IMF WORKING PAPER

WP/87/66

INTERNATIONAL MONETARY FUND

European Department

A New Method for Balancing the National Accounts

Prepared by Kasper Bartholdy*

Authorized for Distribution by Leslie Lipschitz

October 7, 1987

Abstract

The need for balancing the national accounts arises when the original estimates of individual items are inconsistent with the accounting identities. Three strategies for the balancing process dominate the literature: the RAS method, the Friedlander method, and the Generalized Least Squares (GLS) method. The latter is theoretically superior, but the algorithms hitherto developed for its implementation are cumbersome and tax even big computer systems. This paper presents a new, more efficient algorithm, based on the computational skeleton of the conceptually simple Friedlander method.

JEL Classification Number
221

*The author is indebted to F. van der Ploeg of the London School of Economics, Tim Folke of Danmark Statistik, and Eric Swanson of the World Bank for helpful comments on previous drafts of the paper. The usual disclaimer applies.

Contents

		Page
Summary		iii
1.	Introduction	1
2.	The general linear balancing problem	2
3.	A concrete balancing problem	2
4.	The RAS method	4
5.	The Friedlander method	5
6.	Balancing by means of the Generalized Least Squares method	7
7.	The Generalized Friedlander method	9
8.	A practical application	11
9	Efficiency	13
10.	Concluding remarks	13

Summary

Initial estimates of national accounts are usually derived from test samples and may often conflict with basic accounting identities because of random sample errors. A new, consistent and more reliable set of data can be computed on the basis of the original data and the information represented by the identities. The ideal balancing process lets the least reliable of the original estimates bear the brunt of the necessary adjustment, and there is agreement in the literature that the most appropriate framework for achieving this goal is the Generalized Least Squares (GLS) approach.

There are, however, few practical applications of this method as procedures are cumbersome and can tax even large computer systems. This paper derives a new, conceptually simple, algorithm and presents results that indicate that this is computationally more efficient than the GLS algorithms available to date. Because the new approach is much easier to use and can be applied with even relatively small computers, it should provide for substantially wider practical use of GLS-balancing in the future.

1. Introduction

In most industrialized countries, each item of the national accounts is estimated on the basis of several different sources. For example, separately collected data for production, demand and income can be used to calculate three statistically independent estimates of the gross domestic product. Usually such estimates are derived from test samples and it is highly unlikely that independent estimates of the same item have identical values. However, the differing estimates can be used as a basis for the calculation of a new reconciled estimate, which is more reliable than each of the original ones. This process of reconciling conflicting estimates is called "balancing the accounts." In general, the need to balance the national accounts data arises, whenever some of the original estimates conflict with the relevant set of accounting identities. The balancing process eliminates such conflicts and leads to a consistent set of estimates for all items.

Since the second World War, the systems of national accounts in industrialized countries have become increasingly sophisticated, thus the search for ways of systematizing and automizing the balancing process have led to numerous suggestions of methodological approaches. The ones most frequently cited are the RAS method, the Friedlander method, and the Generalized Least Squares (GLS) method. The GLS estimator is the best (that is, the most reliable) 1/ linear unbiased estimator. It makes optimal use of available information about the reliability of the original, inconsistent estimates and is therefore theoretically superior to the RAS and the Friedlander method. The case for GLS balancing was first presented by Stone, Champernowne, and Meade in 1942. For a long time technical problems hindered practical application of the method but these obstacles have been reduced by recent developments in numerical analysis (Byron, 1978; van der Ploeg, 1982; Barker et.al., 1984; and Stone, 1984).

This paper presents an algorithm which provides for yet another step in the direction of facilitating practical use of GLS balancing. The following sections show how a slight generalization of the Friedlander algorithm renders it possible to use that algorithm for GLS balancing. The "Generalized Friedlander" algorithm is conceptually simple and more amenable to programming than previously used algorithms, such as the one suggested by Byron (1978). 2/

Byron's suggestion partly overcame the problems of computer storage. Thus, a computer program, which solves a particular set of balancing problems, along the lines suggested by Byron, has been

---

1/ Here, reliability is defined as the reciprocal value of the sum of the variances of the balanced estimates for the items in the national accounts.

