# DEVELOPMENT OF AN ALGORITHM FOR THE COMPILATION OF NATIONAL ACCOUNTS AND RELATED SYSTEMS OF STATISTICS*

BY JAN W. VAN TONGEREN

*United Nations Statistical Office*

The present article includes a proposal for a national accounts algorithm to be applied in the computerization of the national accounts compilation process. While aiming at the estimation and reconciliation of data from different statistical sources, it is based on the application of a linear programming technique applied to a system of identities and inequalities that define the accounting and analytical relations between the data categories of a national accounting framework. The technique is flexible in the sense that it can be used with any configuration of available statistical sources and data requirements of the national accounts. The algorithm is illustrated graphically with help of a simple example and thereafter applied to an extended but still simple national accounting scheme for Suriname with data for 1965 that was compiled by the author many years ago. As the present study is only a first step in the development of the algorithm, more work is needed to make it operational and the last section of the article includes suggestions about the direction of that further work.

## INTRODUCTION

National accounts compilation procedures are increasingly being computerized, either partly or fully. Such computerization requires some kind of formalization of the compilation procedures, so that they can be translated into computer programmes. Without an established theory of estimation in such complex statistical structures as national accounts, there is the distinct possibility that such computer programmes will differ from country to country, reflecting different manual compilation procedures in the past, and depending on the data sources available and data requirements of the particular national accounts system. Thus, experiences of one country may be of very little help to another as the three elements mentioned may differ considerably among countries. Furthermore, as such computerized systems are presently being designed mainly for developed countries, they tend to be very complex given the many sources available and data required, and are therefore not very useful for countries with much less statistical development.

The present study responds to these concerns by developing a flexible compilation algorithm that is based on the application of a linear programming technique to a system of identities and inequalities that define the functional relations between the variables or data categories included in a national accounting framework. The inequalities which correspond to relations among national accounts categories, and between those and the data available from the basic statistical sources, are similar to the ones that constitute the "assumptions" that

*The views expressed in this study are those of the author and do not necessarily coincide with the official views of the United Nations or any of its member governments.

are used at present to compile national accounts. The difference between the present compilation procedures and the algorithm described here is that the assumed relations are defined within intervals. As this would result in interval estimates of national accounts items, the algorithm incorporates the calculation of so-called central values which are similar to the national accounts estimates arrived at through the traditional compilation procedures. The algorithm is defined such that it can be adapted to any configuration of data sources and data requirements. It is a systematic method that can easily be computerized, given the available software packages on linear programming. The method lends itself in particular to national accounts frameworks that are in their earlier stages of development.

Following this introduction, the paper consists of three sections. In Section I the algorithm is described in abstract terms; it starts with the presentation of the present compilation procedures in formal mathematical terms, and derives the proposed or modified algorithm by adjusting the format of the present compilation procedure. Section II applies the algorithm to two examples. The first one is a very simple one, which serves to illustrate the method graphically. The next application is to a more realistic—but still simple—national accounts framework for Suriname related to data for 1965 that was compiled by the author many years ago. The applications show how maximum and minimum as well as central values can be derived for those national accounts categories for which no direct information is available or for which basic statistical sources do not provide adequate information in terms of coverage and definition of the items. The section ends with a sensitivity analysis, which shows how interval estimates and central values are affected if presumed relations involving unknown national accounts categories are replaced in the course of time by estimates based on direct information from statistical sources. Section III discusses conceptual and operational issues which need to be further examined before the proposed algorithm can be effectively applied to actual compilation procedures for national accounts and related statistics.

## I. The Formal Structure Of The National Accounts Compilation Procedure

### A. *Present National Accounts Compilation Procedure*

The data content of a national accounts scheme and also the national accounts compilation procedures to derive those data are determined by the concepts and classifications used as well as the relations defined between the concepts or variables of the scheme and those available in basic statistical sources. The emphasis of this paper is on the relations between the concepts, as these are the main ingredients of any compilation algorithm. Four kinds of relations can be distinguished:

(i) First there are the identities defined between the variables of the scheme itself. These include the definitions of GDP in terms of expenditure components, activity contributions, income shares and as the difference between gross output and intermediate consumption or as the balance between income and expenditure totals and subtotals in the institutional sector accounts of the national accounts. These identities could be called definitional relationships, as they define one concept or one group of concepts in terms of another.

(ii) A similar group of relationships is defined between information categories in the basic statistical sources and the generally more aggregate categories of the national accounts scheme. These relationships are called here aggregational relations, as they are used in the derivation of the macro-economic categories of the national accounts scheme by aggregating the micro-categories available in the basic sources. For example gross output, intermediate consumption and value added components of the mining sector may be derived—if information is available—by simple aggregation of gross output and the cost components of all establishments that could be classed as mining establishments. Another example is the taxation categories that are used in government statistics, which are derived by simple agregation of the appropriate very detailed tax items that can be identified in the government administrative records.

(iii) A third group of what will be called analytical relations between the variables of the national accounts scheme are those that reflect institutional arrangements such as taxation functions, behaviouristic patterns such as consumption, investment, saving and import functions, and technical relations that are defined between the goods and services flows of a national accounting scheme and in particular between gross output, cost components and capital formation.

(iv) The fourth type are those relations that are typical for the present national accounts compilation process. They are the ones that define the adjustments to the data of the basic sources in order to arrive at macro-economic concepts of the national accounts. They are used in order to correct for lack of coverage or detail in the basic source information, and they involve marginal adjustments needed in order to adapt statistical source concepts to the different coverage defined in the national accounts. These relations are similar to the analytical relations. In fact, use is often made of the analytical relationships in order to correct the deficiencies in the basic data. For example, employment data are used to inflate the survey results for a limited number of production establishments to arrive at the universe as a whole, or production data of similar establishments are utilized to obtain further detail on the intermediate consumption of those establishments for which such information was lacking. Both types of adjustments are based on technical relations between gross output, intermediate consumption and employment data.

If data needed for the derivation of the national accounts aggregates were available in the basic sources, the only type of relation which would be used, and thus would define the national accounts compilation procedure, would be the aggregational relations between the data of the basic sources and the macro concepts. There would be no need to take the national accounts definitional relationships or identities explicitly into account as these would already be present in the basic sources in a disaggregated form. Simple aggregation would guarantee that such identities would then automatically hold in the macro-economic framework of the national accounts as well. Also there would be no need to account explicitly for the analytical relations as they would be available in the basic sources in their micro-format and through simple aggregation would be reflected in the macro concepts. It is also obvious that the adjustments to the basic data would not be relevant, as in the ideal situation, concepts and classifications of the basic sources would coïncide with those of the macro framework of the accounts or at least provide the basic elements for deriving the macro-economic concepts.

In practice, the ideal situation does not exist. Basic sources are lacking in coverage and in detail, the concepts do not coincide with those of the national accounts, and there are inconsistencies between data of different sources. This implies that the macro-economic definitional and analytical relationships will not be fully present in the basic sources and therefore need to be imposed by incorporating those relations in the compilation process. Similarly, adjustments to the basic data are needed in order to bring them in line with the macro-economic concepts.

