios should be available as variables in the framework. Third, all basic data and ratio values must have been assigned reliabilities.

After all error messages have been cleared, the INTERFACE transfers data to the SNAER input files. After SNAER has calculated estimates and precisions, these are converted back into the format of the framework, so that the user can assess the resulting estimates in their framework context.

## **6. Tests of framework**

### *6.1. Comparison of framework options*

We now present various Bayesian estimates for the years 2005 and 2006. The estimates – referred to as options in Table 3 – enable us to assess the impact of

Table 3  
Impact of alternative compilation options on Bayesian estimates

Compilation options	Bayesian conditions of options	Nr. of distortions in Bayesian estimates			Impact of options on Bayesian estimates		
		Nr. of estimates > 2% change	Nr. of estimates with posterior coefficients of variation > 1%	Types of impacts measured	Impact measured by comparing following options	Reference to tables and sections, where impacts are quantified/ assessed	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	
Option 1	2005GV	0	3	Impact of changes, i.e. increases and decreases of prior coefficients of variation, with full 2005 benchmark data	comparing options 2, 3 and 4 with 1, and also reviewing options 5 and 6	Table 8, Section 7.4	
Option 2	2005GV, 50% decrease of prior coefficients of variation	0	4				
Option 3	2005GV, 50% increase of prior coefficients of variation	0	11				
Option 4	2005GV, fictitious prior coefficients of variation (= 100%)	0	64				
Option 5	2006YF, poor (P) reliability of price indices	10	11				
Option 6	2006YF, 50% increase of prior coefficients of variation	13	5	Impact of increases of prior coefficients of variation, with limited 2006 annual data			
Option 7	2005GV, only identities included	0	1	Impact of Bayesian conditions, i.e. identities, price indices, other ratios	comparing options 7 and 8 with 1	Table 4, Section 6.2	
Option 8	2005GV, identities and price indices included	10	1				
Option 9	2005GV, SUT only	0	6	Impact of only compiling SUT or IEA	comparing options 9 and 10 with 1	Table 7, Section 7.1	
Option 10	2005GV, IEA only	2	15				
Option 11	2006YV(2005), 2005 values for missing data	46	6	Impact of using assumptions, i.e. ratios and identities to derive tentative values for missing data	comparing options 14 and 11	Table 4, Section 6.2	

Table 3, continued

(1)	(2)	Bayesian conditions of options		Nr. of distortions in Bayesian estimates		Impact of options on Bayesian estimates	
		Nr. of estimates > 2% change	Nr. of estimates with posterior coefficients of variation > 1%	Nr. of estimates with posterior coefficients of variation > 1%	Types of impacts measured	Impact measured by comparing following options	Reference to tables and sections, where impacts are quantified/ assessed
(1)	(2)	(3)	(4)	(5)	(6)	(7)	
Option 12	2005YF	15	3	Impact lesser availability of basic data in 2005 benchmark year	comparing option 12 with 1	Table 6, Section 6.4	
Option 13	2006GV	5	5	Impact lesser availability of basic data in 2006 current year	comparing options 13 and 14		
Option 14	2006YF	13	3				
Option 15	2006GF, tentative values for missing data treated as basic data	22	2	Impact of assigning basic data status to tentative estimates	comparing options 14 and 15	Table 4 Section 6.2	
Option 16	2006YF, poor quality of household final consumption data and ratios	13	2	Impact of poor quality/non-availability respectively of data on HH final consumption, imports-exports and external sector, services, financial corporations	comparing options 16, 17, 18, 19 with option 20	Table 7, Section 7.2	
Option 17	2006YF, poor quality of imports, exports and external sector data importsexports	15	9				
Option 18	2006YF, poor quality data on financial corporations	32	2				
Option 19	2006YF, poor quality of data on household final consumption restrictions	23	4				
Option 20	2006YF, option 14 with full household final consumption restrictions	11	2				
Option 21	2006YF, aggregate variables and basic data, adjusted aggregate reliabilities	34	Not comparable	Impact of using aggregates in compilation	comparing options 21 with 14	Table 7, Section 7.3	

changing the conditions embedded in the framework on the Bayesian estimates. These impacts as analyzed below and in Section 7.

Table 3 summarizes the results of twenty-one different options. The options are described briefly, highlighting deviations from the six standard frameworks described in Section 3. The quantitative assessment of the estimates in subsequent tables is done in terms of major aggregates of GDP and household disposable income, even though the underlying materials would permit making the assessments for all details of variables and ratios in the BSNA framework. There are seven columns in Table 3. The base year is 2005 and the current year is 2006.

In column 2 we highlight the deviations from one of the four standard frameworks: options 1, and 12–14. The standard format for 2005 (option 1) and 2006 (option 13) includes data for all cells, which are treated as basic data. These include changes in inventories and a full set of secondary products. Standard prior precisions are applied, as presented in Tables 2 and 3. The 2005 and 2006 ‘data’ used in the two frameworks are different. The 2005 benchmark data are based on the conventional practice of using available data in a detailed compilation process in order to establish new data structures for use in future compilations. The 2006 full data set is also based on conventional practices, but applying a more limited data analysis, and using mainly data structures of 2005 to estimate the values of the remaining variables and ratios. The standard format of the annual 2005 or 2006 frameworks (options 12 and 14) includes only a limited data set that is annually available on output, employment, imports and exports, price indices, and data on the government, financial corporate, and rest of the world sectors. The annual frameworks do not include changes in inventories and include only a limited set of data on secondary products. Assumptions based on selected ratios and identities are used to arrive at tentative estimates for missing data. Only annually available data are treated as basic data. Values derived on the basis of assumptions are not considered available and therefore not treated as basic data (except in the 2006GF framework of option 15); they are, however, used in the linear approximation of ratios. Standard precisions are applied to basic data and ratio values.

Columns 3 and 4 present the number of ‘distortions’ in the Bayesian estimates. These are assumed to occur in two instances. First, when Bayesian estimates differ more than 2% in absolute terms (column 3) from the conventional estimates in the full 2005GV (option 1) and 2006GV (option 13); second (column 4) when the posterior coefficient of variation is larger than 1%. The latter case is based on the fact that posterior precisions in terms of coefficients of variation are typically much smaller than 1%.

The types of measured impacts are listed in column 5. They range from testing the impact of compilation assumptions, lesser data availability through surveys and administrative data sources, to lesser data availability on an annual basis. The impacts are measured by comparing the Bayesian estimates of alternative options, listed in column 6. For example, the impact of having less data annually for the current year 2006 is measured by comparing Bayesian estimates of options 14 and 13. Also, the

impact of compiling only the SUT or only the IEA, can be assessed by comparing options 9 and 10 with option 1. A quantitative comparison between the alternative Bayesian results of these options is presented in the tables and sections referred to in column 7.

## *6.2. Tests of framework assumptions*

The BSNA framework is very complex, because of the large number of variables, ratios, and identities. The application of the Bayesian estimation approach to such a complex framework was therefore carefully tested. The results of the most important tests are presented in Table 4.

The first test was to check the internal consistency of ratios and identities in the framework. The test was carried out by applying SNAER and INTERFACE to the 2005GV benchmark data (option 1), and determine whether the Bayesian estimates are close to the conventional estimates. This should be the case, because data, data structures, and identities are fully compatible for that year. The only possible remaining distortion is the effect of the prior reliabilities. This is the additional information that was not (explicitly) available in the 2005 basic data compiled by conventional national accounts methods. The data in the first columns of Table 4 show that there are no distortions in the Bayesian estimates, and that in only three cases there are distortions for the posterior coefficients of variation. Two of these (household income taxes, paid and household other current transfers, received less paid; not shown in the table) are caused by the prior reliabilities, which may indeed be incompatible with the implicit reliabilities of the national accountants who did the benchmark compilation. A similar test for the full 2006GV framework, presented as option 13 in Table 6, shows slightly different results. In particular, the Bayesian estimates of gross fixed capital formation and changes in inventories (not shown in table) present significant changes from the conventional estimates. Also, changes in inventories have a posterior variation coefficient which is larger than 1%. The reason for these distortions may be that not all cells in the 2006GV scheme are treated as basic data. In particular, all output and import subitems of final and intermediate uses are missing (not estimated in conventional national accounting) and 2005 structures are used to estimate these variables.