2/ On the other hand, Byron's algorithm is applicable to a much wider scope of problems than the Generalized Friedlander method.

developed by F. van der Ploeg and applied to systems of up to 262 accounts (see van der Ploeg, 1987). However, datamatic and conceptual complexity still constitute major barriers for practical application of GLS for balancing very large national accounts systems. Thus, the Generalized Friedlander process, which is presented below, provides an important expedient for the proliferation of GLS balancing.

2. The general linear balancing problem

The framework for the presentation below is a specific balancing problem, to which the RAS method, the Friedlander method as well as the Generalized Friedlander method apply. Before outlining this specific case, it is appropriate to define what is generally meant by "a balancing problem."

Let the "true" values of the items in the national accounts comprise the column vector y, and assume that a first estimate, y*, has been computed on the basis of available primary data. The two vectors, y and y*, may differ due to stochastic errors, contained in the vector e:

(1) $y^* = y + e$

The "true" values comply with a set of national accounts identities:

(2) $f_i(y) = h_i$, $i = 1, \ldots, z$,

where, h_i is a scalar.

When f_i is linear for all values of i, the balancing problem is said to be linear. In this case (2) can be rewritten as follows:

(2a) $Cy = h$,

where C is a matrix of constant scalars.

If $f_i(y^*) \neq h_i$ for some value of i, the elements of y* are said to be "inconsistent." The thrust of the balancing process is to calculate a revised estimate, y**, for which

(3) $f_i(y^{**}) = h_i$, $i = 1, \ldots, z$

y** is called "a balanced estimate".

3. A concrete balancing problem

The general balancing problem outlined above can be given a more concrete interpretation on the basis of the following production accounts:

	Supply		Demand					
	Production	Imports	Production inputs	Private consumption	Public consumption	Fixed investment	Stock building	Exports
Commodity 1 Commodity 2 . . . Commodity n	V	m			U			x
Total	i'V	i'm			f'			i'x

Here, V shows the distribution of domestic production by industrial branch while m contains the import figures for each commodity. U specifies the use of commodities in production and for purposes of consumption and investment, while x records the exports of each commodity. The vector f shows the column totals for each category of intermediate and final domestic demand and i is a vector of appropriate dimension in which all elements take on the value 1.

Assume now (realistically) that highly reliable estimates of V, m and x are available while initial estimates of U and f have been computed on the basis of relatively weak information (U and f are assumed to have been estimated independently). A proper approach to the resultant balancing problem would be to adapt the initial estimates of U and f to the much more reliable estimates of V, m and x; that is to find U** and f**, for which

(4) $i'U^{**} - (f^{**})' = 0'$

(5) $U^{**}i - (V^{**}i + m^{**} - x^{**}) = 0$

(6) $V^{**} = V^{*}, \; m^{**} = m^{*}, \; x^{**} = x^{*}.$

In order to simplify the specification of the balancing problem, U* and f* can be concatenated into a matrix A*:

$$(7) \quad A^* = \begin{bmatrix} U^* \\ -(f^*)' \end{bmatrix}$$

Equations (4), (5), and (6) imply that the balancing process should adapt A* to the known sums of the rows and columns of the "true" matrix, A:

$$(8) \quad A^{**}i = \begin{bmatrix} V^{**}i + m^{**} - x^{**} \\ - i'[V^{**}i + m^{**} - x^{**}] \end{bmatrix}$$

$$(9) \quad i' A^{**} = 0'$$

Any A** which satisfies (8) and (9) is a balanced estimate of A. Both the RAS method, the Friedlander method and the Generalized Friedlander method can be used to compute a balanced estimate.