The present national accounts compilation process can be defined symbolically as the solution to the following set of equations:

$$A1 \cdot X = 0$$

$$X = A2 \cdot Y$$

$$B1_p \cdot X = 0$$

$$X = B2_p \cdot Y.$$

The symbols $X$ and $Y$ are vectors, respectively, of variables in the macro framework of the national accounts and of variables available in the basic sources. Vector $Y$ includes many more components than $X$, as it reflects all the detail of information in the basic sources. $A1$, $A2$, $B1$, $B2$ are matrices of coefficients. Matrix $A1$ reflects the definitional relationships among the variables of the national accounting framework. It includes only three types of values, i.e. $+1$, $-1$ and 0. The same holds for the coefficients included in $A2$ of all aggregational relationships between the variables of the basic sources and those of the macro framework that can be defined with help of available basic information. $B1$ includes the coefficients of the analytical relationships that hold between the data within the national accounts framework. The coefficients included can have other values than the ones mentioned for matrices $A1$ and $A2$. Matrix $B2$ is the matrix which includes the coefficients that define the adjustments in terms of coverage, detail and concepts that have to be made to the basic source data categories in order to obtain the categories that are required in the macro framework of the accounts. This matrix may also include coefficients other than the ones mentioned for the definitional and aggregational matrices $A1$ and $A2$. Present national accounts practices generally use ratios between pairs of variables in defining the coefficient matrices $B1$ and $B2$. This implies that such matrices are of a very simple type, which only include two coefficients per row, one of which is equal to 1. The ratios are between data available in the basic sources and unknown categories of the macro framework and are therefore mainly of the adjustment type $B2$. Those analytical relations between pairs of variables within the national accounts framework (matrix $B1$), where one of the variables is derived from the basic data through simple aggregation ($A2$) may also be included. Present national accounts practices do not involve other analytical relations between unknown variables in the national accounts framework.

The definitional relations defined by matrix $A1$ will include only those identities that are not present in the basic statistical sources, either because of lack of detail or coverage or conceptual discrepancies between the basic source categories and those of the macro-economic national accounting framework, or because of inconsistencies between the data of different statistical sources. Identities that are already present in the basic sources do not need to be repeated in defining the compilation procedure. In principle, all independent aggregational relationships between basic sources and the macro framework are included in $A2$. When adding up the number of relations thus incorporated in defining the compilation procedure, one generally would arrive at a number that is smaller than the number of national accounts variables. In order to obtain a "solution" of the national accounts model, the number of relations included is then increased by a selection from the available analytical and adjustment relations, so that the total number of functional relationships is equal to the number of national accounts variables to be compiled. This partial selection is reflected in the $p$ subscripts of $B1$ and $B2$. The national accounts "solution" arrived at is therefore not a unique solution, but changes with the selection of relationships included in $B1$ and $B2$. Present compilation procedures try to offset this disadvantage by checking, in an *ad hoc* manner, estimates based on one selection of analytical and adjustment functions with other relations that also should hold. For instance, if the use of the commodity flow method results is unacceptable values for value added in particular industries, the assumptions built into the commodity flow method are reexamined and revised so that a more "acceptable" level of value added is obtained for those sectors. This iterative checking of estimates is, however, not necessarily a systematic procedure which would involve all relations that could potentially be included in $B1$ and $B2$. The result is that national accounts data may show up relationships that are not acceptable when used in particular types of analysis. If the national accountant had incorporated those relationships in his compilation procedure, such deficiencies might have been avoided.

## B. *General Principles of the Modified Compilation Algorithm*

The major restriction of the present compilation procedure used in national accounting is the one that determines the number of analytical and adjustment relations that can be incorporated in a systematic manner. It is proposed in this paper that this restriction be removed by replacing the single values of the analytical and adjustment parameters in the $B1$ and $B2$ matrices by intervals that are defined by maximum and minimum values of those parameters. This implies that the $B1$ and $B2$ matrices would be replaced by pairs of matrices, one including the minimum values of the $B1$ and $B2$ coefficients and another the maximum values of those parameters. The equality signs in the two expressions that include $B1$ and $B2$ will then be replaced by inequality signs, such that the expressions including the minimum values of the two matrices would be presented with $\geq$ signs and the maximum values with $\leq$ signs. The modified compilation procedure could then be defined as finding the maxima and minima of the variables included in the $X$-vector which are within the bounds of the following set of identities and inequalities.

$$A1 \cdot X = 0$$

$$X = A2 \cdot Y$$

$$B1 \min \cdot X \geq 0$$

$$B1 \max \cdot X \leq 0$$

$$X \geq B2 \min \cdot Y$$
$$X \leq B2 \max \cdot Y.$$

In the symbolic presentation above, $B1$ and $B2$ in their maximum as well as in their minimum value versions no longer include the restrictive subscript $p$. This means that in principle the new procedure may incorporate as many analytical and adjustment relations as required or available, provided of course that the intervals do not conflict with each other. The ultimate intervals of national accounts estimates that result from this computation procedure will then be consistent with many more analytical and adjustment relations than the present procedure could ever satisfy. This formulation furthermore opens the possibility that not only will relations be included between known basic source variables and unknown national accounts aggregates and between unknown national accounts aggregates and aggregates that are derived with the help of the simple aggregational relationships as mentioned above, but also that analytical relations can be defined among two or more unknown variables in the national accounts scheme itself. This enhances the analytical usefulness of the national accounts variables, as they will be consistent with many more analytical and adjustment relationships that have never entered the national accounts compilation process before. There is furthermore the possibility of incorporating, in addition to ratios that are presently used in national accounts compilation procedures, analytical functions that have been derived as a result of econometric or related research. The latter possibility would contribute to further integration of national accounts compilation with econometric and related research.

The application of the linear programming algorithm determines a multi-dimensional region within which all feasible $X$-solution vectors are located. The minimum and maximum values of the components of $X$ are to be found on the edges of this region, and are included in what might be called extreme value vectors. In order to arrive at singular values for each of the components of $X$ similar to the estimates that result from the present national accounts compilation procedures, the algorithm should include additionally the derivation of a solution vector of so-called central values which might be defined as a function of the extreme value vectors that include the minima and maxima of the components $X$. The precise definition of this functional relationship depends on the criteria that single valued estimates should satisfy. No definite conclusions regarding these criteria have been reached in this study, but preliminary suggestions on the theoretical concept of central values are presented in section III, and a practical illustration is worked out in the examples of section II.

## II. Illustration Of The Use Of The Modified Algorithm In The Compilation Of National Accounts Data Schemes

### A. *Simple Example Illustrating the Present and Modified Compilation Procedures*

The more abstract formulations in the previous section of the present and modified national accounts compilation procedures will be first illustrated with the help of a very simple example. The present compilation is reflected in Table 1 and the modified algorithm in Table 2. Both tables assume a definitional relationship based on the expenditure approach to GDP. The variables indicated with a bar in the two tables (for instance $\bar{C}_g$) are those national accounts aggregates that are derived directly from available basic statistical surveys (e.g. economic censuses and surveys, government records and foreign trade and balance of payments information) with help of the simple aggregational relationships earlier

TABLE 1

Alternative Solutions to a Simple National Accounts Model based on Present National Accounts Compilation Practice

Definitional relationship:

$\bar{Y} = Ch + \bar{C}g + If + \bar{I}s + \bar{E}x - \bar{M}$
$\bar{Y}$ = GDP at market prices (230)
$Ch$ = Private final consumption expenditure
$\bar{C}g$ = Governmental final consumption expenditure (37)
$If$ = Gross fixed capital formation
$\bar{I}s$ = Increases in stocks (6)
$\bar{E}x$ = Exports of goods and services (120)
$\bar{M}$ = Imports of goods and services (200)

or

$$Ch + If = 274$$

Analytical and adjustment relations:

$$Ch = 3\ If$$
$$If = 0.2\ \bar{Y}$$
$$Ch = 1.05\ \bar{C}ho\ (\bar{C}ho = 210)$$

Alternative solutions of $Ch$ and $If$:

| | $Ch = 3\ If$ | $If = 0.2\ \bar{Y}$ | $Ch = 1.05\ \bar{C}ho$ |
|---|---|---|---|
| $Ch$ | 205.5 | 228.0 | 220.5 |
| $If$ | 68.5 | 46.0 | 53.5 |

TABLE 2

MAXIMUM, MINIMUM AND CENTRAL VALUES DERIVED FROM A SIMPLE NATIONAL ACCOUNTS MODEL ON THE BASIS OF THE MODIFIED NATIONAL ACCOUNTS COMPILATION ALGORITHM

