Review of Income and Wealth
Series 46, Number 3, September 2000

NATIONAL ACCOUNTS ESTIMATION USING INDICATOR RATIOS

BY JAN R. MAGNUS

CentER, Tilburg University

JAN W. VAN TONGEREN

United Nations Statistics Division

AND

AART F. DE VOS

Free University, Amsterdam

We propose a new approach to national accounts compilation, which also serves as a formalization of current compilation practices. When formalizing the procedure, a distinction is made between (basic) data, national accounts identities and so-called indicator ratios. The latter are ratios of or percentage relations between national accounts variables, such as the relation between output and value added. Indicator ratios are currently used in national accounts compilation practices in order to make adjustments to the basic data or to fill in missing data. The latter use is particularly relevant when basic data are scarce, which is the case not only in many developing countries, but also in developed countries when annual accounts are compiled for recent periods. The (basic) data, indicator ratios and identities together are used in a Bayesian approach to estimate the values of national accounts variables and analytical indicator ratios based thereon. The amendment of the current practices consists in introducing reliability intervals of basic data and indicator ratios, which allows for the use of a much larger number of indicator ratios in the compilation and checking of national accounts data. The Bayesian compilation approach makes it possible—in contrast to current practices—to use indicator ratios both as priors and as analytical indicators.

1. INTRODUCTION

This paper proposes a new and relatively simple Bayesian approach to national accounts compilation. The new approach can (and will) be compared with current compilation practices, which are generally less sophisticated and do not provide reliability intervals of the estimates. The proposed approach takes account of the identities and the indicator ratios used in the national accounts compilation, and establishes a direct link to an important usage of national accounts figures in analysis, namely the use of indicator ratios. It also provides a framework for evaluating the usefulness of extra or better data, for example in estimating and analyzing indicator ratios.

Current national accounts compilation practices use basic data, identities (reflecting internal consistency criteria that the national accounts estimates should satisfy), and structural coefficients (or indicator ratios). The latter are ratios of

Note: We are grateful to Sanne de Boer and Bernhard Conlon for help with the original programming in Gauss, to Anton Markink for professional help in preparing robust computer packages, and to the participants of the "Expert Group Meeting on the Use of Macro Accounts in Policy Analysis," 5–9 October 1998, at the United Nations in New York, and two referees for helpful and constructive comments.

or percentage relations between national accounts variables, such as the relation between output and value added, between value added of large and small establishments, or between product flows and trade-and-transport margins. The indicator ratios are currently used in order to make conceptual adjustments to the basic data, or adjustments for coverage of basic data, to derive missing data, and to perform consistency checks on the data. Indicator ratios are particularly relevant in the compilation of national accounts when basic data are scarce. This is the case in many developing countries, but also in developed countries when annual accounts are compiled for recent periods or quarterly accounts are compiled.

The current paper provides a formalization of current practices, in line with recent UNSD attempts (United Nations, 1999) which support the increasing computerization of national accounts compilation procedures in current practices and link those practices to methodologies of measurement used in analyses based on the national accounts estimates, in particular those used in econometrics.

The discussion in the paper focuses on the national accounts as the medium through which the interaction between indicator ratios used in compilation and analysis is studied. National accounting in this context should be interpreted in a very broad sense, i.e. an approach to integrate and make consistent basic data through the use of an accounting framework with a consistent set of concepts and classifications. This is not the limited concept of national accounts reflected in a large number of country practices, which focuses on the estimation of GDP only. The accounting framework referred to here is that of the international guidelines of the 1993 System of National Accounts (hereafter SNA) described in United Nations *et al.* (1993), which emphasize—in addition to GDP—the compilation of so-called institutional sector accounts and also introduces the possibility of using non-economic data in the compilation of satellite accounts. However, since the example in Section 5 uses only economic data, national accounts in the present context refers to national economic accounts only.

Indicators are widely used by international organizations (World Bank, 1993; United Nations Development Program, 1996) to set policy goals and monitor development in countries and regions. Indicators in this broad sense include basic data, estimates that are reconciled through national accounting, and also ratios between variables. The present paper deals only with indicator ratios. The reason for this limitation is twofold. First, indicator ratios, such as per capita GDP or investments as a percentage of GDP are more relevant for international and intertemporal comparisons than the underlying variables of GDP, investments and population size. Indicator ratios can thus be considered as an important analytical summary of a large national accounting data set. Secondly, indicator ratios are also used in the compilation of national accounts and it is the interaction between these two uses which is the main focus of this paper.

The paper also contributes to an old and still very relevant discussion of reliability of basic data, for which Morgenstern (1963, first published in 1950) laid the foundations. Morgenstern approached national accounting as a branch of descriptive statistics. This is different from the approach taken in this paper, where national accounting is used as the medium for integration of statistics. The approach initiated by Morgenstern was followed by at least three countries which publish their national accounts estimates and related statistics with indications of their reliability: the U.K., Australia and more recently Canada. The reliability of national accounts can also be based on revisions made over time, an approach already suggested by Morgenstern.

A second approach was initiated by Stone, Champernowne and Meade (1942). This involves both the reliability of the data and the problem of balancing the accounts, see Byron (1978); Barker, Van der Ploeg and Weale (1984); and Van der Ploeg (1982, 1984, 1985).