The second test was to measure the impact of alternative values for the variables in the framework for which no basic data are available. The test was applied to the 2006YF framework. As explained earlier, values in cells without basic data are not used in the Bayesian estimation, except in the linearization of ratios. The results are shown in columns 3–6 of the table. In option 11, the 2005 benchmark data are included in the cells without basic data, and in option 14 so-called ‘tentative’ estimates are included in those cells. The latter are derived with the help of assumptions based on a selection of identities and ratios of 2005 structures; see also Section 6.3. The results of the test are convincing: In option 11 the number of distortions of estimates

Table 4  
Testing alternative framework assumptions

	Testing the application to benchmark data (2005GV)		Testing alternative values for variables without basic data (2006GF)		Testing alternative scopes of basic data (2006G)		Testing the accumulative use of Bayesian inputs (2005GV)							
	(D)	(E)	(F)	(G)	(H)	(I)	(J)	(K)	(L)					
	Bayesian estimates of 2005GV option 1		Using values of base year option 11		Using assumptions (selected ratios and identities) to derive alternative data option 14		Treating only selected variables as variables with basic data option 15		Identifying and price indices option 8					
	Posterior coefficient of variation		% Difference from conventional estimates		Posterior coefficient of variation		% Difference from conventional estimates		Posterior coefficient of variation		% Difference from conventional estimates			
GP ECONOMIC ACTIVITIES, CONSTANT	0.0%	0.1%	1.2%	0.10%	0.0%	0.07%	0.0%	0.07%	-1.4%	0.07%	0.0%	0.00%	1.5%	0.00%
GP EXPENDITURES, CONSTANT	0.0%	0.1%	1.2%	0.10%	0.0%	0.07%	0.0%	0.07%	-1.4%	0.07%	0.0%	0.00%	1.5%	0.00%
Agriculture, forestry and fishing	0.0%	0.1%	13.1%	0.23%	1.4%	0.10%	1.4%	0.10%	1.1%	0.10%	0.0%	0.01%	0.3%	0.01%
Mining and quarrying	0.0%	0.1%	21.5%	0.16%	0.1%	0.12%	-0.1%	0.12%	-0.1%	0.12%	0.0%	0.05%	0.1%	0.06%
Manufacturing	0.0%	0.2%	-13.9%	1.04%	0.8%	0.26%	-0.6%	0.26%	-3.8%	0.25%	0.0%	0.02%	3.3%	0.02%
Construction	0.0%	0.1%	20.6%	0.40%	2.8%	0.57%	-7.8%	0.57%	-7.8%	0.55%	0.0%	0.01%	4.7%	0.01%
Wholesale and retail trade	0.0%	0.2%	4.5%	0.27%	1.9%	0.22%	1.9%	0.22%	-2.2%	0.22%	0.0%	0.00%	2.5%	0.00%
Land transport	0.0%	0.5%	-28.9%	0.88%	1.5%	0.47%	1.5%	0.47%	1.0%	0.46%	0.0%	0.01%	1.3%	0.01%
Transport (other), storage, telecommunication	0.0%	0.3%	38.1%	0.25%	1.1%	0.29%	-1.1%	0.29%	-2.5%	0.29%	0.0%	0.02%	1.3%	0.02%
Financial and insurance activities	0.0%	0.0%	-3.0%	0.25%	0.4%	0.11%	-0.4%	0.11%	-0.3%	0.03%	0.0%	0.02%	0.0%	0.02%
Public administration and defence; compulsory social security	0.0%	0.0%	2.3%	0.06%	0.4%	0.04%	-0.4%	0.04%	0.2%	0.02%	0.0%	0.01%	0.0%	0.01%
VALUE ADDED, GROSS, TOTAL, CURRENT	0.0%	0.0%	2.2%	0.05%	-0.1%	0.01%	-0.1%	0.01%	-1.7%	0.01%	0.0%	0.00%	1.4%	0.00%
GP ECONOMIC ACTIVITIES, CURRENT	0.0%	0.0%	2.2%	0.05%	-0.1%	0.01%	-0.1%	0.01%	-1.7%	0.01%	0.0%	0.00%	1.4%	0.00%
Household final consumption	0.0%	0.0%	15.3%	0.09%	2.7%	0.00%	2.7%	0.00%	-0.2%	0.00%	0.0%	0.00%	2.0%	0.00%
Government final consumption	0.0%	0.0%	-1.2%	0.03%	0.2%	0.04%	-0.2%	0.04%	0.2%	0.11%	0.0%	0.05%	0.0%	0.05%
Mixed consumption	0.0%	0.0%	1.2%	0.03%	0.2%	0.04%	1.2%	0.04%	0.2%	0.11%	0.0%	0.05%	0.0%	0.05%
Gross fixed capital formation	0.0%	0.2%	0.0%	0.21%	5.4%	0.20%	-5.4%	0.20%	-3.1%	0.18%	0.0%	0.00%	-1.3%	0.00%
Exports	0.0%	0.0%	2.9%	0.09%	0.7%	0.05%	-0.7%	0.05%	-0.3%	0.05%	0.0%	0.00%	0.0%	0.00%
Imports	0.0%	0.0%	-5.3%	0.06%	1.1%	0.05%	1.1%	0.05%	0.3%	0.03%	0.0%	0.00%	0.0%	0.00%
GP EXPENDITURES, CURRENT	0.0%	0.0%	2.2%	0.05%	-0.1%	0.01%	-0.1%	0.01%	-1.7%	0.01%	0.0%	0.00%	1.4%	0.00%
Compensation of employees	0.0%	0.1%	38.5%	0.65%	0.6%	0.18%	0.6%	0.18%	4.6%	0.17%	0.0%	0.00%	2.8%	0.00%
Net operating surplus	0.0%	0.2%	-38.5%	0.65%	0.6%	0.18%	-0.6%	0.18%	-4.6%	0.17%	0.0%	0.00%	-2.8%	0.00%
Non-HP operating surplus	0.0%	0.2%	44.0%	0.11%	0.6%	0.18%	-0.6%	0.18%	-1.0%	0.17%	0.0%	0.00%	0.0%	0.00%
HP operating surplus	0.0%	0.1%	4.4%	0.13%	0.2%	0.14%	0.2%	0.14%	0.0%	0.14%	0.0%	0.00%	0.0%	0.00%
Mixed income, gross	0.0%	0.2%	-1.4%	0.33%	0.8%	0.22%	-0.8%	0.22%	-4.6%	0.22%	0.0%	0.00%	2.1%	0.00%
VALUE ADDED, GROSS, TOTAL, CURRENT	0.0%	0.0%	2.2%	0.05%	-0.1%	0.01%	-0.1%	0.01%	-1.7%	0.01%	0.0%	0.00%	1.4%	0.00%
HOUSEHOLD DISPOSABLE INCOME, GROSS	0.0%	0.1%	9.5%	0.09%	1.1%	0.08%	1.1%	0.08%	0.4%	0.07%	0.0%	0.00%	1.7%	0.00%
GP growth rate 2006/2005			1.3%	0.0%			0.0%		-1.5%				0.0%	0.00%
Implicit price Index Value Added, total			100.0%	104.7%			100.0%	104.7%					0.0%	0.00%
Propensity to consume			94.1%	93.2%			94.1%	93.2%					0.0%	0.3%
HP disposable income/GDP ratio			92.3%	93.3%			92.3%	93.3%					0.0%	0.2%
HP/GDP (Index of Trade)			100.0%	98.2%			100.0%	98.2%					0.0%	0.0%

was 46, while in option 14 there were only 13 distortions. Also the number of distortions in coefficients of variation was less: a reduction from 6 to 3.

The third test was to determine whether cells without basic data (and therefore with formulas, YF) should be treated as variables without basic data, or alternatively as basic data, with reliabilities attached to those. The test was applied to the 2006 framework in options 14 (2006YF) and 15 (2006GF), and the results are presented in columns 7–10. It is clear that the alternative 2006YF is preferred, as it has less distortions than the 2006GF version (13 compared to 22) in the values of the estimates, while posterior coefficients of variation do not differ significantly between the two options. Based on tests 2 and 3 we conclude that option 14 (2006YF), in which formulas (F) are used to estimate the missing data (Y), is the preferred option for producing estimates for a current year.

The fourth test was to identify separately the impact of different Bayesian inputs, i.e. identities, price indices, and other ratios. The results are presented in options 7 and 8 (columns 11–14), which should be compared with option 1 in columns 1 and 2. When using only identities, the Bayesian estimates of all aggregates are, as expected, close to the conventional estimates (0.0% difference), while there are hardly any distortions in coefficients of variation. When adding price indices to this as an additional input, a few components of GDP by activity (manufacturing, construction, wholesale and retail trade) and household disposable income (mixed income gross and non-household operating surplus, gross, not shown in table) present significant deviations (more than 2%), while posterior coefficients of variation do not change significantly. When adding other ratios in option 1 the  $> 2\%$  deviations in values of components of GDP and household disposable income disappear, but three distortions ( $> 1\%$ ) in posterior coefficients of variation occur. Thus, price indices have large impacts, particularly on the coefficients of variation, but other ratios neutralize this impact. The significant price impact was confirmed in Table 8, where in option 5 of the 2006YF framework the reliability of price indices was significantly reduced from S (superior) to P (poor). When comparing the results of that option with those of option 14 (columns 10 and 11), the number of distortions of coefficients of variation increased significantly (from 3 to 11), but at the same time, somewhat unexpectedly, the number of distorted values of variables decreased (from 13 to 10).

The impacts of alternative options on five analytical indicators are also measured in this and subsequent tables: growth of GDP, implicit price deflator of total value added, propensity to consume of households (ratio of household final consumption and household disposable income), ratio of household disposable income and GDP, and the terms of trade change measured by the ratio between export and import price indices; see [15] for more comprehensive measures of terms of trade. Significant impacts were identified in option 11 (using 2005 values for variables not available in 2006) for the last three indicators, but not on GDP growth, and in option 15 (treating all values in 2006 as basic data) for the ratio of household disposable income to GDP.



### 6.3. Bayesian versus tentative estimates

In the previous subsection we saw that 2006 estimates improved if assumptions (selected ratios and identities) were used to estimate the variables for which no basic data were available. The estimates for variables without basic data, together with the basic data are called ‘tentative’ estimates. Using assumptions to arrive at values for the variables without basic data generates values that are more realistic than when assigning benchmark values to those cells, and this method is also close to the method used by national accountants.

The example presented in Fig. 2, which represents a simple economy, may clarify the relation between tentative and Bayesian estimates. The figure includes three different versions of a data framework: The left panel is the framework for benchmark year  $t$ ; the middle panel includes tentative estimates for the current year  $t + 1$ ; and the right panel is a framework with Bayesian estimates for the same current year  $t + 1$ .