Definitional, analytical and adjustments relations:

$$Ch + If = 274$$
$$2.5 \leq Ch/If \leq 8$$
$$0.1 \leq If/\bar{Y} \leq 0.3$$
$$1.03 \leq Ch/\bar{C}ho \leq 1.08$$

or

$$Ch + If = 274$$
$$Ch - 2.5\ If \geq 0$$
$$Ch - 8\ 8\ If \leq 0$$
$$If \geq 0.1\ \bar{Y}$$
$$If \leq 0.3\ \bar{Y}$$
$$Ch \geq 1.03\ \bar{C}ho$$
$$Ch \leq 1.08\ \bar{C}ho$$

Maximum, minimum and central values:

| | Minimum value | | | Maximum value | |
|---|---|---|---|---|---|
| | *Ex ante* | *Ex post* | Central value | *Ex post* | *Ex ante* |
| *Ch* | — | 216.3 | 221.6 | 226.8 | — |
| *If* | — | 47.2 | 52.4 | 57.7 | — |
| $Ch/If$ | 2.50 | 3.75 | 4.23 | 4.81 | 8.00 |
| $If/\bar{Y}$ | 0.10 | 0.21 | 0.23 | 0.25 | 0.30 |
| $Ch/\bar{C}ho$ | 1.03 | 1.03 | 1.06 | 1.08 | 1.08 |

mentioned. Total GDP ($Y$) in the example is measured with help of the alternative production or income approaches. The aggregational relationships have not been explicitly incorporated in this example; instead the resulting values of the barred variables have been included and indicated in brackets ( ) following the definitions of those variables. The unknown variables in this national accounts model are gross fixed capital formation ($If$) and private final consumption expenditure ($Ch$). Given the values of the known variables, the definitional relationship reduces to a simple sum of $If$ and $Ch$ equal to 274.

Three types of analytical and adjustment relations are used to estimate the two unknown variables. The first one, which is based on the commodity flow method applied to further detail of gross output, imports and known expenditure categories, results in a distribution of the commodity flows between gross fixed capital formation ($If$) and private final consumption expenditure ($Ch$). The second one relates the value of capital formation to GDP on the basis of past measurements of capital output ratios. A third relation assumes a growth rate of private final consumption expenditure ($Ch$) as compared with its last year's value and based on past experience with regard to the link between the growth rates of consumption and GDP.

In the application of the present compilation methodology in Table 1 the parameters of the analytical and adjustment relations have been given specific values. It has been assumed that the application of the commodity flow method results in a value of private final consumption ($Ch$) that is 3 times the amount of gross fixed capital formation ($If$). Past experience has furthermore shown that gross fixed capital formation (*If*) is 20 percent of GDP ($\bar{Y}$) and that the growth rate of consumption might be 5 percent since last year. As only one of these three types of information is needed jointly with the national accounts identity in order to solve for the value of *If* and *Ch*, there are in fact three solutions to the model and all are different. The commodity flow assumption leads to the lowest value of *Ch* and the highest value of *If*, while the assumption regarding the capital-output ratio results in the highest value of *Ch* and the lowest value of *If*. The assumption about the growth of consumption gives results that are intermediate between the two levels.

It is obviously unacceptable that the three assumptions would lead to such different results. This reflects on the quality of the assumptions as well as on the data regarding the known variables. National accountants using the present methodology are aware of this problem and to a limited extent try to resolve it by confronting the varying solutions and making adjustments to the assumed parameters and data in an iterative manner, so that ultimately one acceptable solution evolves. If no reconciliation is feasible, they would generally accept one of the solutions as the most reliable one with which other data are reconciled, or introduce statistical discrepancies to reflect the different values of the solutions.

The modified algorithm as illustrated in Table 2 avoids the irreconciled or irreconcilable solutions by replacing the fixed values of the parameters by ranges and the equality signs in the analytical and adjustments relations by inequality signs. Instead of assuming that consumption is 3 times capital formation it is defined within a range of 2.5 to 8 times the values of capital formation, taking into account the reliability of the data and assumptions used in the commodity flow method. Similarly, the ratio between capital formation and output is not fixed at 0.2 but defined between 0.1 and 0.3, and the growth rate of consumption (based on a consumption figure in the base period $\bar{Cho} = 210$) is assumed to be between 3 and 8 per cent, reflecting variations in the past in these parameters. The solution to the modified compilation model presented in table 2 is not a unique one, but rather consists of a set of intervals for the variables and parameters of the model. These ranges which are called *ex post* intervals are distinguished from the *ex ante* intervals of the parameters which are defined at the outset of the model. The table also presents singular values for each of the variables similar to those resulting from the traditional compilation methods. These are called central values and are defined as the components of a solution vector that is the unweighted average of all extreme solution vectors that include *ex post* minima and maxima of the unknown variables and parameters in conformity with the restrictions of the compilation model.

As the example is simple, the *ex ante* and *ex post* intervals and the central values resulting from the application of the modified compilation algorithm in Table 2 can be graphically illustrated. This is done in diagram A. The co-ordinates of the diagram correspond to the two unknown variables in the compilation model, i.e. *Ch* is presented on the horizontal axis and *If* vertically. The GDP identity is represented by one line in the diagram and the *ex ante* analytical and adjustment restrictions by areas between pairs of lines that are defined by pairs of inequalities of the model. The solution is to be found on that part of the line of the definitional identity of GDP that coincides with an area commonly defined

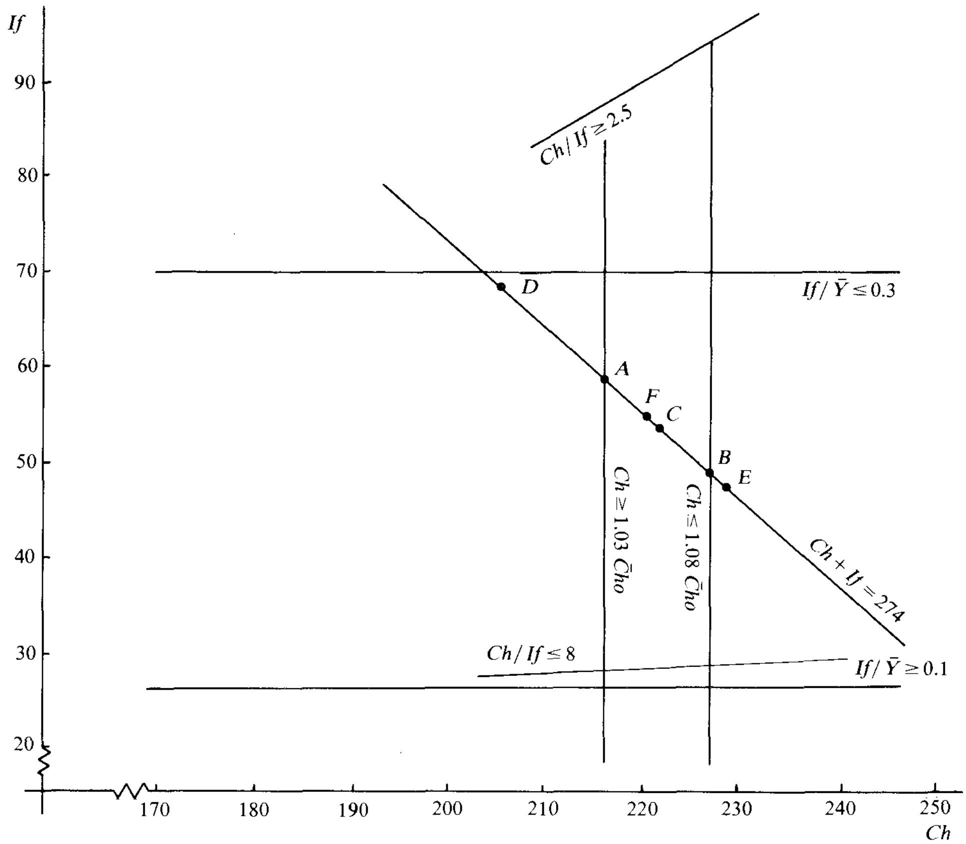


Diagram A. Graphical Representation of the Simple National Accounts Compilation Model

by all analytical and adjustment relations, i.e. between points A and B. Point C is the central value which is defined as the unweighted average of solution vectors A and B. Points D, E and F represent the three solutions shown in Table 1, that would be the result of the present compilation procedure. Each of these three solutions is also on the GDP identity line. Point A corresponds to the solution vector (*Ch*, *If* = 216.3, 57.7), B to the solution vector (226.8, 47.2) and the central value C vector is (221.6, 52.4). The alternative solution vectors of the present compilation method D, E and F are respectively (205.5, 68.5), (228.0, 46.0) and (220.5, 53.5).