Another approach was followed by van Tongeren (1985), who used linear programming techniques not only to balance the accounts, but also to determine the relation between the "prior" reliability intervals of basic data and the "posterior" intervals, once those data are reconciled within an accounting framework.

The present paper builds on these approaches. It is written both for econometricians and other theoretical statisticians, familiar with estimation procedures used in econometrics and related fields, as well as for national accountants, familiar with the estimation methods used in national accounting. The integration of the two approaches might be beneficial to both disciplines, provided national accountants pay more attention to the analyses in which the data are used, and econometricians familiarize themselves more with the intricacies of basic data. National accounting, which was initially designed for analytical purposes, but was later developed by statisticians, may be the ideal framework for such an interdisciplinary approach.

The plan of the paper is as follows. In Section 2 we state our main theoretical result (proved in an Appendix), which gives us the means of combining incomplete data with incomplete prior information (including exact accounting identities). This theorem is based on normality and linearity. In Section 3 we discuss how to use indicator ratios—which are nonlinear—as priors. This leads to an extension of Theorem 1. In Section 4 we discuss how to deal with multiple priors, that is the situation where many priors or several priors on the same variable are available. Section 5 contains a simple but realistic example of the application of the proposed approach, and Section 6 concludes. The proof of our main theorem is provided in the Appendix.

2. The Adjustment of Unreliable Observations

Consider a vector x of n latent variables, to be regarded as a vectorization of a system of accounts. Data are available on $p \leq n$ components (or linear combinations) of x. Let d denote the $p \times 1$ data vector. Our starting point is a *measurement equation*, $d|x \sim N_p(Dx, \Sigma)$. Typically, the $p \times n$ matrix D is a selection matrix, say $D = (I_p, 0)$, so that Dx is a subvector of x. Measurements are unbiased in the sense that $E(d|x) = Dx$. The $p \times p$ matrix Σ denotes a positive definite variance matrix, typically (but not necessarily) diagonal.

In addition to the p data, we have access to two further pieces of information: deterministic accounting constraints and prior (possibly multiple) views concerning the latent variables or linear combinations thereof. These two pieces are combined into one set of linear *priors*, $Ax \sim N_m(h, H)$, where the variance matrix H is singular, because each of the deterministic constraints has variance 0.

We now have data and priors, and we wish to employ Bayes' theorem to combine them and obtain posteriors. We emphasize two complicating features of this problem: we do not, in general, have data on all latent variables ($p < n$), nor do we have priors on all latent variables ($m < n$). If we had data on all latent variables ($p = n$), then the results in Van der Ploeg (1985, Section 2.1) could be applied. If we had priors on all latent variables ($m = n$), then Lemma A1 in the Appendix would give the desired posterior distribution. It is the joint occurrence of lacking data and insufficient prior information, which makes the problem difficult.

Clearly we need an identifiability condition, since each latent variable must be revealed either through the data or through the priors or both. A necessary and sufficient identifiability condition, easy to check, is given in Theorem 1, which also provides a complete solution to the general problem discussed above.

Theorem 1. Let x be an $n \times 1$ vector of latent variables and let d be a $p \times 1$ data vector such that

$$d|x \sim N_p(Dx, \Sigma), \tag{1}$$

where the $p \times n$ matrix D has full row-rank and Σ is positive definite (hence nonsingular). Suppose that prior information is available in the form

$$Ax \sim N_m(h, H), \tag{2}$$

where the $m \times n$ matrix A has full row-rank and H may be singular. If $m < n$, let L be a semi-orthogonal $n \times (n-m)$ matrix (that is, $L'L = I_{n-m}$) such that $AL = 0$, and assume that the identifiability condition

$$r(A) + r(DL) = n \tag{3}$$

is satisfied. Then the posterior distribution of x is given by

$$x|d \sim N_n(\mu, V) \tag{4}$$

with

$$V = A^+HA^{+\prime} - A^+HA^{+\prime}D'\Sigma_0^{-1}DA^+HA^{+\prime} + CKC' \tag{5}$$

and

$$\mu = A^+h - (A^+HA^{+\prime} + CK)D'\Sigma_0^{-1}(DA^+h - d), \tag{6}$$

where

$$\Sigma_0 = \Sigma + DA^+HA^{+\prime}D', \qquad A^+ = A'(AA')^{-1}, \qquad C = I_n - A^+HA^{+\prime}D'\Sigma_0^{-1}D, \tag{7}$$

and

$$K = \begin{cases} L(L'D'\Sigma_0^{-1}DL)^{-1}L', & \text{if } m < n, \\ 0, & \text{if } m = n. \end{cases} \tag{8}$$

Proof. See Appendix. ||

Let us provide a very simple example in order to demonstrate application of the theorem. In this example we have one national accounts identity $y = c + i + g$

and two independent data: $y = 230$ and $g = -44$. In addition, our prior belief is that c should be around 220.5 and i/c around 1/3. We do not yet know how to treat ratios (this is discussed in the next section), but we can deal with this situation naively as follows:

naive example:

data:	$y = 230$ (11.5)	*priors*:	$y = c + i + g$
	$g = -44$ (2.2)		$c = 220.5$ (11.025)
			$i = 73.5$ (3.675).