In each of the three frameworks there are 10 aggregates for output (P), imports (M), intermediate consumption (I), exports (X), value added (Y), gross fixed capital formation (K), final consumption (C), disposable income (R), domestic saving (S), and external saving (B). In addition, there are 6 identities and 5 ratios. In the left panel framework, the included identities are:

- Value added identity ( $E5 - E9 - E13 = 0$ ),
- Supply-use identity ( $(E5 + K5) - (E9 + G9 + I9 + K9) = 0$ ),
- Income distribution identity ( $E13 - I13 = 0$ ),
- Saving identity ( $I13 - I9 - I16 = 0$ ),
- Finance of capital formation identity ( $G9 - (I16 + K16) = 0$ ),
- External deficit identity ( $K5 - K9 - K16 = 0$ ),

and the included ratios are:

- input-output ratio ( $E9/E5$ ),
- import-output ratio ( $K5/E5$ ),
- capital-output ratio ( $G9/E5$ ),
- propensity to consume ratio ( $I9/I13$ ),
- export-output ratio ( $K9/E5$ ).

All identities are equal to zero in the benchmark scheme, which means that the values of the 10 variables satisfy those identities. All ratio values in the benchmark scheme are considered to be the structural ratio values that do not only hold in the benchmark year, but can also be used as assumptions in the estimation for the current year.

The second scheme (middle panel) includes tentative estimates for all variables, based on available basic data and a selection of ratios and identities. In the derivation of tentative estimates, only those ratios that are used have the same value as in the base year, and only those identities that are used have a zero value in the second scheme. The basic values that are included in the second scheme are those for output (P), imports (M), exports (X), and gross fixed capital formation (K). Values for

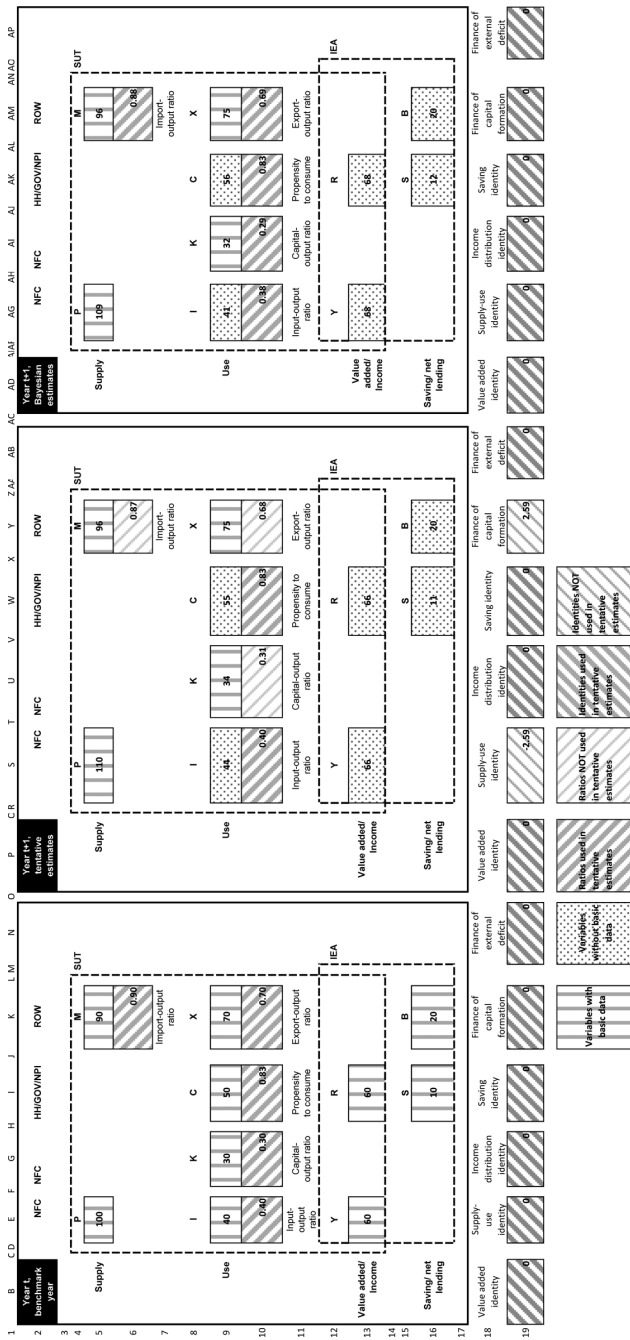


Fig. 2. Comparison of tentative and Bayesian estimates in the BSNA framework, an example.

variables without basic data are derived with the help of assumptions, represented by a selection of ratios and identities, as follows: Intermediate consumption (I) is derived from output (P) with the help of the input-output ratio of the base year (E10); value added is derived from output (P) and intermediate consumption (I) with the help of the value added identity (P19); disposable income (R) is derived from value added (Y) with the help of the income distribution identity (U19); final consumption (C) is derived from disposable income (R) with the help of the propensity to consume ratio of the base year (I10); domestic saving is derived from disposable income (R) and final consumption (C) with the help of the saving identity (W19); and external saving (B) is derived as the difference between imports (M) and exports (X) with the help of the finance of external deficit identity (AB19).

Thus, four basic data, two ratios, and four identities (that is, precisely ten items of prior information) are used to arrive at tentative estimates of the ten variables in the current year  $t + 1$ . The estimates of the variables with basic data are equal to the values of those basic data, and the values of the ratios used for the tentative estimates are equal to the values in the first scheme. However, identities that are not used are not necessarily equal to zero (S19 supply-use and Y19 finance of capital formation identities), and ratios that are not used have values in the tentative estimates that are different from those of the base year (capital-output ratio U10, import-output ratio Y6, and export-output ratio Y10). Hence, the tentative estimates are not compatible with all identities and ratios of the scheme.

In the third scheme this incompatibility between estimates and the not-used identities and ratios has been repaired with the help of the Bayesian approach, which uses *all* ratios and identities. In addition, reliabilities of basic data and ratios are taken into account, so that basic data and ratio values that are less reliable are adjusted more than more reliable data and values. As a consequence, estimates differ between the second and third schemes. For example, we obtain  $I = 44$ ,  $C = 34$ ,  $K = 55$ , and  $X = 75$  in the second scheme, and  $I = 41$ ,  $C = 32$ ,  $K = 56$ , and  $X = 75$  in the third scheme. The differences between the values of variables between the second and third schemes are not large. Thus, small changes in the values of variables (and ratios), as compared to the second framework, make it possible to satisfy all identities, also those that were not satisfied in the second scheme.

The same principles are used in the main BSNA, when making estimates for a current year (2006): first tentative estimates and then Bayesian estimates. The last column of Table 3 shows which of the identities and ratios of the main BSNA scheme used in producing tentative estimates. They include most (within columns or vertical) behaviorist ratios (e.g. input-output coefficients, user coefficients, coefficients of components of household disposable income) in SUT and IEA, and exclude (across columns or horizontal) distributional ratios of value added in SUT and coefficients of distribution across revenue and expenditure items in IEA. We use vertically-defined identities, such as supply use identities, or identities defining value added and operating surplus in SUT and IEA, while horizontally-defined identities in IEA between revenues and expenditures are not used. This implies that in a scheme for

Table 5  
Bayesian versus tentative estimates

	Tentative estimates 2006 VF	% Difference tentative and conventional estimates	Bayesian estimates, 2006VF option 14	% Difference Bayesian and conventional estimates	Prior Coefficient of Variation	Posterior Coefficient of Variation option 14
	(1)		(2)	(3)	(4)	(5)
Nr. of distorted values						
GDP ECONOMIC ACTIVITIES, CONSTANT	219,355,723,268	0.2%	218,918,539,581	0.0%	3.00%	0.07%
GDP EXPENDITURES, CONSTANT	217,212,744,301	-0.8%	218,918,539,581	0.0%	3.00%	0.07%
Agriculture, forestry and fishing	25,909,930,997	0.2%	26,209,217,770	1.4%	12.00%	0.10%
Mining and quarrying	3,315,810,341	1.4%	3,265,850,268	-0.1%	12.00%	0.12%
Manufacturing	42,276,263,192	-1.9%	42,830,905,154	-0.6%	12.00%	0.26%
Electricity, gas and water supply	5,954,046,559	2.6%	5,942,627,996	2.4%	12.00%	0.36%
Construction	12,137,992,674	2.8%	10,890,498,320	-7.8%	12.00%	0.57%
Wholesale and retail trade	33,763,471,805	-0.3%	34,532,219,229	1.9%	12.00%	0.22%
Land transport	4,088,573,858	6.3%	3,904,397,446	1.5%	12.00%	0.47%
Transport (other), storage, telecommunication	11,344,152,665	0.1%	11,210,364,867	-1.1%	12.00%	0.29%
Financial and insurance activities	6,529,536,200	0.4%	6,482,752,094	-0.4%	1.00%	0.11%
Private services	58,130,636,364	0.1%	57,985,829,092	-0.1%	12.00%	0.02%
Public administration and defence; compulsory social security	14,664,192,249	0.0%	14,603,034,376	-0.4%	0.10%	0.04%
VALUE ADDED GROSS, TOTAL, CURRENT	218,114,606,905	0.0%	217,857,696,610	-0.1%	12.00%	0.00%
GDP by ECONOMIC ACTIVITIES, CURRENT	229,859,500,562	0.0%	229,574,850,388	-0.1%	3.00%	0.01%
Household final consumption	199,739,842,843	0.0%	205,142,806,802	2.7%	3.00%	0.00%
Government final consumption	19,237,195,482	0.0%	19,193,854,559	-0.2%	0.10%	0.01%
NPISH final consumption	1,939,433,975	-1.3%	1,976,245,004	0.5%	6.00%	0.45%
Gross fixed capital formation	46,019,465,063	-0.4%	43,740,744,068	-5.4%	3.00%	0.20%
Exports	57,302,114,005	0.0%	56,874,753,165	-0.7%	0.10%	0.09%
Minus: Imports	96,270,197,607	0.0%	97,353,553,210	1.1%	0.10%	0.05%
GDP EXPENDITURES, CURRENT	228,140,677,515	-0.7%	229,574,850,388	-0.1%	3.00%	0.01%
Compensation of employees	72,756,461,528	0.2%	73,112,111,229	0.7%	12.00%	0.14%
Other taxes less subsidies on production	1,372,143,961	9.9%	1,257,100,901	0.6%	0.10%	0.10%
Non-HH operating surplus, gross	76,980,086,024	0.6%	76,089,876,207	-0.6%	12.00%	0.18%
HH Operating surplus (on dwelling services), gross	19,504,504,390	0.1%	19,514,812,150	0.2%	12.00%	0.14%
Mixed income, gross	47,580,064,856	-1.4%	47,883,796,123	-0.8%	12.00%	0.22%
VALUE ADDED GROSS, TOTAL, CURRENT	218,114,606,905	0.0%	217,857,696,610	-0.1%	12.00%	0.00%
HOUSEHOLD DISPOSABLE INCOME, GROSS	221,122,537,153	3.1%	216,780,833,217	1.1%	12.00%	0.08%
GDP growth rate 2006/2005	5.597%	0.2%	5.4%	0.0%		
Implicit price index Value Added, total	104.80%	0.1%	104.7%	0.0%	0.10%	0.01%
Propensity to consume	94.08%	0.9%	94.6%	1.5%	6.00%	-0.02%
HH disposable income/GDP ratio	96.20%	2.9%	94.4%	1.1%		