Table 2 and diagram A show that there is a considerable reduction in the intervals of two of the three parameters after confrontation of the *ex ante* intervals of the analytical and adjustment relations and the definitional identity. The 2.5-8 *ex ante* interval of $Ch/If$ reduces *ex post* to 3.75-4.81 and the 0.10-0.30 interval of $If/\bar{Y}$ reduces to 0.21-0.25. The $Ch/\bar{C}ho$ *ex post* interval remains the same compared to the *ex ante* interval and thus determines the *ex post* intervals of all other parameters and variables in the model. The reason that there are two of such limiting *ex ante* extremes—i.e. the minimum and maximum of $Ch/\bar{C}ho$—is because there are two variables (*If* and *Ch*) to be solved. The central values are simple averages of the two solution vectors A and B, and therefore automatically satisfy the definitional relationships as well as the *ex post* ranges of variables and parameters. From the presentation of the diagram it also becomes clear that two of three alternative solutions D and E that result from the traditional compilation procedure are located outside the *ex post* range AB. These solutions therefore do not satisfy simultaneously all *ex ante* ranges of the parameter intervals, which illustrates the possibility of arriving at erroneous results when applying the traditional compilation methods.

## B. *Application of the Modified Algorithm to an Extended National Accounts Scheme*

The simple example presented above was sufficient to illustrate the general principles of the modified algorithm. However, it cannot reflect the implications for the solution procedure when applied to a more realistic national accounting scheme which would involve many more variables and parameters. To illustrate this, use is made of preliminary data that were compiled by the author for Suriname covering the year 1965, in an earlier attempt to develop the compilation algorithm presented in this study. The data included were later on used and further improved in the actual compilation of national accounts for Suriname for that year on the basis of the traditional national accounts compilation procedure.

The national accounts scheme is limited to production account information. The data categories directly estimated and used in the compilation are presented in sections (a), (b), (c) and (d) of Table 3 below. Section (a) includes production accounts for four key sectors: agriculture, bauxite-aluminum industry, other industries and the public sector. Agriculture includes the traditional small scale agriculture, as well as the larger plantations that apply more advanced production methods. The bauxite-aluminum industry consists of a limited number of enterprises which dominate the mineral exploitation of bauxite and its processing to alumina and aluminum and are also a major source of supply of electricity for the country. Other industries cover the remaining private enterprises that operate in such diverse production processes as timber production and processing, construction, simple production processes as bottling of imported beverages, and also all service activities which are mainly operated on a small scale. The public sector includes general government activities as well as government enterprises such as the postal and telephone system, port authority, Suriname Airlines, agricultural experimental stations, railroad and bus services, government car repair shops, etc. The production accounts only include aggregate categories of output and cost: gross output, intermediate consumption, and value added broken down by compensation of employees and the rest of value added covering operating surplus as well as depreciation, indirect taxes and subsidies. In each of the production accounts imported goods and services are separately identified within intermediate consumption.

Section (b) of Table 3 presents the breakdown of GDP by cost components. It includes compensation of employees for all sectors together and shows a breakdown of the remaining value added by gross operating surplus and indirect taxes and import duties. Section (c) of the table includes a breakdown of GDP

TABLE 3

DATA CATEGORIES OF THE EXTENDED NATIONAL ACOUNTING SCHEME

| | Agriculture | Bauxite-Aluminum Sector | Other Industries | Public Sector |
|---|---|---|---|---|
| (a) PRODUCTION ACCOUNTS | | | | |
| Gross output | 48,628 | 172,311 | $Xo$ | 71,634 |
| Intermediate consumption | 9,505 | 103,472 | $Co$ | 30,182 ($\bar{C}g$) |
| Of which: Imports | 4,283 ($\bar{M}ca$) | 77,838 ($\bar{M}cb$) | $Mco$ | |
| Value added | 39,163 $\bar{Y}a$) | 68,839 | $Yo$ | 41,452 |
| Compensation of employees | $Sa$ | 25,902 ($\bar{S}b$) | $So$ | 41,452 ($\bar{S}g$) |
| Other value added | $Pa$ | 42,937 ($\bar{P}b$) | $Po$ | — |
| (b) COST COMPONENTS OF GDP | | | | |
| Compensation of employees | | $S$ | | |
| Other value added | | $P'$ | | |
| Indirect taxes | | 37,471 | | |
| Of which: Import duties | | 28,380 | | |
| Operating surplus, gross | | $P$ | | |
| Gross domestic product | | $Y$ | | |
| (c) GDP BY TYPE OF FINAL EXPENDITURES | | | | |
| Final consumption expenditure | | | | |
| Households | | $Ch$ | | |
| Government | | 59,734 ($\bar{X}g$) | | |
| Gross output | | 71,634 | | |
| Minus: Government sales | | −11,900 | | |
| Of which: Imports | | 43,800 ($\bar{M}c$) | | |
| Gross fixed capital formation | | $If$ | | |
| Dwellings | | 9,782 ($\bar{I}fh$) | | |
| Bauxite-aluminum sector | | 78,061 ($\bar{I}fb$) | | |
| Agriculture and other industries | | $Ifao$ | | |
| Government | | 25,703 ($\bar{I}fg$) | | |
| Of which: Imports | | $Mif$ | | |
| Changes in stocks | | 6,014 ($\bar{I}s$) | | |
| Exports | | 123,049 ($\bar{E}x$) | | |
| Agriculture | | 11,981 | | |
| Bauxite-aluminum sector | | 87,729 | | |
| Other industries | | 21,882 | | |
| Government | | 1,457 | | |
| Minus: Imports | | 199,100 ($\bar{M}$) | | |
| Gross Domestic Product | | $Y$ | | |
| (d) OTHER BASIC DATA | | | | |
| Compensation of employees of large enterprises in other industries | | 39,221 ($\bar{S}o$) | | |
| Investments made by large enterprises, other industries | | 11,848 ($\bar{I}fo$) | | |
| Imports of investment goods identified, other industries | | 41,600 ($\bar{M}if$) | | |
| Corporate income tax (other than paid by the bauxite-aluminum sector) | | 1,568 ($\bar{T}dao$) | | |
| Income tax paid by individuals | | 6,451 ($\bar{T}dh$) | | |

by categories of final demand, i.e. final expenditure of households and government, gross fixed capital formation, changes in stocks, exports and imports. The import components of final consumption expenditure and gross fixed capital formation are separately identified. Government final consumption expenditures are deconsolidated into gross output minus sales of the public sector (including sales by government enterprises incorporated in the public sector), and gross fixed capital formation and exports are broken down by sector.