Standard errors are given in parentheses. In each case the coefficient of variation (standard error divided by mean) is assumed to be 5 percent. We have $n = 4$, $p = 2$, and $m = 3$. The vector of latent variables is $x = (y, c, i, g)'$, and

$$D = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} (11.5)^2 & 0 \\ 0 & (2.2)^2 \end{pmatrix}, \quad h = \begin{pmatrix} 0 \\ 220.5 \\ 73.5 \end{pmatrix}, \tag{9}$$

$$A = \begin{pmatrix} 1 & -1 & -1 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \quad H = \begin{pmatrix} 0 & 0 & 0 \\ 0 & (11.025)^2 & 0 \\ 0 & 0 & (3.675)^2 \end{pmatrix}. \tag{10}$$

The semi-orthogonal matrix L is $(1/\sqrt{2})(1, 0, 0, 1)'$ and the identifiability condition is satisfied. Application of Theorem 1 then gives the following results:

posterior moments:	$y = 239.7$ (8.2)
	$c = 211.6$ (8.2)
	$i = 72.5$ (3.6)
	$g = -44.4$ (2.2).

Clearly the accounting identity is exactly satisfied. One can experiment with this simple example to see how different assumptions on the precisions of data and priors affect the results. Equally, we could allow for correlations between priors or measurements (data).

It is easy to see that the identifiability condition in Theorem 1 is the most general possible—it is necessary and sufficient.

In classical statistics, it is well-known that there exists a close and non-trivial connection between generalized least squares and best linear unbiased estimation; see Rao (1973, pp. 294–302) and Magnus and Neudecker (1999, Section 13.18). An analogous result holds in Bayesian statistics. If we partition the prior information $Ax \sim N_m(h, H)$ into two parts:

$$\pi_1 : A_1 x \sim N_{m_1}(h_1, H_1), \qquad \pi_2 : A_2 x = h_2 \quad \text{(a.s.)}, \tag{11}$$

where h_1 has m_1 components, h_2 has $(m - m_1)$ components, and H_1 is positive definite, then a generalized least squares procedure would minimize

$$(d - Dx)'\Sigma^{-1}(d - Dx) + (A_1 x - h_1)'H_1^{-1}(A_1 x - h_1) \tag{12}$$

subject to the linear constraints

$$A_2 x = h_2. \tag{13}$$

Since (12) is proportional to the exponential term of the posterior when only π_1 is used, while also using the linear constraints (13) is equivalent to adding π_2, we see that minimization leads to the posterior mode, which, in view of the normality assumption, equals the posterior mean $x = \mu$ of Theorem 1. Hence, there exists a close connection between our Bayes solution and generalized least squares; see also Van der Ploeg (1985).

Analogous to the classical result that a generalized least squares estimator is best linear unbiased in the absence of the normality assumption (by the Gauss–Markov theorem), there is a Bayesian result that μ is a Linear Bayes estimator, obtained by minimizing the expected (with respect to data) quadratic loss in the class of linear functions of the data. This result does not depend on normality assumptions, only on first and second moments. As our posterior mean is linear in d, it may be justified as a Linear Bayes estimator; see e.g. Goldstein (1988). O'Hagan (1994, pp. 163–66) gives a short and critical review of Linear Bayes estimators.

Note that other assumptions than normality, like more heavy-tailed priors, lead to Bayesian posterior means that are not linear in the data. Numerical methods like the Gibbs sampler are then required to estimate the posteriors. However, the Linear Bayes estimator may still be used as a near-optimal simple device in such cases.

3. Indicator Ratios

In practice many of the priors will be nonlinear. In particular, many of the priors used in the construction of national accounts are "indicator ratios," that is, ratios of two latent variables. In this section we shall see how indicator ratios can be linearized in a suitable manner, so that Theorem 1 can still be applied.

Let x and y be latent variables and consider an indicator ratio $R = y/x$. Let r denote the prior expectation of R. Assume for the moment that we have prior information on R and x and that these priors are independent. This will not be strictly true in practice, but is nevertheless reasonable in many applications. Then we can easily show that

$$E(y - rx) = 0, \qquad \mathrm{var}(y - rx) = \mathrm{var}(R) \cdot (\mathrm{var}(x) + (Ex)^2). \tag{14}$$

Our strategy is to replace the prior R by its linearization $y - rx$. We have prior knowledge about the mean and variance of R, but *not* about the mean and variance of x. If we would know the mean and variance of x, then we could replace the prior on R by a prior on $y - rx$, and Theorem 1 could be applied. Since we do not know the moments of x, we use a simple iterative procedure, as follows. First, use (14) taking Ex to be the value of x in the previous year and let $\mathrm{var}(x) = 0$. Then, apply Theorem 1. This gives posterior moments of all latent variables, and hence in particular of x. In step 2 we use the posterior moments Ex and $\mathrm{var}(x)$ obtained in step 1 and recalculate the prior variance of $y - rx$ from (14).

Using this updated prior variance, we apply Theorem 1 again and continue this process until convergence.

We may or may not have access to last year's values. This is not important, since the resulting posterior estimates will be independent of the starting values of the iteration. In practice we will have not one but several indicator ratios. The iteration procedure is then applied to all of them simultaneously.

The linearization just described involves an approximation. Our experience shows that the approximation works well in practice, but of course there may be situations where it does not work well. Below we provide a further justification for the procedure.