Equations (8) and (9) are derived solely on the basis of the production accounts. This may appear to conflict with basic ideas in the guidelines laid down in the UN System of National Accounts (1968), which provide for a consistent set of tables covering production, income-outlay, accumulation and financial accounts. Ideally, the whole set of tables should be balanced simultaneously so as to apply all available information to the adjustment of each item. (A practical example of such an approach can be found in Barker et. al., 1984.) However, in many countries it is extremely difficult to compile data for nonwage income, which is, therefore, derived residually in the full accounting system. Under such circumstances, the production accounts can be balanced separately, in line with the set-up presented above, without loss of information.

4. The RAS method

The RAS method can be used whenever an initial estimate A* of a matrix A has been generated, and the "true" row and column totals are known. 1/ The method is based on a model for the elements of the matrix A. The model contains a descriptive and a definitional part. First it is assumed that the "true" value of each individual element, a_{ij} is a product of the initial estimate, a^*_{ij}, a row factor, r_i, and a column factor, s_j.

$$(10) \quad a_{ij} = r_i a^*_{ij} s_j + e_{ij} \quad (i = 1, \ldots, m;\ j = 1, \ldots, n)$$

Here e_{ij} denotes a stochastic variable whose mean is zero. According to (10) a set of unbiased balanced estimates can be calculated as follows:

1/ The method was originally proposed in Deming and Stephan (1940), but the mathematical characteristics of the method were only explored much later (Gorman, 1963, and Bacharach, 1970).

$$(11) \quad a_{ij}^{**} = r_i a_{ij}^{*} s_j \quad (i = 1, \ldots, m;\ j = 1, \ldots, n)$$

Equation 11 constitutes the descriptive part of the RAS model. The identities of the RAS model express that the row and column totals of the balanced matrix-estimate must equal the exogenously given values. Let the vectors u and v contain these values. Then

$$(12) \quad \sum_{j=1}^{n} a_{ij}^{**} = u_i \quad (i = 1, \ldots, m;\ j=1, \ldots, n)$$

$$(13) \quad \sum_{i=1}^{m} a_{ij}^{**} = v_j \quad (i=1, \ldots, m;\ j=1, \ldots, n)$$

The standard method for solving the system (11), (12) and (13) with respect to A** is iterative. During each iteration two estimates of the matrix A and one estimate of r and s, respectively, are computed. In the p'th iteration, the formulas applied are the following:

$$(14) \quad a_{ij}^{2p-1} = \left(1 + \frac{u_i - \sum_j a_{ij}^{2p-2}}{\sum_j a_{ij}^{2p-2}}\right) \cdot a_{ij}^{2p-2} = r_i^{p} \cdot a_{ij}^{2p-2}$$

$$(15) \quad a_{ij}^{2p} = a_{ij}^{2p-1} \left(1 + \frac{v_j - \sum_i a_{ij}^{2p-1}}{\sum_i a_{ij}^{2p-1}}\right) = a_{ij}^{2p-1} \cdot s_j^{p}$$

a_{ij}^{2p} is the 2p'th approximation to a_{ij}^{**}. In the first iteration, p equals 1 and a_{ij}^{o} equals a_{ij}^{*}.

In (14) the difference between the i'th row sum (in the most recently calculated A-estimate) and the value u_i is distributed proportionately to the elements of that row. Similarly (15) distributes the difference between the current j'th column sum (in the most recently computed A-estimate) and the value v_j proportionately to the elements of that column. Under normal circumstances the iterative process described by (14) and (15) will converge toward a unique balanced estimate of A. 1/ This balanced estimate is the solution A** to (11), (12) and (13).

1/ In some cases it is possible to break down the RAS model into several independent sub-models. If for each of these subproblems the sum of the known row totals equals the sum of the known column totals ($\Sigma u_i = \Sigma v_j$) then a unique solution to the RAS model, (11)-(13) exists and the process described by (14) and (15) will converge toward that solution (see Gorman, 1963 and Bacharach, 1970).