The intermediate data available for the exercise was the following. Annual agricultural surveys provided information on gross output by type of agricultural products and a survey of agricultural holdings included data on intermediate consumption. Direct production account data were available from the few enterprises operating in the bauxite-aluminum industry. This included data on gross output, intermediate consumption, imports, the breakdown of value added, exports, gross fixed capital formation and changes in stocks. Government administrative records were the main source of data for the public sector. For the remaining industries an annual survey of large enterprises was conducted in co-operation with an auditing firm, which supplied data on compensation of employees and capital formation only. Foreign trade statistics included all needed detail on exports by each of the sectors distinguished. Among imports of goods it was possible to identify imports for the agricultural and bauxite-aluminum sectors, and a limited number of other items could be identified as either consumption goods or goods destined for capital formation. The remaining import goods could not be further identified. Balance of payments statistics compiled by the Central Bank could distinguish between imports of services for each of the sectors. Additional information available from tax records supplied data on direct taxes paid by individuals, enterprises of the bauxite and aluminum industry, and enterprises operating in agriculture and other industries, as well as total indirect taxes and import duties collected by the government. All direct information on national accounts categories that was used in the application of the algorithm is presented in quantitative form in sections (a), (b) and (c) of Table 3. Data that does not conform to the definitions and/or coverage of the categories of the national accounting scheme but is used in the compilation is described and quantified in section (d) of the table.

To estimate the unknown variables, definitional, analytical and adjustment relations have been defined between known and unknown variables as well as among unknown variables. The functional relationships and the intervals for the parameters are presented in Table 4. The variables that are used in the relations are defined in the four sections of Table 3. The main variables are gross output ($X$), value added and GDP ($Y$), intermediate and final consumption ($C$), gross capital formation ($I$), exports ($Ex$), imports ($M$), compensation of employees ($S$) and other value added ($P$). Subscripts and primes (′) are added to denote further subdivisions, sectors and different definitions of the same concept. Known categories of information that could be directly derived from the basic data are indicated with a bar (e.g. $\bar{C}g$) and unknown variables are presented without a bar (e.g. $S$).

Four definitional relationships have been defined that are all independent of each other and include at least some unknown variables. The first one is the

TABLE 4

IDENTITIES AND ADJUSTMENT AND ANALYTICAL RELATIONS DEFINED IN THE EXTENDED NATIONAL ACCOUNTS SCHEME*

*Identities*

$$Pa + \bar{P}b + (Xo + Co - So) + S = Ch + \bar{X}g + (Ifao + \bar{I}fh + \bar{I}fb + \bar{I}fg) + \bar{I}s + \bar{E}x - (\bar{M}g + \bar{M}s) \quad (1)$$

$$\bar{M}g = \bar{M}c + \bar{M}ca + \bar{M}ob + Mco + Mif \quad (2)$$

$$S = Sa + \bar{S}b + So + \bar{S}g \quad (3)$$

$$\bar{Y}a = Sa + Pa \quad (4)$$

*Adjustment relations*

$$1 \leq Mif/\bar{M}if \leq 2 \quad (5)$$
(or *lmif*)

$$1.5 \leq Ifao/\bar{I}fo \leq 2.5 \quad (6)$$
*lifao*)

$$1.9 \leq So/\bar{S}o \leq 2.4 \quad (7)$$
(or *lso*)

*Analytical relations*

$$0.25 \leq \bar{M}c/(Ch + \bar{C}g) \leq 0.6 \quad (8)$$
(or *mc*)

$$0.25 \leq Mif/(Ifao + \bar{I}fh + \bar{I}fb + \bar{I}fg) \leq 0.6 \quad (9)$$

$$0.25 \leq Mco/Co \leq 0.6 \quad (10)$$
(or *mco*)

$$0.3 \leq Co/Xo \leq 0.7 \quad (11)$$
(or *co*)

$$0.2 \leq So/Xo \leq 0.7 \quad (12)$$
(or *so*)

$$0.55 \leq Sa/\bar{Y}a \leq 0.95 \quad (13)$$
(or *sa*)

$$0.7 \leq Ch/S \leq 0.95 \quad (14)$$
(or *ch*)

$$0.8 \leq Ifao/(Xo - Co - So + Pa) \quad (15)$$
(or *ifao*)

$$0.03 \leq \bar{T}dh/S \leq 0.06 \quad (16)$$
(or *tdh*)

$$0.03 \leq \bar{T}dao/(Pa + Xo - Co - So) \leq 0.25 \quad (17)$$
(or *tdao*)

*For an explanation of the symbols of the variables, refer to Table 3.

identity between expenditure minus imports and the total of income shares of GDP. The second identity shows how the total imports of goods are broken down by goods imported for purposes of intermediate consumption for the three production sectors, imported consumption goods for public and private consumption, and imported capital goods. The third identity stands for the breakdown of total compensation of employees by industrial sectors. The last identity used defines value added of the agricultural sector as the sum of compensation of employees and the non-labour income component of value added in that sector.

Three adjustment relations are defined between the incomplete data of the basic sources and the complete coverage of the items that should be reflected in the national accounts scheme. The first one assumes that actual imports of capital goods are between 1 and 2 times the amount that could be identified as capital goods imports in the foreign trade statistics. The second relation assumes that gross fixed capital formation in agriculture and other industries is between 1.5 and 2.5 times the amount of capital goods purchased by the large industrial enterprises surveyed in co-operation with an auditing firm. It was further assumed that compensation of employees paid out by the large enterprises surveyed should be multiplied by a factor between 1.9 and 2.4 in order to arrive at compensation of employees for the total of other industries. The intervals are either based on expert judgement (using large intervals) or past experience.

Expert judgement and past experience were also used to define lower and upper bounds for ratios reflecting six analytical relations between variables. Identified imports of consumption goods were assumed to be between 0.25 and 0.6 of final consumption by households and intermediate consumption by the government sector. The same interval was assumed for the ratio between total imports of capital goods and capital formation in agriculture and other industries and between imports of goods and services for intermediate consumption by other industries and intermediate consumption of those industries. Intermediate consumption of other industries was assumed to be between 0.3 and 0.7 of gross output of other industries and compensation of employees between 0.2 and 0.7 of gross output of other industries. In agriculture it was assumed that compensation of employees was between 0.55 and 0.95 of total value added of agriculture. Household consumption was considered to be between 0.7 and 0.95 of compensation of employees. Gross fixed capital formation in agriculture and other industries was assumed to be larger than 0.8 of non-labour income generated in these sectors. Direct taxes paid by individuals were fixed between 0.03 and 0.06 of compensation of employees and direct taxes paid by corporations and other enterprises operating in agriculture and other industries between 0.03 and 0.25 of non-labour income generated by those sectors.

The identities and lower and upper bounds of parameters of the adjustment and analytical relations are rewritten in Table 5 in linear programming (lp) format in order to apply a linear programming technique for deriving the *ex post* minima and maxima of the unknown variables of the national accounts model. This technique replaces the graphical analysis earlier applied to the simple model in section II A, which is not suitable for complex national accounts models with a large number of variables. The relations are divided in the table in three groups: the first group are the identities; the second group are the inequalities that include more than one unknown variable; the third group includes those inequalities with one unknown variable which is defined in terms of known estimates directly derived from available statistical sources. The reference codes of the relations used in this table are the same as those of Table 4. The analytical and adjustment relations are presented in pairs of two, which define the minimum ($\geq$) and maximum ($\leq$) values of the unknown national accounts categories in terms of the known ones. The unknown categories are presented on the left-hand side of the equality and inequality signs and the known categories on the right-hand side.