Consider a very simple set-up with two latent variables x and y and one indicator ratio $R = y/x$, as follows:

$$\text{(15)} \qquad \textit{data}: \quad d|x \sim (x, \sigma^2)$$

$$\text{(16)} \qquad \textit{prior } \pi: \quad R \sim (r, \tau^2).$$

We wish to replace the prior π in (16) by its linearization π',

$$\text{(17)} \qquad \textit{prior } \pi': \quad y - rx \sim (0, \tau_*^2),$$

and we wish to choose τ_*^2 optimally in some sense. Given (15) and (17), the posterior moments of y given d can be calculated from Theorem 1 or directly. They are

$$\text{(18)} \qquad y|d \sim (rd, r^2\sigma^2 + r_*^2).$$

Given (15) and (16), what are the posterior moments of $y|d$? The prior π directly implies that

$$\text{(19)} \qquad E(y|x) = rx, \qquad \operatorname{var}(y|x) = \tau^2 x^2.$$

Let $z = x|d$. Then the posterior moments of $y|d$ are

$$\text{(20)} \qquad \begin{aligned} E(y|d) &= E_z E(y|d, z) = E_z E(y|d, x) \\ &= E_z E(y|x) = E_z(rx) = rd \end{aligned}$$

and

$$\text{(21)} \qquad \begin{aligned} \operatorname{var}(y|d) &= E_z \operatorname{var}(y|d, z) + \operatorname{var}_z E(y|d, z) \\ &= E_z \operatorname{var}(y|x) + \operatorname{var}_z E(y|x) \\ &= E_z(\tau^2 x^2) + \operatorname{var}_z(rx) \\ &= \tau^2(d^2 + \sigma^2) + r^2\sigma^2. \end{aligned}$$

We see that the posterior mean of $y|d$ is rd, both in the linearized and the non-linearized version. We now choose the prior variance τ_*^2 such that also the posterior variance of $y|d$ for the linearized prior (17) equals that for the non-linearized prior (16). This yields

$$\text{(22)} \qquad \tau_*^2 = \tau^2(d^2 + \sigma^2).$$

Hence, in this simple example, the iterative procedure is justified by the fact that it leads to the correct first two posterior moments. We notice that this method is not Linear Bayes, since the estimator is not linear in d. A Linear Bayes estimator would have a larger variance.

Continuing our simple example from the previous section, let us assume, instead of $i = 73.5$, that $i/c = 1/3$ or alternatively that $c/i = 3$. The results are presented in Table 1.

TABLE 1

POSTERIOR MEANS AND STANDARD ERRORS FOR THREE PRIOR SPECIFICATIONS OF THE SIMPLE MODEL

	Prior		
Posterior	$i = 73.5$	$i/c = 1/3$	$c/i = 3$
y	239.7 (8.2)	237.2 (9.2)	237.2 (9.2)
c	211.6 (8.2)	211.6 (7.1)	211.6 (7.0)
i	72.5 (3.6)	69.9 (3.7)	69.9 (3.7)
g	−44.4 (2.2)	−44.3 (2.2)	−44.3 (2.2)

The difference between the three specifications is very small. The two alternative prior indicator ratios ($i/c = 1/3$ and $c/i = 3$) do not necessarily yield the same posterior estimates, because different approximations are involved in linearizing the ratios. Nevertheless, the difference between $i/c = 1/3$ and $c/i = 3$ as prior indicator ratio is negligible.

We conclude that an iterative version of Theorem 1 can be suitably applied to linearized indicator ratios in order to obtain posterior moments of all latent variables in the system.

4. MULTIPLE PRIORS

There is one further possible complication. We may have many priors or several priors on the same latent variable. In such situations the $m \times n$ matrix A in the prior specification

$$Ax \sim N_m(h, H) \tag{23}$$

will have rank $r < m$ and the conditions of Theorem 1 are not satisfied. For example, if $y = c + i$, we may have priors on both c/y and i/y, and these will generally not add up to one. In such situations we will need to reflect carefully about the source of the conflict. Also it may occur that such priors are *not* in conflict. Then we need to assess how dependent the information is. Maybe the information on c/y and i/y originated from the same source, in which case they are perfectly correlated. Much has been written about combining expert's opinions; see Genest and Zidek (1986); Wiper and French (1995); and Clemen and Winkler (1999).

The most common approach to this problem uses different expert opinions like "data." We follow this approach and generalize it to a format that suits our goals. Let S be an $m\times(m-r)$ matrix and T an $m\times r$ matrix such that (S, T) is orthogonal,

$$S'S = I_{m-r}, \qquad T'T = I_r, \qquad S'T = 0, \tag{24}$$

and satisfies $A'S = 0$. (Hence S contains the eigenvectors associated with the $m-r$ zero eigenvalues of AA' and T contains the remaining r eigenvectors.) Premultiplying (23) by $(S, T)'$ gives the equivalent prior specification

$$\begin{pmatrix} S'Ax \\ T'Ax \end{pmatrix} \sim N_m\left[\begin{pmatrix} S'h \\ T'h \end{pmatrix}, \begin{pmatrix} S'HS & S'HT \\ T'HS & T'HT \end{pmatrix}\right], \tag{25}$$

and hence $T'Ax|(S'Ax=0) \sim N_r(h^*, H^*)$ with

$$h^* = T'h - T'HS(S'HS)^+S'h, \qquad H^* = T'HT - T'HS(S'HS)^+S'HT, \tag{26}$$

where $(S'HS)^+$ denotes the Moore–Penrose inverse of $S'HS$. Instead of the prior information on Ax contained in (23) we now consider the prior information on $T'Ax$. The matrix $T'A$ has full row-rank and thus satisfies the conditions of Theorem 1.