the current year (2006) many of the identities that were not selected for use in the tentative estimates do not hold in the tentative estimates, and can only be satisfied when applying the Bayesian integration approach.

The differences in the BSNA scheme of the tentative and Bayesian estimates from the conventional estimates of the full 2006GV scheme are presented in Table 5. There are slightly more distortions in the Bayesian than in the tentative estimates (13 versus 11), but the distortions concern different aggregates and are also smaller. Thus, the estimate of household disposable income is distorted in the tentative estimates and not anymore in the Bayesian estimates; the difference in the tentative estimates changes from 3.1% to 1.1% in the Bayesian estimates. Also the differences from conventional estimates between the Bayesian estimates of GDP in constant prices reduces considerably between tentative and Bayesian estimates; that of GDP in current prices only improved for the GDP total by expenditures. The same applies to the contribution

Table 6  
Bayesian estimates based on complete versus partial data

	Nr. of distorted values	Bayesian estimates in 2005YF with full vs. partial framework scope				Bayesian estimates in 2006YF with full vs. partial framework scope			
		Bayesian estimates of 2005GV option 1		Bayesian estimates of 2005YF option 12		Bayesian estimates of 2006YF option 14		Bayesian estimates of 2006GV option 13	
		% Difference from conventional estimates	Posterior coefficient of variation	% Difference from conventional estimates	Posterior coefficient of variation	% Difference from conventional estimates	Posterior coefficient of variation	% Difference from conventional estimates	Posterior coefficient of variation
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
GDP ECONOMIC ACTIVITIES, CONSTANT		0.0%	0.07%	0.3%	0.07%	0.0%	0.07%	-0.6%	0.07%
GDP EXPENDITURES, CONSTANT		0.0%	0.07%	0.3%	0.07%	0.0%	0.07%	-0.6%	0.07%
Agriculture, forestry and fishing		0.0%	0.10%	0.1%	0.10%	1.4%	0.10%	-0.1%	0.10%
Mining and quarrying		0.0%	0.15%	0.0%	0.15%	-0.1%	0.12%	0.0%	0.12%
Manufacturing		0.0%	0.25%	1.2%	0.26%	-0.6%	0.26%	-1.0%	0.25%
Electricity, gas and water supply		0.0%	0.35%	0.4%	0.35%	2.4%	0.36%	-0.3%	0.36%
Construction		0.0%	0.53%	1.4%	0.54%	-7.8%	0.57%	-1.2%	0.52%
Wholesale and retail trade		0.0%	0.22%	-0.7%	0.22%	1.9%	0.22%	-0.7%	0.23%
Land transport		0.0%	0.46%	0.3%	0.48%	1.5%	0.47%	-0.4%	0.47%
Transport (other), storage, telecommunication		0.0%	0.28%	0.2%	0.29%	-1.1%	0.29%	-0.4%	0.30%
Financial and insurance activities		0.0%	0.03%	0.3%	0.12%	-0.4%	0.11%	0.0%	0.03%
Private services		0.0%	0.01%	0.3%	0.02%	-0.1%	0.02%	-0.2%	0.01%
Public administration and defence; compulsory social security		0.0%	0.02%	-0.5%	0.04%	-0.4%	0.04%	0.0%	0.02%
VALUE ADDED GROSS, TOTAL, CURRENT		0.0%	0.00%	0.3%	0.00%	-0.1%	0.00%	-0.5%	0.00%
GDP ECONOMIC ACTIVITIES, CURRENT		0.0%	0.01%	0.3%	0.01%	-0.1%	0.01%	-0.1%	0.01%
Household final consumption		0.0%	0.00%	-2.0%	0.00%	2.7%	0.00%	0.7%	0.00%
Government final consumption		0.0%	0.01%	-0.4%	0.01%	-0.2%	0.01%	0.0%	0.01%
NPISH final consumption		0.0%	0.42%	1.2%	0.45%	0.5%	0.45%	-0.2%	0.42%
Gross fixed capital formation		0.0%	0.20%	-8.5%	0.22%	-5.4%	0.20%	-4.0%	0.20%
Exports		0.0%	0.02%	5.4%	0.08%	-0.7%	0.09%	-0.1%	0.01%
Minus: Imports		0.0%	0.02%	-9.0%	0.05%	1.1%	0.05%	-0.2%	0.00%
GDP EXPENDITURES, CURRENT		0.0%	0.01%	0.3%	0.01%	-0.1%	0.01%	-0.4%	0.01%
Compensation of employees		0.0%	0.13%	0.4%	0.14%	0.7%	0.14%	-0.5%	0.14%
Other taxes less subsidies on production		0.0%	0.07%	0.0%	0.10%	0.6%	0.10%	0.0%	0.07%
Non-HH operating surplus, gross		0.0%	0.17%	-1.0%	0.18%	-0.6%	0.18%	-0.7%	0.18%
HH Operating surplus (on dwelling services), gross		0.0%	0.14%	0.1%	0.14%	0.2%	0.14%	-0.1%	0.14%
Mixed income, gross		0.0%	0.21%	2.2%	0.21%	-0.8%	0.22%	-0.3%	0.21%
VALUE ADDED GROSS, TOTAL, CURRENT		0.0%	0.00%	0.3%	0.00%	-0.1%	0.00%	-0.5%	0.00%
HOUSEHOLD DISPOSABLE INCOME, GROSS	2006 Prior values	0.0%	0.08%	-2.7%	0.08%	1.1%	0.08%	-0.1%	0.08%
GDP growth rate 2006/2005	5.4%					0.0%		-0.6%	
Implicit price index Value Added, total	104.7%	0.0%	0.01%	0.0%	0.01%	0.0%	0.01%	0.0%	0.00%
Propensity to consume	93.2%	0.0%	-0.01%	0.7%	-0.02%	1.5%	-0.02%	0.7%	-0.02%
HH disposable income/GDP ratio	93.3%	0.0%		-2.8%		1.1%		0.3%	

to GDP in current prices of land transport and other taxes less subsidies on production; in both cases the differences from the conventional estimates are considerably reduced. Also the distortion in the estimate of the ratio of household disposable income to GDP reduces to a non-distorted value. Furthermore, the differences from conventional estimates reduce for the GDP growth rate and implicit price deflator of GDP. The deviation for the propensity to consume, however, increases. New distortions are also found in the Bayesian estimates of household final consumption and gross fixed capital formation. It is expected that if tentative estimates are improved by the national accountant, the differences from conventional estimates will become smaller not only for the tentative estimates but also for the Bayesian estimates.

#### 6.4. Complete versus partial data in base year

The most important question is how well the Bayesian estimation method generates estimates that are close to conventional estimates, when annually only a partial data

set is available. This question is examined on the basis of alternative sets of estimates presented in Table 6.

The table compares Bayesian estimates for 2005 and 2006, alternatively based on full options 1 and 13, and limited annual data sets for those years (options 12 and 14). The limited data set corresponds to what is normally available annually. From this table several conclusions can be drawn.

When treating a large number of cells in the 2005 framework as variables for which annually no data are available (2005YF, option 12), Bayesian estimates of some GDP expenditures (gross fixed capital formation and also exports and imports) and some components of household disposable income (not shown in the table) differ significantly ( $> 2\%$ ) from the conventional estimates. Bayesian estimates of GDP totals, however, differ only slightly more for the 2005YF framework than for the 2005GV framework, but the total of household disposable income deviates significantly from its conventional estimate in the 2005YF framework. Posterior coefficients of variation do not differ significantly between the 2005GV and 2005YF options, and in both cases are much smaller than their prior equivalents.

A similar pattern is observed when comparing 2006YF and 2006GV estimates (options 14 and 13), but some differences should be noted. Bayesian estimates of GDP are slightly better for 2006YF than for the 2006GV version. The opposite is true for household disposable income, but the deviation from the conventional estimates is significantly less in the 2006YF framework than in the 2005YF framework. There are also differences between the 2005YF and 2006YF frameworks with regard to the components of GDP and household disposable income.