By comparing (12) and (13) it is seen that the specific balancing problem given by (8) and (9) can be solved by means of the RAS method. The primary advantage of choosing the RAS approach is its great computational and conceptual simplicity. The price for this simplicity is non-optimal use of available information on the relative reliability of the individual estimates in A*. The j'th column error is distributed proportionately on the elements of the j'th column even if the variance of the estimator for one element is significantly larger than the variance of the estimator for another. Thereby, the RAS method violates the intuitively reasonable requirement of any balancing process that the most unreliable estimates should bear the lion's share of the balancing burden.

5. The Friedlander method

The Friedlander-method is very similar to the RAS-method. 1/ The precondition for using the Friedlander method is the existence of an initial estimate A* of some "true", non-negative matrix A and perfect knowledge of the "true" row and column totals. The model on which the Friedlander method is based can be formulated as follows:

$$(16) \quad a_{ij}^{**} = a_{ij}^{*}(1 + r_i + s_j) \qquad (i = 1, \ldots, m;\ j = 1, \ldots, n)$$

$$(17) \quad \sum_j a_{ij}^{**} = u_i \qquad (i = 1, \ldots, m;\ j = 1, \ldots, n)$$

$$(18) \quad \sum_i a_{ij}^{**} = v_j \qquad (i = 1, \ldots, m;\ j = 1, \ldots, n)$$

As in the formulation of the RAS model above, r_i and s_j are row and column factors, respectively, while u_i and v_j are the "known" values of i'th row sum and the j'th column sum, respectively.

The Friedlander model can be solved by using the following iterative procedure (corresponding to equation (14) and (15) of the RAS process):

$$(19) \quad a_{ij}^{2p-1} = a_{ij}^{2p-2} + \left(\frac{u_i - \sum_j a_{ij}^{2p-2}}{\sum_j a_{ij}^{*}}\right) \cdot a_{ij}^{*} = a_{ij}^{2p-2} + r^{p} \cdot a_{ij}^{*}$$

$$(20) \quad a_{ij}^{2p} = a_{ij}^{2p-1} + a_{ij}^{*} \cdot \left(\frac{v_j - \sum_i a_{ij}^{2p-1}}{\sum_i a_{ij}^{*}}\right) = a_{ij}^{2p-1} + a_{ij}^{*} \cdot s_j^{p}$$

This process keeps the coefficients used for distribution of the discrepancies constant. The discrepancy in each row is distributed on

1/ Like the RAS method, the Friedlander method was originally presented in Deming and Stephan (1940) but the mathematical properties of the method were first analyzed in Friedlander (1961).

the individual element a_{ij}^{2p-1} in proportion to the size of the initial estimate a_{ij}^{*}. The same principle applies to the distribution of column differences.

If the process converges (that is, if all the row and column discrepancies approach zero for $p \to \infty$) then the estimate A^p for $p \to \infty$ is a balanced estimate.

It can be proven that the Friedlander balanced estimate solves the following minimization problem:

$$(21) \quad \mathrm{Min}_{(i,j)\varepsilon\phi} \Sigma \frac{(a_{ij}^{*} - a_{ij}^{**})^2}{a_{ij}^{*}} \quad \text{subject to } \Sigma_j a_{ij}^{**} = u_i \text{ for all } i, \quad \Sigma_i a_{ij}^{**} = v_j \text{ for all } j.$$

Here $\phi = \{ (i,j) \mid a_{ij}^{*} > 0 \}$.

The Friedlander process will always converge toward the solution to this problem if such a solution exists.

Like the RAS method, the Friedlander method yields conceptual and computational simplicity. But it also contains the same inherent weakness; namely that the column and row discrepancies are distributed automatically without regard to the relative reliability of the initial estimates.