A trivial example may illustrate this procedure. Suppose we have two pieces of information on a single latent variable x,

$$x\begin{pmatrix} 1 \\ 1 \end{pmatrix} \sim N\left[\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & \tau^2 \end{pmatrix}\right]. \tag{27}$$

According to one expert, $Ex = 1$; according to another expert with independent information, $Ex = 2$. Then,

$$S = (1/2)\sqrt{2}\begin{pmatrix} 1 \\ -1 \end{pmatrix}, \qquad T = (1/2)\sqrt{2}\begin{pmatrix} 1 \\ 1 \end{pmatrix}. \tag{28}$$

Substituting into (23) gives

$$x \sim N(2-\lambda^2, \lambda^2), \qquad \lambda^2 = \tau^2/(1+\tau^2), \tag{29}$$

the standard weighted average of the two pieces of information.

5. An Example

We illustrate the proposed methodology by investigating a simple but realistic accounting framework with data based on a Western-type economy. The framework includes national aggregates, such as GDP, expenditures and external transactions, that allow for the derivation of national disposable income and national saving (not explicitly included in the presented accounts). It also includes the main institutional sector accounts, as published in the 1993 System of National Accounts (United Nations *et al.*, 1993).

The framework and data, involving 38 variables distributed over 5 sectors, is presented in Table 2 together with last year's data. The framework and the data are described in detail in United Nations (2000, Chapter 4A) and are consistent with the international guidelines of the 1993 System of National Accounts (SNA) described in United Nations *et al.* (1993).

TABLE 2

National Economic Accounts, Example with Last Year's Data

Total Economy		Resident Sectors		
Industries	Rest of the World	Government	Non-financial and Financial Corporations	Households
[1] Output, incl. product taxes, less subsidies 3,737	[10] Imports 499			
[2] Intermediate consumption 1,883	[11] Exports 540	[17] Final consumption, government 368		[26] Final consumption, housholds 1,031
[3] Gross capital formation, total economy[1] 414		[18] Gross capital formation, government[1] 40	[22] Gross capital formation, corporations[1] 287	[27] Gross capital formation, households[1] 87
[4] GDP, market prices (current prices) 1,854 [5] GDP, market prices (constant prices) 1,160	[12] External balance of goods and services 41			[28] Disposable income, gross[5] 1,259 [29] Disposable income before taxes, gross[6] 1,437 [30] Savings, gross 228
[6] Compensation of employees, paid and mixed income, gross 1,204	[13] Compensation of employees received by residents, less paid to non-residents 4			[31] Compensation of employees and mixed income received 1,208
[7] employment (1,000 m/years worked) 33,350				[32] Size of population (× 1000), beginning of year 88,700 [33] Size of population (× 1000), end of year 90,000 [34] Population increase (× 1000) 1,300
[8] Taxes on production and imports, less subsidies[2] 191	[14] Taxes on production less subsidies plus taxes on income and wealth, received by resident government less paid to non-resident government 1	[19] Taxes on production less subsidies plus taxes on income and wealth, received by government 404	[23] Taxes on income and wealth, paid by corporations 34	[35] Taxes on income and wealth, paid by households 178

TABLE 2—continued

Total Economy		Resident Sectors		
Industries	Rest of the World	Government	Non-financial and Financial Corporations	Households
[9] Operating surplus, gross (excl. mixed income) 459	[15] Other incomes, receipts by residents less payments to non-residents[3] −8	[20] Other outlays, payments less receipts, by government[4] 46	[24] Other incomes, receipts less payments, by corporations[3] 257	[36] Other incomes, receipts less payments, by households[3,7] 229 [37] Capital transfers, receipts less payments, households 11
	[16] Net lending to abroad 38	[21] Net lending, government −50	[25] Net lending, corporations −64	[38] Net lending, households 152

Note: All entries, except [7] and [32]–[34], are in millions of US$.

[1]Gross capital formation includes the value of improvements to land and the cost of ownership transfers of non-produced assets.

[2]Production taxes less subsidies have not been allocated to sectors, but have only been recorded for the total economy.

[3]Other incomes, receipts less payments, include operating surplus gross, property income, and non-tax current and capital transfers. Capital transfers include acquisition less disposal of non-produced non-financial assets.

[4]In the case of the government, other incomes have been replaced by other outlays, which include payments less receipts of property income and non-tax current and capital transfers, less operating surplus gross.

[5]Includes adjustment for the change in net equity of households on pension funds, and is after deduction of taxes on income and wealth.

[6]No tax deductions have been made.

[7]In the case of households, operating surplus excludes mixed income and capital transfers, which are presented separately.