When comparing the estimates of the 2005YF and 2006YF (options 12 and 14), we see that in the 2005YF framework there are no significant deviations in the industry breakdown of GDP, while in the 2006YF framework contributions to GDP of construction, wholesale and retail trade, and land transport show significant deviations from the conventional 2006 estimates. For the expenditures the distortions in the 2005YF framework are in gross fixed capital formation, exports, and imports; in the 2006YF framework they are in household and non-profit institutions final consumption and in gross fixed capital formation. With regard to the IEA variables explaining household disposable income, the deviations from conventional estimates are almost the same in the 2005YF and 2006YF frameworks. There are no significant differences in the coefficients of variation between the 2005YF and 2006YF frameworks. Thus, aggregates of GDP in current and constant prices, when estimated with limited annual data, show insignificant ( $< 2\%$ ) differences from the 2005 and 2006 conventional estimates, while the aggregate of household disposable income differs more from the conventional estimates in the 2005YF and 2006YF options, and is therefore more dependent than GDP on the availability of basic data. Components of GDP by activities and expenditures, and also components of household disposable income, are more dependent on the availability of basic data than the totals.

It should be noted that there is hardly any effect of limited data availability on the measurement of GDP growth in 2006, the implicit price deflator of total value added,

Table 7  
Dependence of Bayesian estimates on framework scope and data availability

	Bayesian estimates in 2005GV with full vs. partial framework scope										Bayesian estimates of alternatives, based on different data availability and precision, measured in 2006YF									
	Bayesian estimates of option 1		SUT only option 9		IEA only option 10		High quality of data on -Household final consumption -Imports and exports -Financial corporations -Services option 20		Poor quality of household final consumption data and ratios option 16		Poor quality of export and import data option 17		Poor quality of data on services option 18		Poor quality of financial corporations data option 19		Aggregate data option 21			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)		
	No. of distorted values																			
	0	1	0	0	0	0	2	15	11	2	13	2	15	0	2	25	4	34		
GDP DOMESTIC ACTIVITIES, CURRENT	3.00%	0.0%	0.01%	0.0%	0.10%				-1.1%	0.03%	-0.2%	0.07%	-0.2%	0.08%	2.2%	0.00%	-1.1%	0.00%		
GDP EXPENDITURES, CONSTANT	3.00%	0.0%	0.03%	0.0%	0.10%				-1.1%	0.03%	-0.2%	0.07%	-0.2%	0.08%	2.2%	0.00%	-1.1%	0.00%		
Agriculture, forestry and fishing	12.00%	0.0%	0.10%	0.0%	0.41%				1.1%	0.10%	1.3%	0.10%	1.4%	0.10%	0.3%	0.10%	1.2%	0.10%		
Mining and quarrying	12.00%	0.0%	0.10%	0.0%	0.44%				-0.1%	0.12%	-0.1%	0.12%	-0.1%	0.12%	-0.3%	0.12%	-0.2%	0.12%		
Manufacturing	12.00%	0.0%	0.25%	0.0%	0.94%				-3.3%	0.23%	-1.0%	0.27%	-1.0%	0.27%	10.9%	0.26%	-3.3%	0.24%		
Electricity, gas and water supply	12.00%	0.0%	0.35%	0.0%	0.67%				1.6%	0.36%	2.3%	0.36%	2.3%	0.36%	-0.2%	0.36%	1.5%	0.36%		
Construction	12.00%	0.0%	0.55%	0.0%	0.70%				-0.3%	0.57%	-0.1%	0.56%	-0.2%	0.57%	25.1%	0.70%	-0.4%	0.56%		
Wholesale and retail trade	12.00%	0.0%	0.22%	0.0%	0.41%				0.5%	0.21%	1.7%	0.22%	0.9%	0.22%	67.5%	0.26%	-2.5%	0.22%		
Land transport	12.00%	0.0%	0.46%	0.0%	0.67%				0.6%	0.48%	1.1%	0.48%	1.2%	0.48%	4.0%	0.71%	0.7%	0.48%		
Transport (air, rail, storage, telecommunication)	12.00%	0.0%	0.28%	0.0%	0.66%				-2.0%	0.29%	-1.2%	0.29%	-1.1%	0.29%	30.9%	0.66%	1.7%	0.29%		
Financial and insurance activities	1.00%	0.0%	0.03%	0.0%	0.15%				-0.5%	0.11%	-0.4%	0.12%	-0.5%	0.12%	-2.7%	0.24%	-13.9%	0.75%		
Private services	12.00%	0.0%	0.01%	0.0%	0.36%				-0.6%	0.06%	-0.2%	0.09%	-0.2%	0.09%	-11.1%	0.15%	-0.5%	0.05%		
Public administration and defence; compulsory social security	0.10%	0.0%	0.00%	0.0%	0.06%				-1.4%	0.04%	-0.4%	0.04%	-1.4%	0.04%	-1.5%	0.04%	-1.5%	0.04%		
VALUE ADDED GROSS, FINAL, CURRENT	12.00%	0.0%	0.01%	0.0%	0.00%				-1.2%	0.01%	-1.3%	0.01%	-1.3%	0.01%	2.2%	0.00%	-1.0%	0.00%		
GDP DOMESTIC ACTIVITIES, CURRENT	3.00%	0.0%	0.01%	0.0%	0.01%				-1.2%	0.01%	-1.3%	0.01%	-1.3%	0.01%	2.2%	0.00%	-1.0%	0.01%		
Household final consumption	3.00%	0.0%	0.00%	0.0%	0.13%	0.1%	1.65%		1.0%	0.05%	2.6%	0.05%	0.1%	0.09%	3.1%	0.05%	0.0%	0.05%		
Government final consumption	0.10%	0.0%	0.01%	0.0%	0.04%	-2.4%	0.04%		1.2%	0.01%	-0.2%	0.03%	1.2%	0.01%	1.1%	0.01%	0.0%	0.01%		
NRISH final consumption	6.00%	0.0%	0.42%	0.0%	1.50%	-0.1%	1.79%		0.4%	0.44%	0.7%	0.47%	-0.4%	0.48%	1.3%	0.68%	-1.2%	0.45%		
Gross fixed capital formation	3.00%	0.0%	0.20%	0.0%	0.03%	0.1%	1.61%		-4.4%	0.19%	-5.6%	0.20%	-6.0%	0.21%	5.8%	0.19%	-5.3%	0.19%		
Exports	0.10%	0.0%	0.02%	0.0%	0.08%	0.0%	0.09%		-0.8%	0.09%	-0.7%	0.09%	-0.8%	0.09%	1.1%	0.09%	-0.5%	0.09%		
Minus: Imports	0.10%	0.0%	0.02%	0.0%	0.08%	0.0%	0.06%		1.1%	0.06%	1.1%	0.07%	-8.3%	1.31%	2.0%	0.06%	0.6%	0.06%		
GDP EXPENDITURES, CURRENT	3.00%	0.0%	0.01%	0.0%	0.01%				-1.2%	0.01%	-1.3%	0.01%	-1.3%	0.01%	2.2%	0.00%	-1.0%	0.01%		
Compensation of employees	12.00%	0.0%	0.10%	0.0%	0.21%	0.2%	3.44%		-0.5%	0.13%	0.5%	0.14%	0.5%	0.14%	0.8%	0.13%	-1.0%	0.14%		
Other taxes less subsidies on production	0.10%	0.0%	0.07%	0.0%	0.10%	-0.2%	0.02%		0.5%	0.10%	0.6%	0.10%	0.6%	0.10%	1.5%	0.09%	0.3%	0.10%		
Non-ferrous operating surplus, gross	12.00%	0.0%	0.17%	0.0%	0.17%	-0.2%	0.02%		-2.1%	0.17%	-0.8%	0.18%	-0.9%	0.18%	0.8%	0.17%	-3.2%	0.18%		
MI Operating surplus (on dwelling services), gross	12.00%	0.0%	0.14%	0.0%	0.14%	0.0%	11.25%		0.1%	0.14%	0.2%	0.16%	0.1%	0.16%	-0.1%	0.14%	0.1%	0.14%		
Mixed income, gross	12.00%	0.0%	0.21%	0.0%	0.32%	0.1%	6.42%		-2.0%	0.21%	-0.9%	0.22%	-1.2%	0.22%	8.1%	0.20%	-2.7%	0.21%		
VALUE ADDED GROSS, FINAL, CURRENT	12.00%	0.0%	0.00%	0.0%	0.00%	0.0%	1.29%		-1.7%	0.00%	-1.7%	0.00%	-1.7%	0.00%	2.2%	0.00%	-2.0%	0.00%		
HOUSEHOLD DISPOSABLE INCOME, GROSS	12.00%	0.0%	0.00%	0.0%	0.00%	0.1%	1.01%		-0.1%	0.00%	1.0%	0.00%	-1.2%	0.00%	2.0%	0.00%	-0.7%	0.00%		
GDP growth rate 2006:2005	0.10%	0.0%	0.01%	0.0%	0.05%				-1.2%	0.01%	-0.2%	0.01%	-0.2%	0.01%	2.0%	0.00%	-1.0%	0.01%		
Implicit price index Value Added, total	6.00%	0.0%	-0.01%	0.0%	0.05%				0.3%	0.03%	0.0%	0.03%	0.1%	0.04%	-0.3%	0.03%	0.1%	0.04%		
Propensity to consume		0.0%				0.0%	0.09%		1.0%		1.2%		0.8%		0.0%		0.1%	0.01%		
MI disposable income/GDP ratio		0.0%							1.0%		1.2%		0.8%		0.0%		0.5%	0.0%		

the propensity to consume of households, and the terms of trade effect. As the effect on household disposable income in the 2006YF framework is much smaller than in the 2005YF framework, the impact on the GDP/household disposable income ratio is also much lower in the 2006YF framework than in the 2005YF framework.