The GLS method, which is presented below, overcomes this weakness and can be combined with the Friedlander iteration scheme to avoid loss of computational simplicity.

6. Balancing by means of the generalized least squares method

The GLS solution to the general linear balancing problem is based on the following equation:

$$(22) \quad y^* = y + e, \; E(ee') = W$$

Here y^* is the first estimate of the "true" national accounts vector y, and W is the covariance matrix for the residuals, contained in the vector e. Since W is positive definite (as are all nonsingular covariance matrices), it can be expressed in the form PP', where P is a nonsingular matrix:

$$(23) \quad W = PP'$$

This leads to:

$$(24)\quad (P^{-1})'\, P^{-1} = W^{-1} \qquad \text{and}$$

$$(25)\quad P^{-1}W(P^{-1})' = I,$$

where I is a unit matrix.

Premultiplying (22) by P^{-1} gives

$$(26)\quad P^{-1}y^{*} = P^{-1}y + P^{-1}e$$

These steps convert (22) into an ordinary least squares model, since the covariance matrix of the vector $P^{-1}e$ is equal to the unit matrix:

$$(27)\quad E[P^{-1}e\ (P^{-1}e)'] = P^{-1}W(P^{-1}) = I$$

The principle of GLS is that the estimate, y**, should minimize the sum of the squared residuals in (26). Thus, the balanced estimate solves the following problem:

$$(28)\quad \text{Min } (P^{-1}y^{*} - P^{-1}y^{**})'\ (P^{-1}y^{*} - P^{-1}y^{**})$$

$$\text{s.t.}\quad Cy^{**} = h,$$

where the constraints express the national accounts identities (which are assumed to be linear). Using (24), (28) can be simplified:

$$(29)\quad \text{Min } (y^{*} - y^{**})'\ W^{-1}\ (y^{*} - y^{**})$$

$$\text{s.t.}\quad Cy^{**} = h$$

The first order conditions for a minimum are the following:

$$(30)\quad W^{-1}(y^{*} - y^{**}) - C'\alpha = 0 \quad \text{and}$$

$$(31)\quad Cy^{**} = h$$

Here α is a vector of Lagrange multipliers. The solution to the equation system, (30) and (31), is:

$$(32)\quad \alpha = -(CWC')^{-1}\ (Cy^{*}-h) \text{ and}$$

$$(33)\quad y^{**} = y^{*} + WC'\alpha$$

The vector y^{**} in (33) is the balanced BLUE estimate, if the initial estimate, y^{*}, is unbiased.

In practice the values of the elements in the covariance matrix W are rarely known. Moreover, it is usually difficult to ensure that all estimates in y^{*} are unbiased. However, given the knowledge about biases and relative reliabilities, an "assumed" BLUE estimate (i.e., an estimate which uses the statisticians' knowledge optimally) can be calculated by

(i) eliminating "assumed" biases from y^* before inserting it into (32) and (33) and

(ii) using the statisticians' "guessed" covariance matrix for W in these formulas. 1/

However, the inversion of (CWC') may often not be feasible in practice because of the enormous dimensions of that matrix. For instance, the A matrix in the Danish national accounts would contain 4,000 rows and 200 columns. The constraint matrix C generally contains one row for each column or row in A. Thus, when balancing the Danish national accounts the dimension of (CWC') would be 4,200 x 4,200. Frequent inversion of such matrices is rarely possible on computers, accessible to a country's national accounts department.

In his 1978 article, Byron suggests a possible way of escaping this problem. The computation of y^{**} can be divided into two phases; one in which only α is calculated (via equation (32)) and one in which the value of α is inserted into (33) to give the balanced estimate of y. The problematic inversion only concerns phase 1, and can be totally evaded by use of the so-called conjugate gradient algorithm. By using this algorithm, computation or storage of the inverted matrix $(CWC')^{-1}$ can be avoided. Only the non-inverted version, (CWC'), needs to be stored in the memory of the computer, and that can be done compactly on a data machine (since under normal circumstances the non-inverted version is dominated by zeros).