The columns of the Table 2 refer to sectors of the economy and the rows to accounts. The first column contains the aggregate data on industries, while the other columns refer to the rest of the world and three aggregate resident institutional sectors, i.e. government, non-financial and financial corporations, and households. The rows of accounting data for each sector are grouped together by four accounting segments. The first segment, contained in the first three row-blocks, refers to the data elements of the supply and use table, i.e. output, imports, exports, intermediate and final consumption, and capital formation. The second segment, covering row-blocks 4 and 6, refers to the main product, income and related aggregates including GDP at current and constant prices, disposable income before and after taxes, saving, and the corresponding employment and population data needed to derive product and income aggregates per worker and per capita. The third segment, covering row-blocks 5, 7 and 8, refers to receipts and payments of compensation of employees including mixed income, taxes, operating surplus and other income and outlay data. The fourth segment, contained in row-block 9, includes net lending for each sector. This is used as the main analytical balancing item in each sector, except in the household sector where also disposable income and savings are shown.

The 38 variables of the accounting framework must satisfy 16 linear accounting restrictions: 10 "vertical" restrictions that define the accounting constraints within each sector, and 6 "horizontal" restrictions that correspond to accounting restrictions between the sectors. These are given in Table 3. In fact, there is one further "horizontal" restriction referring to net lending between sectors: [16] = [21] + [25] + [38], but this identity is linearly dependent on the other restrictions.

TABLE 3

ACCOUNTING IDENTITIES

Industries	[4] = [1] − [2]
	[9] = [4] − [6] − [8]
Rest of the world	[12] = [11] − [10]
	[16] = [12] + [13] + [14] + [15]
Government	[21] = [19] − [17] − [18] − [20]
Corporations	[25] = [24] − [23] − [22]
Households	[30] = [28] − [26]
	[29] = [28] + [35]
	[36] = [28] − [31] + [35]
	[38] = [30] + [37] − [27]
Supply and use	[1] + [10] = [2] + [3] + [11] + [17] + [26]
Gross capital formation	[3] = [18] + [22] + [27]
Compensation of employees	[6] + [13] = [31]
Population	[34] = [33] − [32]
Taxes	[19] = [8] + [14] + [23] + [35]
Other incomes and outlays	[24] + [36] + [37] − [9] − [15] − [20] = 0

An essential ingredient in estimating the 38 latent variables is prior, but uncertain, knowledge about some indicator ratios. We have selected, on economic grounds, 16 indicator ratios on which we believe reasonable prior knowledge to be available. These were selected from a larger list of indicator ratios presented in United Nations (2000, Chapter 4A). The 16 indicator ratios, together with their prior moments, are presented in Table 4.

Each indicator ratio except one has a prior mean which is assumed to be equal to last year's value, that is, our prior belief is that the indicator ratio remains the same. The only exception is the inflation rate, which we give a prior value of 2 percent, based on the results of a limited consumer price survey. As a consequence, the GDP price deflator [4]/[5] has a prior mean of last year's value multiplied by 1.02. We emphasize that the choice of prior means based on last year's values is *not* demanded by some theory. In the presence of more insights into the economy, other prior means could be used.

The assumed precision (uncertainty) of the indicator ratios is given as high (H), medium-high (MH), medium-low (ML), or low (L). These four categories indicate the coefficient of variation, that is, standard error divided by mean. In particular, "H" indicates a coefficient of variation of 0.5 percent, "MH" of 2.5 percent, "ML" of 5 percent, and "L" of 10 percent. For example, the first indicator ratio [3]/[4] has a prior mean of 0.2233 and a coefficient of variation of 2.5 percent. Hence the standard error is $0.025 \times 0.2233 = 0.0056$. In our example, the precision of the indicator ratios is "H," "MH" or "ML," but never "L."

In addition to the indicator ratios, we have four "other" priors, namely variables which we assume a priori will not change compared to last year. These are given in Table 5.

TABLE 4

INDICATOR RATIOS WITH PRIOR MOMENTS

				Prior Moments		
			Indicator Ratios	Mean	Coefficient of Variation	Standard Error
Indicator ratios vertically defined within industries and sectors	Production by industries	[3]/[4]	Investment share in GDP	0.2233	MH	0.0056
		[4]/[1]	Value added/output coefficient, total	0.4961	H	0.0025
		[4]/[5]	GDP price deflator	1.6302	H	0.0081
		[5]/[7]	GDP constant prices per worker, labor productivity (1,000 US$ per m/year)	0.0348	MH	0.0009
		[6]/[4]	Labor share in GDP	0.6494	H	0.0032
		[8]/[4]	Production taxes less subsidies/GDP	0.1030	MH	0.0026
	Corporations	[23]/[24]	Taxes/revenues of corporations	0.1323	MH	0.0033
		([24] − [23])/[22]	Earnings (after taxes)/gross capital formation, corporations	0.7770	MH	0.0194
	Households	[26]/[28]	Propensity to consume of households	0.8189	MH	0.0205
		[27]/[30]	Capital formation/saving, households	0.3816	MH	0.0095
		[31]/[29]	Labor income as share of disposable income of households, before taxes	0.8406	MH	0.0210
		[35]/[29]	Tax ratio of household disposable income, before taxes	0.1239	MH	0.0031
		[34]/[32]	Population growth	0.0147	ML	0.0007
Indicator ratios horizontally defined across industries and sectors		[10]/([1] + [10])	Import/supply-use	0.1178	MH	0.0029
		[17]/[26]	Government/household consumption ratio	0.3569	MH	0.0089
		[7]/[33]	Number of employees/population	0.3706	MH	0.0093

TABLE 5

OTHER PRIORS USED

Variable	Prior Moments		
	Mean	Coefficient of Variation	Standard Error
[13] Compensation of employees received by residents, less paid to non-residents	4	L	0.4
[14] Taxes on production less subsidies plus taxes on income and wealth, received by resident government less paid to non-resident government	1	L	0.1
[15] Other incomes, receipts by residents less payments to non-residents	−8	L	0.8
[37] Capital transfers, receipts less payments, households	11	L	1.1

Tables 3–5 give the accounting restrictions, the indicator ratios, and the "other" priors, 36 priors all together. However, we also have data. In fact we have data on 19 of our 38 variables. These are given, together with their assumed precisions (measurement errors), in Table 6.