## 7. Sensitivity analysis

### 7.1. Sensitivity to SUT or IEA scope

Many countries limit their SNA compilation to the SUT. At the other extreme, we may consider compilation of only the IEA, which may be relevant if analysts are interested in sector accounts and their monetary extension. The question raised here is how reliable estimates are if they are based solely on the compilation of the SUT or alternatively of the IEA.

The Bayesian estimates in Table 7 answer this question for the 2005GV framework of the benchmark year, by comparing options 9 and 10 to option 1. The results show that if only the SUT is compiled (option 9), the number of distortions in the estimates remains 0, as in option 1, while the number of distortions of the coefficients of

variation increases slightly (from 3 to 6). Hence the precision of the compilation of the SUT is nearly the same as when the whole framework is compiled. This is not the case, however, when only the IEA is compiled. In that case, unexpectedly perhaps, Bayesian estimates of government final consumption deviate considerably from the conventional estimates, and the same holds (not shown in the table) for household social transfers, received (+) less paid (-). The number of distortions in the posterior coefficient of variation increases dramatically from 3 to 15, if only the IEA is compiled. The latter is to be expected, because much fewer basic data and also identity and ratio restrictions are used in this compilation.

### *7.2. Sensitivity to data available from surveys and administrative data sources*

The next question is whether some basic data influence the precision of the estimates more than others, i.e. which surveys and administrative data should be considered essential for compiling reliable national accounts aggregates. In Table 7 we quantify the effect for the 2006YF framework: in option 16 for data availability on household final consumption data (columns 10 and 11), in option 17 for data on exports and imports (columns 12 and 13), in option 18 for data on services (columns 14 and 15), and in option 19 for data on financial corporations (columns 16 and 17). In all cases, impacts are assessed by comparing the distortions in the Bayesian estimates and the posterior coefficients of variation with those for option 20 in columns 8 and 9. Option 20 is close to option 14, but includes all final consumption restrictions.

The impact of household final consumption restrictions is measured by comparing option 16 (without restrictions) to option 20 (with restrictions). The absence of the restrictions in option 16 consists in not using the household survey data, the assignment of a poor (P) reliability to the weights of household final consumption items in the benchmark household survey, and also the assignment of a poor (P) reliability to the use coefficients (in order to eliminate the impact of structural coefficients that may mitigate the influence of not having household survey data and ratios). The overall impact is limited: Without using household survey data, the number of distortions in the values of the Bayesian estimates increases from 11 to 13, while the number of distortions in the coefficients of variation remains the same (2). The additional distortions are mainly in the two items of household final consumption in current and constant prices, as expected. There are some shifts in distortions for individual activity categories of GDP in current and constant prices, but overall there are no serious distortions in the main aggregates of GDP in current and constant prices and household disposable income, and in the details thereof.

Not having import and export details would be the case in a regional economy in which it is difficult to register incoming and outgoing flows of goods, in countries where foreign trade statistics are little developed, or in countries belonging to a customs union such as the European Union. This situation is simulated in option 17, which assigns poor (P) reliability to import and export data, balance of payments, and external sector data, and also, as before, to the user coefficients. To measure the



impacts, a comparison is made between the distortions in options 17 and 20. The impacts in this case are also limited, and are mainly found in the four items of exports and imports in current and constant prices, leading to 15 distortions in the values of the estimates in option 17 against 11 in option 20. The main aggregates of GDP in current and constant prices and household disposable income are not in any major way affected by the lack of export and import and external sector data. The number of distorted values of coefficients of variation in this case is higher than in option 20: There are 9 instead of 2 distorted values of those coefficients, and the additional ones are in exports and imports in current and constant prices, and in several components of household disposable income (not shown in the table).

Impacts are much larger in a third scenario, in which it is assumed that there are no reliable statistics on services. This is simulated in option 18, in which data on services are assigned poor (P) reliability. When comparing the distortions in option 18 with those in option 2, we observe that the number of distortions in the values of the Bayesian estimates increases from 11 to 32, while the number of distortions in the coefficients of variation remains the same (2). The distortions in the values of the Bayesian estimates occur in all three major aggregates of GDP in current and constant prices and also in household disposable income. Several subcomponents of the three aggregates are affected. Posterior coefficients of variation are not affected.

The impact of not having reliable data on financial corporations was simulated in option 19. In this option it was assumed that all data on financial corporations in the SUT and IEA had poor (P) reliability. When comparing the distortions in this option 19 with those in option 20, we found that the number of distortions in the Bayesian estimates increased significantly from 11 to 23, while the number of distortions in the posterior coefficients of variation increased slightly from 2 to 4. In this case Bayesian estimates of the main aggregates of GDP in current and constant prices and household disposable income are not significantly distorted. The distortions only occur in the subcomponents of GDP by activities that are related to services, and also in several subcomponents of household disposable income (not shown in the table). The additional distortions in the coefficients of variation also occur in these subcomponents.

### 7.3. Sensitivity to aggregation

The present framework with 2719 variables is a very large one. The number of aggregate variables is, however, relatively small (62), and a subset of these are presented in the tables. This follows national accounts practices, which is generally also carried out in much detail leading to a small set of estimates, sometimes just one: GDP. This procedure is based on the assumption that it is more precise to estimate aggregates by using much detail in the compilation. Added complexity does not *necessarily* improve the estimates, and this is tested in option 21 in the last column of Table 7.

To carry out this test a small framework was designed, containing the main aggregates of GDP in current and constant prices and household disposable income, but also including subcomponents that are needed to link these variables conceptually and quantitatively. The main aggregates are output, intermediate consumption, value added, value added components classified by aggregate ISIC categories, and total imports and exports, as well as sector components that link GDP to household disposable income. The total number of variables in the aggregate framework is 105, of which 29 (27.6% versus 19.5% in the extended framework) are supported by basic data, and there are 29 identities and 83 ratios. The number of information items (basic data, identities, and ratios) per variable is therefore 1.34 (versus 1.45 in the extended BSNA framework). Reliabilities are attached to basic data and ratio values. Although similar in nature, these reliabilities are not quantitatively linked to those of the extended framework presented in Tables 1 and 2.

Column 18 in Table 7 shows that the number of distortions in the estimated values increases to 34, when an aggregate 2006YF framework is used, as compared to 11 distortions in option 21 for the extended 2006YF framework. Distortions in the coefficients of variation are not presented for the aggregate framework, as they are not comparable with those of the extended framework. We conclude that Bayesian estimates compiled in a disaggregated framework are generally closer to the conventional estimates than those compiled through aggregates.

There is another aspect of aggregation, the results of which are reflected in Tables 4–8. In several options in those tables, it is shown that the values and coefficients of variation of GDP and household disposable income are generally not distorted, while distortions are observed for components of GDP and household disposable income. For example, in option 14 of the 2006YF framework in Table 6, there are no significant distortions in the values of GDP in current and constant prices and household disposable income, and also the growth rate is not much affected. However, for some details of these aggregates by industries and other categories, there are significant distortions, such as for contributions to GDP of electricity, gas and water and construction, household final consumption and gross fixed capital formation.

#### *7.4. Sensitivity to prior precisions*

Prior reliabilities play an important role in the Bayesian estimation procedure. The prior coefficients of variation are subjectively determined by national accountants based on insight and some detailed studies; see also [2]. We now ask how much influence these priors have on the Bayesian estimates and how much posterior coefficients of variation are reduced when compared to their prior values. We will also consider the question to what extent values of Bayesian estimates of variables with basic data are located outside the range of the basic values plus/minus a percent variation that is permitted by the national accountant, when quantifying the prior coefficient of variation.

Table 8  
Impact of prior reliability on Bayesian estimates and posterior reliability