However, even in the light of Byron's suggestions, the problems of storage optimization and programming remain important barriers for practical widespread use of GLS for balancing large matrices. When using the conjugate gradient algorithm, the statisticians and their assistants still have to do some difficult programming, and still run into computer capacity problems in connection with large balancing exercises. Therefore it remains worthwhile to look for alternative algorithms which may simplify the GLS computations. The following section shows how the Friedlander method can be used for that purpose.

7. The generalized Friedlander method

As mentioned above, the Friedlander process solves balancing problems which are characterized by perfect knowledge of the row and column totals for a matrix A, given that (i) a first estimate, A^*, has

---

1/ Note that only relative values--not absolute values--of the elements in W matter for the solution to (33). This property can be utilized in the case in which no initial estimate for a particular element is available. An arbitrary value for the initial estimate can be chosen, and the variance given a very large value.

been computed and (ii) a solution to the balancing problem exists. The minimization problem, solved by the Friedlander process may be written as follows:

$$(34)\quad \text{Min} \sum_{(i,j)\varepsilon\phi} \frac{(a^{*}_{ij} - a^{**}_{ij})^{2}}{a^{*}_{ij}} \qquad \text{s.t.} \quad \sum_{j} a^{**}_{ij} = u_{i} \text{ for all } i, \quad \sum_{i} a^{**}_{ij} = v_{j} \text{ for all } j,$$

where $\phi = \{ (i,j) \mid a^{*}_{ij} > 0 \}$.

By comparing (34) to (29) it is easily seen that the solution, A**, to (34) is actually the GLS estimate in the special case for which $\text{var}(a^{*}_{ij}) = \zeta a^{*}_{ij}$ and $\text{cov}(a^{*}_{ij}, a^{*}_{kl}) = 0$ for all i, j, k and l (ζ is a scalar constant).

With this in mind, it seems straight-forward to try to use the computationally simple Friedlander algorithm for GLS balancing in more general cases. In most practical applications of GLS balancing it is natural to assume that $\text{cov}(a^{*}_{ij}, a^{*}_{kl}) = 0$ for all values of i,j, k, and l (this assumption is often unavoidable because of lack of knowledge about covariances). But it is rarely in accordance with the knowledge of the statisticians involved to assume, as in the Friedlander case above, that the variance of the stochastic error of the individual estimate, a^{*}_{ij}, is proportional to the initial estimate, $(\text{var}(a^{*}_{ij}) = \zeta a^{*}_{ij})$.

Thus it is desirable to find an algorithm which is computationally similar to the Friedlander method, but makes it possible to solve the following GLS minimization problem for any value of the variances, w_{ij}:

$$(35)\quad \text{Min} \sum_{(i,j)\varepsilon\phi} \frac{(a^{*}_{ij} - a^{**}_{ij})^{2}}{w_{ij}} \qquad \text{s.t.} \quad \sum_{j} a^{**}_{ij} = u_{i}, \quad \sum_{i} a^{**}_{ij} = v_{j}$$

Here $\phi = \{ (i,j) \mid w_{ij} > 0 \}$

It turns out that only a slight generalization of the Friedlander process is needed. The necessary generalization is reflected in the following formulas for the calculations of the p'th iteration:

$$(36)\quad a^{2p-1}_{ij} = a^{2p-2}_{ij} + \frac{w_{ij}}{\sum_{j} w_{ij}} \left(u_{i} - \sum_{j} a^{2p-2}_{ij}\right)$$

$$(37) \quad a_{ij}^{2p} = a_{ij}^{2p-1} + \frac{w_{ij}}{\sum_i w_{ij}} (v_j - \sum_i a_{ij}^{2p-1})$$