To arrive at the *ex post* minima and maxima, the inequalities of Table 5 are converted to equalities by introducing slack variables that are defined to be equal

TABLE 5

LINEAR PROGRAMMING FORMAT OF THE IDENTITIES AND ANALYTICAL AND ADJUSTMENT RELATIONS OF THE EXTENDED NATIONAL ACCOUNTS SCHEME*

| Unknown national accounts variables | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *Xo* | *Co* | *So* | *Ifao* | *Pa* | *Sa* | *Mif* | *Mco* | *Ch* | *S* | | |
| +1 | −1 | −1 | −1 | +1 | | | | −1 | +1 | $= -\bar{P}b + (\bar{I}fh + \bar{I}fb + \bar{I}fg) + \bar{I}s + \bar{E}x - (\bar{M}g + \bar{M}s)$ | (1) |
| | | | | | | +1 | +1 | | | $= Mg - (\bar{M}c + \bar{M}ca + \bar{M}cb)$ | (2) |
| | | −1 | | | −1 | | | | +1 | $= \bar{S}b + \bar{S}g$ | (3) |
| | | | | +1 | +1 | | | | | $= \bar{Y}a$ | (4) |
| | | | −0.25 | | | +1 | | | | $\geq 0.25\,(\bar{I}fh + \bar{I}fb + \bar{I}fg)$ | (9) |
| | | | −0.6 | | | +1 | | | | $\leq 0.6\,(\bar{I}fh + \bar{I}fb + \bar{I}fg)$ | |
| | | | | | | | | +1 | −0.7 | $\geq 0$ | (14) |
| | | | | | | | | +1 | −0.95 | $\leq 0$ | |
| −0.2 | | +1 | | | | | | | | $\geq 0$ | (12) |
| −0.7 | | +1 | | | | | | | | $\leq 0$ | |
| −0.8 | +0.8 | +0.8 | +1 | −0.8 | | | | | | $\geq 0$ | (15) |
| −0.3 | +1 | | | | | | | | | $\geq 0$ | (11) |
| −0.7 | +1 | | | | | | | | | | |
| | −0.25 | | | | | | +1 | | | $\geq 0$ | (10) |
| | −0.6 | | | | | | +1 | | | $\leq 0$ | |
| +1 | −1 | −1 | | +1 | | | | | | $\geq \bar{T}dao/0.25$ | (17) |
| +1 | −1 | −1 | | +1 | | | | | | $\leq \bar{T}dao/0.03$ | |
| | | | | | | +1 | | | | $\geq \bar{M}if$ | (5) |
| | | | | | | +1 | | | | $\leq 2\,\bar{M}if$ | |
| | | | +1 | | | | | | | $\geq 1.5\,\bar{I}fo$ | (6) |
| | | | +1 | | | | | | | $\leq 2.5\,\bar{I}fo$ | |
| | | +1 | | | | | | | | $\geq 1.9\,\bar{S}o$ | (7) |
| | | +1 | | | | | | | | $\leq 2.4\,\bar{S}o$ | |
| | | | | | | | | +1 | | $\geq \bar{M}c/0.6 - \bar{C}g$ | (8) |
| | | | | | | | | +1 | | $\leq \bar{M}c/0.25 - \bar{C}g$ | |
| | | | | | +1 | | | | | $\geq 0.55\,\bar{Y}a$ | (13) |
| | | | | | +1 | | | | | $\leq 0.95\,\bar{Y}a$ | |
| | | | | | | | | | +1 | $\geq \bar{T}dh/0.06$ | (16) |
| | | | | | | | | | +1 | $\leq \bar{T}dh/0.03$ | |

*For explanation of the symbols of the variables and parameters, refer to Tables 3 and 4, respectively.

to the difference between the left and right hand side values of the inequality expressions. The *ex post* extremes are then obtained as the optimum values of objective functions that are subsequently defined in terms of each of the unknown variables separately, and subject to all restrictions defined in the table. The subsequent solutions can be represented by vectors, each of which includes the maximum or minimum of at least the one variable that was optimized and extreme or non-extreme values for all other unknown variables. These so-called extreme value vectors are then used to derive a central value vector which contains for each of the unknown variables the unweighted averages of the corresponding values included in each of the extreme value vectors. The values of the components of the central value vector thus defined automatically satisfy the *ex ante* restrictions of the national accounts model and are also consistent with the *ex post* maxima and minima of the variables.

The complexity of the modified algorithm is determined by the size of the lp matrix. This size is dependent on the number of unknown variables and slack variables and on the number of identities and inequalities which define the compilation procedure. It should be noted here that the inequalities always appear in pairs, which considerably increases the size of the matrix. Not reflected in the lp matrix are the inequalities with only one unknown variable, as these can be redefined as minimum and maximum values of unknown variables, and accommodated without enlarging the lp matrix by using a modified lp technique [1] based on upper and lower bounds of unknown variables. Further reductions in the size of the matrix were effected by not including those unknown variables that can be defined as linear combinations of the other variables. This reduced the number of unknown variables from the 16 explicitly presented in Table 3 to the 10 included in the lp presentation of Table 5. For instance, GDP ($Y$) was not included by defining it as the sum of the value added components in Tables 3(a) and 3(b) or as the sum of the expenditure categories presented in Table 3(c). Total operating surplus ($P$) was left out by defining it as the sum of operating surplus of each of the sectors separately. Other value added of other industries ($P_0$) was not included by defining it as the difference between gross output and intermediate consumption and compensation of employees of other industries.

The results of the lp iteration have been presented in Table 6. The first section of the table shows the *ex post* extremes as well as central values of the unknown national accounts categories. Also included in this part of the table is the relative size of the *ex post* intervals which, for each of the variables is expressed as a percentage of their central values. The second part of the table shows the central values and *ex ante* and *ex post* minima and maxima of the parameters or ratios defined in Table 4, and also includes in the last two columns the relative sizes of the *ex ante* and *ex post* intervals expressed as percentages of the central values of these parameters.

Conclusions similar to those for the simple example examined in section IIA can be drawn from Table 6 for the application of the lp algorithm to the extended national accounts model. Of the *ex ante* extremes of the parameters, 10 remain unchanged when measured *ex post* (i.e. the minimum and maximum values of *lifao*, *lso* and *sa* and the minimum values of *mco*, *co*, *ch* and *ifao*). The number of these extremes correspond to the number of unknown variables in the national accounts model. They are the ones that determine changes in the *ex post* as compared to the *ex ante* values of the remaining extremes of the parameters shown in the second section of table, and also the *ex post* intervals of the national accounts variables presented in the first section. Some of the parameter intervals have been considerably reduced as compared to their *ex ante* length. This applies in particular to the parameters of imports (*lmif*, *mif*, *mc*, *mco*), final consumption (*ch*), intermediate consumption of other industries (*co*) and the tax parameters (*tdh* and *tdao*). With regard to the unknown variables in the model (the first section of Table 6), the relative *ex post* intervals are fairly small for imports (*Mco*, *Mif*), gross output, intermediate consumption and value added of other industries (*Xo*, *Co*, *So*), for the main expenditure and cost components of GDP (*Ch*, *If*, *S*, *P*) and for GDP ($Y$) itself. Further breakdowns of items by detailed cost components and expenditure categories (*Sa*, *Pa*, *So*, *Po* and *Ifao*) have *ex post* intervals that are much wider. The central values presented in the table are different (sometimes lower, sometimes higher) from the simple mid-points