	Prior coefficients of variation	Bayesian estimates in 2005GV								Bayesian estimates in 2006YF					
		Standard prior coefficients of variation option 1	50% Increase of prior coefficients of variation option 3	50% Decrease of prior coefficients of variation option 2	Fictitious prior coefficients of variation (100%) option 4		Standard prior coefficients of variation option 14	50% Increase of prior coefficients of variation option 6	Prior reliability of price indices poor (P) option 5						
			Posterior coefficient of variation	% Difference from conventional estimates	Posterior coefficient of variation	% Difference from conventional estimates	Posterior coefficient of variation	% Difference from conventional estimates	Posterior coefficient of variation	% Difference from conventional estimates	Posterior coefficient of variation	% Difference from conventional estimates	Posterior coefficient of variation	% Difference from conventional estimates	
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Nr. of distorted values		0	3	0	11	0	4	0	61	13	3	13	5	10	11
<p>                     GDP ECONOMIC ACTIVITIES, CONSTANT                      3.00% 0.0% 0.07% 0.0% 0.16% 0.0% 0.05% 0.0% 8.18% 0.0% 0.07% -0.4% 0.13% 0.8% 0.47%                      GDP EXPENDITURES, CONSTANT                      3.00% 0.0% 0.07% 0.0% 0.16% 0.0% 0.05% 0.0% 8.18% 0.0% 0.07% -0.4% 0.13% 0.8% 0.47%                      Agriculture, forestry and fishing                      12.00% 0.0% 0.10% 0.0% 0.42% 0.0% 0.14% 0.0% 10.80% 1.4% 0.10% 1.3% 0.19% 1.9% 0.12%                      Mining and quarrying                      12.00% 0.0% 0.15% 0.0% 0.46% 0.0% 0.15% 0.0% 5.55% -0.1% 0.12% -1.1% 0.23% -1.7% 0.17%                      Manufacturing                      12.00% 0.0% 0.25% 0.0% 1.35% 0.0% 0.37% 0.0% 21.63% -0.6% 0.26% -1.0% 0.48% -1.3% 0.30%                      Electricity, gas and water supply                      12.00% 0.0% 0.35% 0.0% 0.86% 0.0% 0.23% 0.1% 22.98% 2.4% 0.36% 2.7% 0.69% 2.6% 0.45%                      Construction                      12.00% 0.0% 0.53% 0.0% 1.08% 0.0% 0.28% 0.0% 30.42% -2.8% 0.57% -7.1% 1.10% -9.7% 0.63%                      Wholesale and retail trade                      12.00% 0.0% 0.22% 0.0% 0.46% 0.0% 0.13% 0.0% 17.24% 1.9% 0.22% 0.1% 0.43% -1.3% 0.27%                      Land transport                      12.00% 0.0% 0.46% 0.0% 0.96% 0.0% 0.25% 0.0% 28.80% 1.5% 0.47% 2.2% 0.90% -0.8% 0.55%                      Transport (other), storage, telecommunication                      12.00% 0.0% 0.28% 0.0% 0.82% 0.0% 0.23% 0.0% 22.70% -1.1% 0.29% -1.2% 0.55% -2.1% 0.39%                      Financial and insurance activities                      1.00% 0.0% 0.03% 0.0% 0.09% 0.0% 0.06% 0.0% 18.95% -0.4% 0.11% 0.4% 0.18% -1.3% 0.20%                      Private services                      12.00% 0.0% 0.01% 0.0% 0.43% 0.0% 0.12% -0.1% 24.19% -0.1% 0.02% -0.2% 0.17% -0.3% 0.03%                      Public administration and defence; compulsory social security                      0.10% 0.0% 0.02% 0.0% 0.04% 0.0% 0.04% 0.0% 9.35% -0.4% 0.04% -0.5% 0.04% -0.5% 0.04%                      VALUE ADDED GROSS, TOTAL, CURRENT                      12.00% 0.0% 0.00% 0.0% 0.14% 0.0% 0.00% 0.0% 4.02% -0.1% 0.00% -0.7% 0.13% -1.0% 0.00%                      GDP ECONOMIC ACTIVITIES, CURRENT                      3.00% 0.0% 0.01% 0.0% 0.13% 0.0% 0.00% 0.0% 2.97% -0.1% 0.03% -0.4% 0.12% -1.0% 0.01%                      Household final consumption                      0.10% 0.0% 0.00% 0.0% 0.18% 0.0% 0.05% 0.0% 5.84% 2.7% 0.00% 2.2% 0.00% 1.9% 0.00%                      Government final consumption                      1.00% 0.0% 0.01% 0.0% 0.01% 0.0% 0.01% 0.0% 9.20% -0.2% 0.01% -0.3% 0.03% -0.3% 0.01%                      NPISH final consumption                      6.00% 0.0% 0.42% 0.0% 0.81% 0.0% 0.22% 0.0% 6.92% 0.5% 0.45% 0.6% 0.87% 0.4% 0.46%                      Gross fixed capital formation                      3.00% 0.0% 0.20% 0.0% 1.76% 0.0% 0.45% 0.0% 24.87% -5.4% 0.20% -6.2% 0.39% -5.9% 0.20%                      Exports                      0.10% 0.0% 0.02% 0.0% 0.05% 0.0% 0.05% 0.0% 15.25% -0.7% 0.09% -0.4% 0.09% -0.7% 0.09%                      Minus: Imports                      0.10% 0.0% 0.02% 0.0% 0.04% 0.0% 0.03% 0.0% 10.07% 1.1% 0.05% 0.7% 0.06% 1.1% 0.05%                      GDP EXPENDITURES, CURRENT                      3.00% 0.0% 0.01% 0.0% 0.13% 0.0% 0.00% 0.0% 3.97% -0.1% 0.01% -0.4% 0.13% -1.0% 0.01%                      Compensation of employees                      12.00% 0.0% 0.13% 0.0% 0.27% 0.0% 0.07% 0.0% 7.76% 0.7% 0.14% 0.4% 0.27% -0.1% 0.15%                      Other taxes less subsidies on production                      0.10% 0.0% 0.07% 0.0% 0.07% 0.0% 0.06% 0.0% 4.90% 0.6% 0.10% 0.3% 0.10% 0.5% 0.10%                      Non-IH operating surplus, gross                      12.00% 0.0% 0.17% 0.0% 0.38% 0.0% 0.07% 0.0% 5.14% -0.6% 0.18% -1.2% 0.34% -1.7% 0.19%                      IH Operating surplus (on dwelling services), gross                      12.00% 0.0% 0.14% 0.0% 0.29% 0.0% 0.06% 0.1% 28.61% 0.2% 0.14% 0.2% 0.30% 0.2% 0.13%                      Mixed income, gross                      12.00% 0.0% 0.21% 0.0% 0.59% 0.0% 0.15% 0.0% 14.77% -0.8% 0.22% -0.8% 0.43% -1.9% 0.22%                      VALUE ADDED GROSS, TOTAL, CURRENT                      12.00% 0.0% 0.00% 0.0% 0.14% 0.0% 0.00% 0.0% 4.02% -0.1% 0.00% -0.7% 0.13% -1.0% 0.00%                      HOUSEHOLD DISPOSABLE INCOME, GROSS                      12.00% 0.0% 0.08% 0.0% 0.16% 0.0% 0.05% 0.0% 5.59% 1.1% 0.08% 0.7% 0.15% 0.3% 0.09%                      GDP growth rate 2006/2005                      0.10% 0.0% 0.01% 0.0% 0.13% 0.0% 0.03% 0.0% 8.67% 0.0% 0.01% 0.0% 0.01% -1.0% 0.17%                      Implicit price index Value Added, total                      6.00% 0.0% -0.01% 0.0% 0.07% 0.0% 0.02% 0.0% 1.75% 1.5% -0.02% 1.4% -0.03% 1.4% -0.02%                      IH disposable income/GDP ratio                      0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 1.1% 1.1% 1.1% 1.2% 1.2% 1.2%                      P/PW (Terms of trade)                      0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% -0.3% -0.3% -0.3% -0.3% -0.3% -0.3%                 </p>															

In Table 8 a comparison is made between posterior coefficients of variation of the full 2005GV benchmark using standard prior reliabilities and the same framework in which these standard prior reliabilities are changed. The same exercise is carried out for the 2006YF framework with limited data availability. For the 2005GV framework the impact is measured by increasing or decreasing the standard prior reliabilities proportionally with  $\pm 50\%$  and by using fictitious prior reliabilities of  $\pm 100\%$  for all basic data and ratio values. For the 2006YF framework the impact is measured by lowering the S (superior) reliability of price indices to P (poor) and by increasing proportionally all coefficients of variation by 50%.

The impact of the changes in prior reliabilities on the estimates is low. In the 2005 framework (options 2, 3, 4 versus 1) the number of distortions in estimates (i.e.  $> 2\%$  change as compared to conventional estimates) remains zero, even when fictitious prior reliabilities are used. A similar pattern is observed for the 2006YF framework (options 5 and 6 versus 14). In that framework the number of distortions in estimates of variables remains the same (13), and when the reliability of price indices is lowered from S to P, the number of distortions in estimates is even reduced to 10.

The change in the prior reliabilities does, however, have a significant impact on the posterior coefficient of variation of the estimates. In the 2005GV framework, this effect is largest in the case of using fictitious prior coefficients of variation (64 instead of 3 distortions), a little smaller when proportionally increasing the coefficients of variation (11 distortions), and hardly any impact when proportionally decreasing the prior coefficients of variation (4 distortions).

In the case of the 2006YF framework, the impact on posterior coefficients of variation is insignificant, when prior coefficients of variation are increased proportionally with 50%. Lowering the reliability of price indices in the 2006YF framework has no (or even a decreasing) effect on the number of distortions in the Bayesian estimates, but the number of distortions in the posterior coefficients of variation has increased (from 3 to 11).

In Table 8, as in earlier tables, we note the significant change in the posterior reliability as compared to the prior reliability of variables. In most instances, prior reliabilities of 3, 5, and 12% change to much lower posterior reliabilities of less than 1%, independent of their prior value. In the 2006YF framework of option 14 the variation coefficient of GDP in current prices changes from 3% prior to 0.01% posterior, and in constant prices from 3% to 0.07%. For household disposable income the reduction is from 12% to 0.08%. Posterior coefficients of variation are generally higher for details than for the totals of GDP and household disposable income, and also higher when using only the IEA in the estimation of the two aggregates (option 10 in Table 7). We emphasize that, while posterior variances are always smaller than prior variances (in accordance with Bayesian theory), this is not necessarily the case for coefficients of variation. In fact, we found that in 38% of the cases (1173 of the 3053 variables estimated), the deviation of posterior estimates from conventional estimates deviated more than the percent value of the prior coefficient of variation. Overall, the findings confirm the utility of using a Bayesian approach to national accounts integration; the approach makes the posterior estimates much more reliable.

## **8. Conclusions**

Based on experiences and sensitivity experiments with the Bayesian estimation approach as described in this paper we draw eight conclusions.

After many INTERFACE/SNAER runs applied to various versions of the 2005 and 2006 framework of Guatemala and other countries in Central America, the BSNA framework and the Bayesian estimation approach can be considered as operational, not only in Guatemala but, with minor modifications, in all countries using the SNA. In other words, the software programmes can be applied to any framework of data and variables, for which identity and ratio relations and reliabilities can be defined and presented in Excel format, leading to a consistent set of Bayesian estimates and posterior coefficients of variation. The selection of basic data, identities, ratios, and reliabilities are essential ingredients in the Bayesian approach presented here. As

long as these four elements can be well identified, the Bayesian estimation method does not compete but can be well combined with alternative methodologies such as ERETES developed by EUROSTAT, and the method proposed in [17] to construct input-output tables or the entropy method used in [16] to reconcile household survey data and national accounts.