Of course, the coefficient of variation is not defined when the mean is zero, and is not very useful when the mean is "close" to zero. Here we let the standard error be 0.5 when the mean (in absolute value) is smaller than 10.

Thus, we have 38 variables ($n = 38$), 19 data ($p = 19$), and 36 priors ($m = 36$) of which 16 identities. Using the iterative version of Theorem 1 on linearized indicator ratios (see Section 3), we obtain posterior means and variances of all 38 latent variables. Of the total number of 38 variables, we selected 11 "key" variables and 11 "key" indicator ratios. These are considered to be the essential elements for assessing the state and development of the economy under investigation.

5.1. *Key Variables*

Bayesian estimates (together with their standard errors) of key variables are presented in Table 7 (under "comprehensive compilation"). The key variables include, among others, GDP at market prices, gross capital formation, disposable income of households, and net lending of all sectors, including corporations, government, households and abroad. Some of the key variables can be estimated very accurately: gross capital formation in the total economy [3], GDP [4], final consumption government [17], final consumption households [26], and gross disposable income [28]. However, some of the key variables, in particular the balancing items, are estimated less accurately, as was to be expected. These include net lending by government [21], corporations [25] and households [38], and also gross household savings [30]. The most difficult to estimate are two key variables relating to the rest of the world: external balance of goods and services [12] and net lending to abroad [16]. Qualitatively these results appear to be sensible. The power of our approach, however, lies in the quantification of estimates and precisions.

TABLE 6

THE DATA AND THEIR PRECISIONS

Data	Mean	Moments: Coefficient of Variation	Standard Error
[1] Output, incl. product taxes, less subsidies	4,034	ML	201.7
[3] Gross capital formation, total economy	449	L	44.9
[6] Compensation of employees, paid and mixed income, gross	1,252	MH	31.3
[7] Employment (1000 m/years worked)	33,657	MH	841.4
[10] Imports	543	MH	13.6
[11] Exports	567	MH	14.2
[13] Compensation of employees received by residents, less paid to non-residents	3	*	0.5
[14] Taxes on production less subsidies plus taxes on income and wealth, received by resident government less paid to non-resident government	0	*	0.5
[15] Other incomes, receipts by residents less payments to non-residents	−10	ML	0.5
[17] Final consumption, government	385	MH	9.6
[18] Gross capital formation, government	41	MH	1.03
[19] Taxes on production less subsidies plus taxes on income and wealth, received by government	366	MH	9.2
[20] Other outlays, payments less receipts, by government	48	MH	1.2
[22] Gross capital formation, corporations	316	ML	15.8
[23] Taxes on income and wealth, paid by corporations	29	ML	1.45
[24] Other incomes, receipts less payments, by corporations	275	ML	13.75
[26] Final consumption, households	1,070	L	107.0
[32] Size of population (× 1,000), beginning of year	90,000	MH	2,250
[33] Size of population (× 1,000), end of year	91,200	MH	2,280

Note: When the prior mean (in absolute value) is smaller than 10, we set the standard error at 0.5.

We now compare our Bayesian estimates with the estimates based on a formalization of current practices (hereafter "SNA estimates"). These current practices do not involve the use of reliability intervals, and as a consequence can accommodate fewer structural coefficients (indicator ratios). The SNA estimates were obtained by simulating current compilation practices. The compilation is carried out in two stages. The first stage is based on the 19 data (the same 19 data that are used in the Bayesian approach). To "estimate" the 38 variables, we add the 16 restrictions (the same restrictions that we used in the Bayesian approach) and a small number (5 in this case) of the available indicator ratios. The five selected indicator ratios are those that are closest to the ones used in current practices, and consist of value added/output coefficient, total; the GDP price deflator; GDP in constant prices per worker, labor productivity; production taxes less subsidies/GDP; and also capital formation/saving of households (see

TABLE 7

Bayesian Estimates for "Key" Variables

Key Variable	Previous Year	Comprehensive Compilation: Current SNA Practice	Comprehensive Compilation: Bayes	Reduced Compilation: Bayes
[3] Gross capital formation, total economy	414	490	424 (8.8)	434 (11.2)
[4] GDP, market prices (current prices)	1,854	2,001	1,890 (22.9)	1,938 (30.8)
[12] External balance of goods and services	41	25	29 (13.1)	43 (15.0)
[16] Net lending to abroad	38	18	24 (13.1)	40 (15.0)
[17] Final consumption, government	368	385	380 (7.2)	384 (8.0)
[21] Net lending, government	−50	−107	−71 (8.2)	−51 (9.4)
[25] Net lending, corporations	−64	−87	−67 (5.6)	−67 (6.6)
[26] Final consumption, households	1,031	1,102	1,058 (19.6)	1,076 (24.4)
[28] Disposable income, gross, households	1,259	1,435	1,302 (20.4)	1,313 (24.2)
[30] Savings, gross, households	228	333	244 (16.0)	237 (17.6)
[38] Net lending, households	152	211	162 (10.1)	158 (11.1)