TABLE 6

*EX ANTE* AND *EX POST* MINIMUM, MAXIMUM AND CENTRAL VALUES OF VARIABLES AND PARAMETERS OF THE EXTENDED NATIONAL ACCOUNTS SCHEME, PRIOR TO AND OBTAINED AS A RESULT OF THE APLICATION OF THE MODIFIED NATIONAL ACCOUNTS COMPILATION ALGORITHM

| | Minimum Values | | Central Values | Maximum Values | | Interval Length as a Percentage of the Central Value | |
|---|---|---|---|---|---|---|---|
| | *Ex ante* | *Ex post* | | *Ex post* | *Ex ante* | *Ex ante* | *Ex post* |
| Variables | | | | | | | |
| *Sa* | | 21,540 | 32,914 | 37,205 | | | 47.6 |
| *Pa* | | 1,958 | 6,249 | 17,623 | | | 250.7 |
| *Xo* | | 145,583 | 153,294 | 159,720 | | | 9.2 |
| *Co* | | 43,675 | 46,087 | 47,916 | | | 9.2 |
| *Mco* | | 10,919 | 11,776 | 11,979 | | | 9.0 |
| *Yo* | | 101,908 | 107,207 | 111,804 | | | 9.2 |
| *So* | | 74.520 | 85,744 | 94,130 | | | 22.9 |
| *Po* | | 7,778 | 21,463 | 35,067 | | | 127.1 |
| *S* | | 176,337 | 186,012 | 198,689 | | | 12.0 |
| *P'* | | 61,408 | 70,648 | 79,962 | | | 26.2 |
| *P* | | 23,937 | 33,177 | 42,491 | | | 55.9 |
| *Ch* | | 123,436 | 130,539 | 140,243 | | | 12.9 |
| *If* | | 131,318 | 136,425 | 143,166 | | | 8.7 |
| *Ifao* | | 17,772 | 22,879 | 29,620 | | | 51.8 |
| *Mif* | | 41,600 | 41,803 | 42,660 | | | 2.5 |
| *Y* | | 251,362 | 256,661 | 261,258 | | | 3.9 |
| Parameters | | | | | | | |
| *lmif* | 1 | 1 | 1.005 | 1.025 | 2 | 99.5 | 2.5 |
| *lifao* | 1.5 | 1.5 | 1.931 | 2.5 | 2.5 | 51.8 | 51.8 |
| *lso* | 1.9 | 1.9 | 2.186 | 2.4 | 2.4 | 22.9 | 22.9 |
| *mif* | 0.25 | 0.291 | 0.306 | 0.325 | 0.6 | 114.3 | 11.1 |
| *mc* | 0.25 | 0.257 | 0.273 | 0.285 | 0.6 | 128.2 | 10.3 |
| *mco* | 0.25 | 0.25 | 0.256 | 0.274 | 0.6 | 136.7 | 9.4 |
| *sa* | 0.55 | 0.55 | 0.840 | 0.95 | 0.95 | 47.6 | 47.6 |
| *so* | 0.2 | 0.476 | 0.559 | 0.647 | 0.7 | 89.4 | 30.6 |
| *co* | 0.3 | 0.300 | 0.301 | 0.310 | 0.7 | 132.9 | 3.3 |
| *ch* | 0.7 | 0.700 | 0.702 | 0.715 | 0.95 | 35.6 | 2.1 |
| *ifao* | 0.8 | 0.800 | 0.826 | 0.964 | — | — | 19.9 |
| *tdh* | 0.03 | 0.032 | 0.035 | 0.037 | 0.06 | 85.7 | 13.2 |
| *tdao* | 0.03 | 0.42 | 0.057 | 0.080 | 0.25 | 386.0 | 13.3 |

between the minimum and maximum value for each of the variables. The latter types of averages cannot be used as central values, as they generally would not satisfy the identities and analytical constraints of the national accounts model.

Conclusions can also be drawn with regard to the direction in which national accounts estimates could be improved, once new sources of data become available or existing records are analysed in a more effective manner. Such improvements imply that the indirect measurement of selected national accounts items based on the lp algorithm is replaced by direct measurement of those items on the basis of newly available statistical information. Some of the effects of such replacement can be appreciated from the information included in Table 7. The first column of the table is the same as the last column of Table 6. It quantifies for each of

TABLE 7

*EX POST* INTERVALS OF VARIABLES OF EXTENDED NATIONAL ACCOUNTS SCHEME AS COMPARED WITH REDUCED INTERVALS THAT RESULT ALTERNATIVELY FROM DIRECT MEASUREMENT OF IMPORTS (*Mco, Mif*) AND PRODUCTION ACCOUNT TRANSACTIONS OF OTHER INDUSTRIES (*Xo, Co, Yo, So, Po*)

| | | Reduced intervals, based on direct measurement of: | |
|---|---|---|---|
| | Unreduced *Ex Post* Intervals | Imports | Production Account Transactions of Other Industries |
| *Sa* | 47.6 | 52.0 | 22.9 |
| *Pa* | 250.7 | 173.2 | 91.1 |
| *Xo* | 9.2 | 7.3 | — |
| *Co* | 9.2 | 7.3 | — |
| *Mco* | 9.0 | — | 4.1 |
| *Yo* | 9.2 | 7.3 | — |
| *So* | 22.9 | 22.9 | — |
| *Po* | 127.1 | 130.0 | — |
| *S* | 12.0 | 11.4 | 3.9 |
| *P'* | 26.2 | 24.9 | 10.7 |
| *P* | 55.9 | 51.1 | 24.0 |
| *Ch* | 12.9 | 11.4 | 3.8 |
| *If* | 8.7 | 8.6 | 3.8 |
| *Ifao* | 51.8 | 48.5 | 25.4 |
| *Mif* | 2.5 | — | 1.1 |
| *Y* | 3.9 | 3.0 | — |

the variables the lengths of the *ex post* intervals as a percentage of their central values. The remaining two columns include, as illustrations, the reductions of those intervals as a result of introducing additional direct estimates of variables into the compilation procedure. Column 3 reflects direct estimation of all import categories (including *Mif* and *Mco*) on the basis of further detailed analysis of foreign trade data. Column 3 shows how the intervals are reduced, when direct estimates are made of all production account transactions of other industries with help of more comprehensive survey or census information covering all or a large part of the establishments that belong to this sector. As none of the improvements could be implemented in practice with the data available, it was assumed in the table that direct measurement resulted in estimates that were equal to the central values earlier calculated as presented in Table 6. The direct estimates in Table 7 are those that have zero intervals (—) in columns 2 and 3. The zero intervals also apply to those variables that could be derived as residuals, once the direct information became available. The latter holds for GDP (*Y*) in column 3 of the table, which can be derived once value added of other industries (*Yo*) is measured directly.

The information presented in columns 2 and 3 shows that direct estimates of imports or production account transactions of other industries reduce the *ex post* intervals of a majority of the variables, including those variables that are not estimated directly. What was to be expected is that the reduction in the intervals depends on the type of variables estimated directly. For instance, the table shows that there is much more impact from direct estimates of production account transactions of other industries (column 3) than of direct estimates of import categories (column 3). In the latter instance intervals are generally reduced a few percentage points only, while intervals are reduced to half or even less when direct estimates are made of production account transactions of other industries. The difference in impact between the two columns may give an indication of how to improve national accounts estimates in the future. If one has to choose between further studies of foreign trade records to improve the import estimates, or improved surveys of industries to expand the coverage of establishments in the other industries sector, the latter course clearly is preferable. The effects may be examined similarly for direct estimates of compensation of employees through social security records, direct estimates of private final consumption expenditure through household surveys, or improvement of gross fixed capital formation estimates through investment surveys of enterprises.

## III. Evaluation and Further Development of the Modified Algorithm

This study is only a first step in the development of the lp algorithm for national accounts compilation. Much more work is needed to establish it more firmly from the conceptual point of view and make it operational beyond the application to the mere illustrative examples presented in section II. Some preliminary thoughts on some issues that need further elaboration, such as the theoretical justification for the central value concept, the practical and conceptual considerations in defining *ex ante* intervals for parameters and variables, and the required data processing facilities needed for the application of the algorithm in terms of the size of the linear programming matrix, are presented below. Finally some specific applications of the algorithm which might be considered in the field of national accounts as well as in other statistical systems are indicated.