It proved feasible to define an internally consistent and very large BSNA framework of 2719 variables, including not only the SUT and the major aggregate of GDP, but also the IEA which incorporates household disposable income as another major aggregate. The internal consistency of the framework means that there is full compatibility of Bayesian inputs, and this is confirmed by applying INTERFACE and SNAER to the 2005GV framework, resulting in Bayesian estimates that are close to the conventional estimates and without any distortions in the posterior coefficients of variation. The internal consistency of the framework is also reflected in estimates of price indices that remain close to 1.00 in the base year.

The use of assumptions (i.e. a selection of ratios and identities) to arrive at tentative estimates for all variables of the framework, is found to be the best procedure to prepare the framework for Bayesian integration and arrive at current year estimates (2006YF framework). It leads to Bayesian estimates for a current year that are close to conventional estimates for that year, in particular for major aggregates such as GDP and household disposable income. More efforts are needed by national accountants to improve the tentative estimates prior to Bayesian integration.

A set of basic data for a current year (2006YF) on output, imports and exports, employment, data on the government and financial corporate sectors, and the rest of the world, constituting approximately 20% of the total number of 2719 variables in the system, is adequate for making Bayesian estimates of the comprehensive framework of SUT and IEA, as long as the basic data are supplemented by a large number of identities and ratio values including a full set of price indices. This would apply not only to the compilation of annual accounts, but may also hold for early estimates, quarterly accounts, or even projections. When the number of basic data gradually increases between early (flash), preliminary, semi-final, and final estimates, while the number of identities and ratios remains the same, Bayesian estimates gradually improve, i.e. discrepancies between Bayesian and conventional estimates reduce and the number of estimates with distorted posterior coefficients of variation also reduces. In the present framework for annual accounts the ratio of the number of information items (basic data, ratios and identities) and the number of variables is 1.45, while the maximum ratio (in this framework) is 2.02 when basic data are available on all variables.

Compilation through only the SUT yields Bayesian estimates that are close to those of the full framework. This means that the IEA data and their restrictions do not add much to the information that is needed to arrive at GDP and other main aggregates. Compilation through the IEA only is much less reliable.

The best procedure to compile major aggregates is through the compilation of details. Compilation on the basis of aggregate basic data is found to be much less

reliable. Within this preferred procedure major aggregates are compiled much more reliably than details by economic activities and sectors and by expenditure and income components of GDP and household disposable income.

By analyzing the sensitivity of Bayesian estimates to the availability of basic data from selected survey and administrative data sources, we find that direct data on the financial corporate sector and on services have most impact. Direct information from household surveys on household final consumption and from foreign trade and balance of payments statistics on exports, imports, and external sector transactions has less impact.

The posterior coefficients of variation are much smaller than the prior values. This means that integration of estimates, as being pursued by the Bayesian integration method in line with similar practices by national accountants, results in a considerable improvement of the reliability of Bayesian (and also conventional) estimates. The reduction in posterior coefficients of variation is the most in household disposable income and GDP in current prices, less in GDP in constant prices, and even less in details of both aggregates.

In addition to these eight conclusions, we mention a few thoughts for future work. Regarding the use of basic data, more work is required. We need data that are 'basic', hence not already processed. In practice, this is not fully possible, because all data are processed. But there are degrees of processing, and the more basic the better in our method; see also the distinction made in [24] between the use of national accounts frameworks for observation and analysis.

Regarding the Bayesian method, more calibration of Bayesian estimates with conventional estimates is needed in order to specify the ideal features (ratios, identities, and reliabilities) of a framework of this size, in which Bayesian estimates are close to conventional estimates, not only for aggregates but also for details. More experiments are also needed in order to arrive at estimates for current years (later than 2006) that are further away from the base year (2005). The normal distribution has its limitations, especially the symmetry implied by this distribution. The possibility of introducing inequalities in the estimation should be investigated; see [1]. Finally, the possibility of using parameters of multivariate regression functions instead of the present binary ratios would be of interest.

Regarding the framework, alternatives may be considered. The present use is in annual accounts, but work is already underway to apply our methods to quarterly accounts. The use of frameworks for alternative monetary analysis has been applied to regional accounts, their use in monetary analysis in which the IEA is extended to financial accounts is being investigated, and also their use in projections. Furthermore, instead of only using monetary variables as in the present BSNA framework, non-monetary variables may be introduced in satellite frameworks. Efforts are underway in defining and implementing a Bayesian integration of health-environmental satellite frameworks with (non)monetary variables, and frameworks for demographic analysis, as in [6]. In general, any study based on large data sets may be supported by specialized data frameworks, in which the data of the study are made compatible

with each other through Bayesian integration, and missing variables are estimated as well.

## Acknowledgements

We are grateful to Magda Ascues, Carlos Blanco, Victor Flores, Arthur Giesberts, and Ruud Picavet for constructive comments and help in the preparation of documents, and to the editor and referees of JESM for their positive response. Responsibility for the content of this paper is with the authors and does not reflect in any way the point of view of the UN Statistics Division.

## References

- [1] H.J. Boonstra, C. de Blois and G.J. Linders, Macro-integration with inequality constraints: An application to the integration of transport and trade statistics, Discussion Paper 10006, Statistics Netherlands, The Hague/Heerlen, 2010.
- [2] F. Bos, The art and craft of compiling national accounts statistics and their implications for reliability, *The Review of Income and Wealth* **55** (2009), 930–958.
- [3] D. Danilov and J.R. Magnus, Some equivalences in linear estimation (in Russian), *Quantile* **3** (2007), 83–90. (English original can be downloaded from <http://www.janmagnus.nl/papers/JRM077E.pdf>.)
- [4] D. Danilov and J.R. Magnus, On the estimation of a large sparse Bayesian system: The Snaer program, *Computational Statistics & Data Analysis* **52** (2008), 4203–4224.
- [5] N.I.M. Gould and J. Nocedal, The modified absolute-value factorization for trust-region minimization, in: *High Performance Algorithms and Software in Nonlinear Optimization*, R. De Leone, A. Murli, P.M. Pardalos and G. Toraldo, eds, Kluwer Academic Publishers, Dordrecht, 1998, pp. 225–241.
- [6] K. Gross, B.A. Craig and W.D. Hutchison, Bayesian estimation of a demographic matrix model from stage frequency data, *Ecology* **83** (2009), 3285–3298.
- [7] D.W. Jorgenson, A new architecture for the U.S. national accounts, *The Review of Income and Wealth* **55** (2009), 1–42.
- [8] D.W. Jorgenson, S. Landefeld and W. Nordhaus, A new architecture for the U.S. national accounts: A reply to André Vanoli, *The Review of Income and Wealth* **56** (2010), 752–761.
- [9] J.R. Magnus and J.W. Van Tongeren, Assessing and optimizing the impact of national accounts compilation methods on indicator analysis, in: *United Nations Statistics Division, Handbook of National Accounting: Use of Macro Accounts in Policy Analysis*, Studies in Methods Series F, No. 81, Chapter VI-D, United Nations, New York, 2002, pp. 263–281.
- [10] J.R. Magnus, J.W. Van Tongeren and A.F. De Vos, National accounts estimation using indicator ratios, *The Review of Income and Wealth* **46** (2000), 329–350.
- [11] B.D. McCullough and B. Wilson, On the accuracy of statistical procedures in Microsoft Excel 97, *Computational Statistics & Data Analysis* **31** (1999), 27–37.
- [12] B.D. McCullough and B. Wilson, On the accuracy of statistical procedures in Microsoft Excel 2000 and Excel XP, *Computational Statistics & Data Analysis* **40** (2002), 713–721.
- [13] C.R. Rao, Unified theory of linear estimation, *Sankhyā A* **33** (1971), 371–477. (Corrigenda, *Sankhyā A* **34** (1972), 477.)
- [14] C.R. Rao, *Linear Statistical Inference and Its Applications*, (2nd edition), John Wiley, New York, 1973.
- [15] M.B. Reinsdorf, Terms of trade effects: Theory and measurement, *The Review of Income and Wealth* **56** (2010), S177–S205.

- [16] A.-S. Robilliard and S. Robinson, Reconciling household surveys and national accounts data using a cross entropy estimation method, *The Review of Income and Wealth* **49** (2003), 395–406.
- [17] J.M. Rueda-Cantuche and T. ten Raa, Testing the assumptions made in the construction of input-output tables, International Input-Output Association, Working Paper WPIOX 09-011, 2009.
- [18] J.E. Stiglitz, A. Sen and J.-P. Fitoussi, Report by the Commission on the Measurement of Economic Performance and Social Progress, [www.stiglitz-sen-fitoussi.fr](http://www.stiglitz-sen-fitoussi.fr), 2009.
- [19] United Nations, European Commission, OECD, IMF and World Bank, *System of National Accounts 1993*, Sales No. E.94.XVII.4, United Nations, New York, 1993.
- [20] United Nations, European Commission, OECD, IMF and World Bank, *System of National Accounts 2008*, Sales No. 08.XVII.29, United Nations, New York, 2008.
- [21] United Nations Statistics Division (UNSD), *Classification by Broad Economic Categories, Defined in terms of SITC, Rev. 3 (BEC Rev. 3)*, Series M, No. 53, United Nations, New York, 1989.
- [22] J.W. Van Tongeren, Designing analytical data frameworks, *The Review of Income and Wealth* **50** (2004), 279–297.
- [23] J.W. Van Tongeren, *Bayesian Integration of Large SNA Data Frameworks*, Final report of BSNA Application in CMCA Region, Delft, The Netherlands, 2010.
- [24] A. Vanoli, The new architecture of the U.S. national accounts and its relationship to the SNA, *The Review of Income and Wealth* **56** (2010), 734–751.