This "Generalized Friedlander process" is started by setting p equal to 1 and $A^{\circ} = A^*$. Thereafter, (36) and (37) are solved in turn during the p'th iteration. It is obvious that if the process converges (that is if the row and column differences tend to zero for $p \to \infty$) then $\lim_{p\to\infty} A^{2p}$ is a balanced estimate of A. The resultant matrix is the GLS balanced estimate for the case in which there is no covariance between errors on initial estimates. Whenever a solution to the minimization problem (35) exist, the generalized Friedlander process will tend toward that solution. 1/

In sum, if (i) the problem of balancing a national accounts matrix does not contain other restrictions on the balanced estimate than those dictated by predefined values of row and column sums, and (ii) the covariance matrix of stochastic estimation errors is diagonal, then the GLS estimate can be computed in a simple and efficient manner by means of the generalized Friedlander method, defined by equations (36) and (37) above.

8. A practical application

The acquisition of data for a detailed set of national accounts is a time consuming process. Therefore, many countries set up preliminary accounts which may be published far earlier than the final accounts. An example of that practice can be found in the Danish Bureau of Statistics. One of the balancing problems which faces Danish statisticians in connection with the preliminary national accounts corresponds to the balancing problem presented above in equations (8) and (9) and depicted in the table below. 2/

Initial estimates for the supply-matrix, V, the import vector, m, and the export-vector, x, are far more reliable than estimates for domestic absorption (the matrix U and the vector f). Therefore, the Danish Bureau of Statistics has decided to let the balancing process adapt the estimates of U and f to the initial estimates for V, m and x. The procedure currently used in practice is semi-automatic. The value zero is used as an initial estimate for all types of stockbuilding. As a first balancing step, the item "total stockbuilding" (which constitutes the last element in f*) is set equal to the difference between the estimates for total supply and total final demand, respectively. Thereafter, the order of magnitude of the resultant stockbuilding estimate is evaluated on a discretionary basis. This evaluation may cause some manual readjustments of elements in the f* vector. The resultant f estimate is considered final. Now,

---

1/ This can be shown via a slight modification of the appendix in Friedlander (1961).

2/ In the Danish national accounts, the non-wage income is calculated as a residual item, as described in section 3. Therefore, the production accounts are balanced separately.

the U estimate goes through a RAS iteration (one distribution of row errors and one distribution of column errors). Subsequently, all remaining row differentials are attributed to the stockbuilding column of the U estimate, whereby a balanced estimate, U**, is formed.

The Danish Bureau of Statistics is considering changing this process and instead use the principles of GLS balancing. In cooperation with the Bureau, the author carried out a GLS balancing exercise, applying the "Generalized Friedlander" algorithm to an aggregated version of the unbalanced preliminary national accounts for 1982. The level of aggregation in the exercise is shown in the table below.

Aggregated Version of the Preliminary
Danish National Accounts for 1982

	68 industries	66 types of private consumption	1 type of public consumption	4 types of fixed investment	1 type of stockbuilding	Totals
157 commodity groups 1/			U*			V**i+m**-x**
Totals			(f*)'			i'(V**i+m**-x**)

1/ In the preliminary accounts used by the Danish Bureau of Statistics, 1,600 commodity groups are involved. The final accounts comprise almost 4,000 commodity groups.

The variances of stochastic errors associated with U* and f* were calculated on the basis of 95 percent confidence-intervals, specified by specialists in the Danish Bureau of Statistics. The most reliable initial estimates in the U* matrix were deemed to be the ones related to input usage in industries (the error margin to the corresponding individual f* element was set at 2 percent of the size of the element). Fixed investment and stockbuilding data were considered the least reliable by far (the f* margins for fixed investment were set at 20 percent of the size of the elements, while a guessed 95 percent confidence interval for stockbuilding was computed as ±5 percent of total supply). 1/ It was decided that the iterative generalized Friedlander computations should end, when the largest row or column difference was smaller than DKr 100,000 (less than a ten millionth of

1/ The procedure as well as the results are documented in detail in Bartholdy (1983).

total supply). With this stop criterion, 13 iterations were required to reach an approximately balanced solution. As shown below, this result bodes well for the efficiency of the generalized Friedlander method.