The emphasis of the present study is on finding *ex post* intervals for variables and parameters which define the limits of the solution space. As no further information is available in the present model to arrive at optimal one point estimates, a central value vector was defined in the examples of section II as a simple unweighted average of all extreme solution vectors that were needed in order to arrive at the *ex post* minimum and maximum values. The advantage of this approach is that it is simple and easy to use. A more complex alternative is to apply the central value criteria defined in other recent similar studies [2, 3, 4] which are based on a modified version of the RAS method [5]. These approaches if used jointly with the present algorithm would minimize an objective function which might include the sum of the squared differences between the central values of all variables and parameters and their preliminary values before a reconciliation, weighted by the elements of the inverse of a covariance matrix. Such an approach, however, would complicate the use of the present algorithm because it would introduce non-linearities in the objective function and would require additional information on preliminary estimates in addition to the *ex ante* intervals that were used as the basis for the present algorithm. The incorporation of the covariance matrix furthermore seems to be superfluous in the present algorithm because the variances are already reflected in the restrictions of the present algorithm, while the covariances are difficult to measure and are generally assumed to be zero in the alternative approaches quoted.

The possibilities of establishing intervals for national accounts estimates [6, 7] have generally been limited for two reasons. The first is that a large part of the basic data used in national accounts estimation is a by-product of administrative records, for which it is difficult to assign reliability intervals on the basis of objective criteria. Only in instances where sample surveys are used as a source of statistical information can confidence intervals be determined with help of criteria developed in sampling theory. The second reason is a consequence of the present compilation procedures used in national accounts. Even if one could establish reliability intervals for the basic data, it would still be difficult to determine appropriate intervals for the ultimate national accounts estimates, given the very complex relation between these estimates and the basic data as a result of the adjustments applied to the latter. Through systematization of the national accounts compilation procedure based on the modified algorithm, the subjectivity of assigning reliability intervals can be reduced, however. The modified method accurately measures the ultimate *ex post* intervals of the variables in the national accounts framework, once the *ex ante* intervals of the basic data and the parameters of the analytical and adjustment relations have been defined. As this implies that the second obstacle mentioned above is no longer valid, one can concentrate on defining the reliability intervals of the basic data and particularly of the parameters of the analytical and adjustment constraints.

Three types of *ex ante* intervals could be used in the application of the algorithm:

(a) The first type are the intervals obtained from sample surveys. Probability functions and reliability intervals would thus be available for some of the national accounts variables and in some instances for parameters of the analytical relations.

(b) The second type of reliability intervals is based on judgemental criteria. Efforts have been made in some national accounting systems to assign subjective intervals to national accounts variables based on national accountants' experience [8, 9]. The proposed algorithm, however, opens the possibility not only to incorporate intervals for variables, but also for parameters in analytical functions, such as the ones used in the examples of section II. Such intervals may be established more easily, as national accountants continually make use of assumed ratios in the present compilation procedures and also make judgements on which ratio values are acceptable and which ones are not. Further development of the algorithm may require the identification of those parameters for which fairly short *ex ante* intervals can be set.

(c) Finally regression analysis may aid in defining intervals of parameters on the basis of a systematic analysis of relations between time series in the past. These intervals are generally indicated when analytical functions are measured. The proposed algorithm thus opens the possibility of incorporating the results of econometric analysis in the national accounts compilation procedure.

The size of the lp matrix is an important element in determining the data processing requirements and thus would influence the operationality of the lp algorithm. Experiments with the illustrative examples of section II have shown what the determining factors of the lp matrix size are. The number of columns of the matrix is equal to the number of unknown variables in the solution vector that are not equal to zero, plus one column for the solution vector itself. The number of variables that are not zero is equal to the total number of unknown variables minus the number of identities and analytical and adjustment constraints. To the number of unknown variables must be added the slack variables that are introduced in order to convert the analytical and adjustment inequalities to equalities required for the lp technique. In more precise terms, if the number of unknown variables (excluding slack variables) is $X$, the number of identitites is $I$, the number of analytical and adjustment constraints (pairs of inequalities) is $S$, then the number of columns of the lp matrix is equal to the number of unknowns including the slack variables $(X+2S)$ minus the number of identities and analytical constraints $(I+2S)$, plus 1 for the solution vector, or $(X-I+1)$. The number of rows of the matrix is equal to the number of national accounting identities and analytical constraints, or $(I+2S)$. Based on a number of 16 unknown variables represented in Table 3, and 4 identities and 13 adjustment and analytical relations shown in Table 4, the size of the lp matrix would thus be $13 \times 30$.

Further reductions of the size of the matrix, however, can be obtained in a number of ways. The number of unknown variables can be reduced by not including aggregates such as GDP and value added that can be defined as the sum of other unknown variables that can be included. This reduces not only the magnitude of $X$, but also the number of identities $I$ because if an aggregate is included in the compilation structure, it also needs to be defined in terms of other component items that are included as unknown variables. Elimination of these aggregates therefore reduces the number of rows of the matrix, not the number of columns. Such reduction can be considerable, as is apparent from the example presented in section IIB, where the total number of variables including aggregates was 16 (in Table 3) and the number of unknown variables included in the national accounts model is only 10 (in Table 4). Another reduction in the size of the lp matrix can be obtained by removing from $S$ all analytical and adjustment constraints that define one unknown national accounts category in terms of known basic data only. Such constraints, which can be reformulated as maxima and minima of the unknown variables, do not have to be incorporated in the matrix itself, but can be handled externally on the basis of the modified lp technique mentioned earlier (see Reference 1). This reduces $S$ and therefore the number of rows of the matrix further. In the example of section IIB the number of 13 analytical and adjustment constraints is thus reduced to only 7 in the final formulation of the lp algorithm. In order to improve the operationality of the algorithm, it is essential that additional reductions of the lp matrix be achieved. Further study is therefore directed into the structure of the matrix in order to reveal dependencies between columns which would permit elimination of some of the rows and columns of the matrix, to identify parts of the matrix that could be processed in an iterative manner or independent parts of the matrix that could be processed separately.

The algorithm was applied in section II to compile production accounts for a limited number of sectors and to show a breakdown of GDP by cost components and expenditure categories. Its flexibility, however, allows it to be applied to any statistical system that is characterized by consistency requirements between the different types of statistical information. For instance, it could be applied to reconcile institutional sector data between different institutional sectors, or to input-output compilation supplementary to the earlier mentioned RAS method in order to eliminate some of the unacceptable results of that method. Furthermore, it could be used to reconcile information of statistical subsystems of financial flows and balance sheets, income distribution, balance-of-payments statistics and government finance statistics with related information of the national accounts. Reconciliation between micro data and macro national accounts aggregates could be another field of application. The uses are not necessarily restricted to national accounts only, but could be also directed to internally consistent statistical systems that cover demographic and social information.

## References

[1] Hadley, G. *Linear Programming*, Addison-Wesley, Reading, Mass. 1963, pp. 387-395.

[2] van der Ploeg, F. Reliability and the Adjustment of Sequence of Large Economic Acounting Matrices, *Journal of the Royal Statistical Society*, Series A (General), Volume 145, Part 2, 1982, pp. 169-194.

[3] Barker, Terry, Frederick van der Ploeg and Martin Weale, A Balanced System of National Accounts for the United Kingdom, *The Review of Income and Wealth*, Series 30, Number 4, December 1984, pp. 461-485.

[4] Arkhipoff, Oleg, The Synthesis and the Reliability of National Accounts by Means of Electronic Data Processing: Reflections on an Experience, published in *Models and Decision Making in National Economics*, Proceedings of the 2nd International Conference on Dynamic Modeling and Control of National Economies, North-Holland Publishing Co., 1979.

[5] United Nations, *Input-Output Tables and Analysis*, Series F, No. 14, Rev. 1, 1973, pp. 65-80.

[6] Morgenstern, Oskar, *On the Accuracy of Economic Observations*, Princeton, 1970, pp. 242-276.

[7] Arkhipoff, Oleg, Le Paradigme de la Mesure et la Fiabilité de la Compatibilité Nationale, *Journal de la Societé de Statistique de Paris*, ler trimestre, 1984, pp. 25-41.

[8] Central Statistical Office, *National Accounts Statistics*, London 1968.

[9] Australian Bureau of Statistics, *The Accuracy and Reliability of the Quarterly Australian National Accounts*, Occasional Paper No. 1982/2.