9. Efficiency

In Van Der Ploeg (1982), it is shown that for big social account mxn- matrices, the conjugate gradient algorithm requires approximately $2(m^2n + mn^2)$ elementary operations (multiplications or divisions). The Generalized Friedlander method initially requires 2mn divisions to compute the weight matrices for error distribution. Thereafter, 2mn elementary operations are carried out in each iteration, which implies that (for large values of m and n) the generalized Friedlander algorithm is cheaper than the conjugate gradient algorithm when the number of iterations required is smaller than (m+n). In the balancing exercise on the Danish National Accounts above, approximately 2*13*158*140 (=575120) multiplications were carried out. The conjugate gradient algorithm would have required $2*[158^2*140+158*140^2]$ (=13183520) elementary operations (almost 23 times as many). This gives a clear indication of the generalized Friedlander method's efficiency.

10. Concluding remarks

This paper has presented a new method for computing a GLS balanced estimate of a matrix, whose elements have to match a given set of row and column totals. The new method, the "Generalized Friedlander" method, has one major weakness compared to the conjugate gradient algorithm: it cannot handle covariances between errors of original estimates. Thus, if for instance the total of two items in the interior of the matrix is known more precisely than the two items themselves, the possibility of using the Generalized Friedlander method may have to be dismissed. However, in practice, the "best guess" covariance matrix is often diagonal. When that is the case, the Generalized Friedlander seems preferable to the conjugate gradient algorithm on account of conceptual simplicity and computational efficiency.

Bibliography

Bacharach, M., Biproportional Matrices and Input-Output Change, Cambridge University press, 1970.

Barker, T., F. Van Der Ploeg, and M. Weale, "A Balanced System of National Accounts for the United Kingdom," Rev. of Income and Wealth, 1984, pp. 461-485.

Bartholdy, K., Metoder til Afstemning af Nationalregnskabsmatricer, University of Copenhagen, 1983.

Beckman, F.S., "The Solution of Linear Equations by the Conjugate Gradient Method," Mathematical Methods for Digital Computers, (*Editors: Ralston and Wolf), Wiley, New York, 1962.*

Byron, R., "The Estimation of Large Social Account Matrices," J.R. Stat. Soc. A, 1978, pp. 359-367.

Deming, V. and F. Stephan, "On A Least Squares Adjustment of a Sampled Frequency Table when the Expected Marginal Totals are Known," Annals of Mathematical Statistics, 1940, pp. 427-444.

Friedlander, D., "A Technique for Estimating a Contingency Table, Given the Marginal Totals and Some Supplementary Data," J.R. Stat. Soc. A, 1961, pp. 412-420.

Gorman, W.M., Estimating Trends in Leontieff Matrices, 1963, duplicated note, Oxford University.

Stone, R., "Balancing the National Accounts. The Adjustment of Initial Estimates: A Neglected Stage in Measurement," in Demand, Equilibrium and Trade: Essays in Honour of Ivan Pearce, (Editors: Ingham, A. and Ulph, A), MacMillan, London, 1984.

_____, D. Champernowne, and J. Meade, "The Precision of National Income Estimates," Rev. of Ec. Studies, 1942, pp. 111-125.

United Nations, A System of National Accounts, Studies in Methods, Ser. F., No. 2, Rev. 3, New York, 1968.

van der Ploeg, F., "Reliability and the Adjustment of Sequences of Large Economic Accounting Matrices," J.R. Stat. Soc. A, 1982, pp. 169-194.

_____, "Econometrics and Inconsistencies in the National Accounts," Economic Modelling, 1985, pp. 8-16.

_____, Balancing Large Systems of National Accounts, unpublished article, London School of Economics, 1987.