Table 4). Prior knowledge of the accuracy of the data is largely ignored at this stage. We have now 40 (19 + 16 + 5) pieces of information to estimate the 38 variables, and hence we have two degrees of freedom. In order to identify the 38 variables, two of the exact restrictions are ignored at this stage: supply and use, and other incomes and outlays. The "estimates" obtained in this way do not satisfy these two restrictions, and hence inconsistencies (statistical discrepancies) occur at this first stage.

These statistical discrepancies are eliminated in the second stage of the compilation, by adjusting the estimates of the variables, and indirectly also the values of the indicator ratios used in the first stage. Implicit in the latter procedure are qualitative reliability criteria regarding data and indicator ratios, which the national accountants take into account. In particular the estimates of two variables are adjusted at this stage, which are considered to be less reliable, i.e. gross capital formation and final consumption of households. For a more comprehensive description of the formalized approach to current national accounts compilation practices, the reader is referred to United Nations (2000, V.D.1).

The SNA estimates of the key variables obtained in this way are also presented in Table 7. The discrepancy from our Bayesian estimates is substantial.

5.2. *Key Indicator Ratios*

The key indicator ratios are presented in Table 8. They are closely related to the key variables and include, among others, GDP per capita, household disposable income per capita, GDP real growth, and the share of investment and total

TABLE 8

BAYESIAN ESTIMATES FOR "KEY" INDICATOR RATIOS

	Key Indicator Ratio	Previous Year	Comprehensive Compilation: Current SNA Practice	Comprehensive Compilation: Bayes	Reduced Compilation: Bayes
[28]/[33]	Household disposable income/capita	13,989	15,739	14,305 (300)	14,399 (392)
[4]/[33]	Per capita GDP	20,600	21,944	20,760 (388)	21,246 (544)
[12]/[4]	Export–import gap/GDP (%)	2.2	1.2	1.5 (0.7)	2.2 (0.8)
[3]/[4]	Investment share in GDP (%)	22.3	24.5	22.4 (0.4)	22.4 (0.5)
$([5]-[5]_{-1})/[5]_{-1}$	GDP real growth (%)	***	5.8	0.0 (1.3)	1.5 (3.1)
−[21]/[4]	Government net borrowing/ GDP (%)	2.7	5.3	3.8 (0.4)	2.6 (0.5)
[19]/[4]	Total taxes/GDP (%)	21.8	18.3	21.0 (0.3)	21.8 (0.3)
[26]/[28]	Propensity to consume of households (%)	81.9	76.8	81.2 (1.1)	82.0 (1.3)
[17]/[26]	Government/household consumption ratio (%)	35.7	34.9	35.9 (0.7)	35.7 (0.7)
[26]/[4]	Household consumption/ GDP (%)	55.6	55.1	56.0 (0.7)	55.5 (0.8)
[38]/([38]+[16])	Net lending of households/total net lending (%)	80.0	92.4	87.2 (5.7)	79.7 (5.3)

taxes in GDP. Some of the key indicator ratios ([3]/[4], [26]/[28] and [17]/[26]) are also used as priors, but most are not. The Bayesian estimates of the key indicator ratios are presented under "comprehensive compilation" in Table 8, and can be compared with the SNA estimates. Some of the differences between the SNA and our Bayesian estimates are quite large. For example, we estimate the export–import gap/GDP ratio at 1.5 percent, while the SNA estimate is 1.2 percent. Also, we estimate the government net borrowing/GDP ratio at 3.8 percent, while the SNA estimate is 5.3 percent. The largest difference between the two approaches is in GDP real growth, where our Bayesian estimate indicates zero growth, whereas the SNA approach estimates 5.8 percent growth. The difference is so large, because in current practices, as interpreted here, GDP real growth is solely dependent on the observed growth of output to which it is linked through another indicator ratio, namely the output value-added coefficient. In contrast, the Bayesian approach takes into account the values of all indicator ratios listed in Table 4, and thus also the much more limited growth of population/employment (1.3 percent). This large difference in GDP growth between the two estimation approaches influences, of course, the value of all other indicator ratios that are dependent on GDP. It thus explains the large difference between the two estimates for the export-import gap/GDP ratio (1.5 percent versus 1.2 percent) and net borrowing/GDP (3.8 percent versus 5.3 percent).

In Table 8 we present standard errors of the estimated Bayesian indicator ratios. These standard errors are approximations based on (14). Thus,

$$\text{var}(y/x) \approx \frac{\text{var}(y) + r^2\,\text{var}(x) - 2r\,\text{cov}(y,x)}{\text{var}(x) + (Ex)^2}, \tag{30}$$

where

$$r = Ey/Ex \tag{31}$$

and Ex, Ey, var(x), var(y), and cov(y, x) denote posterior moments. We see that the export–import gap/GDP ratio and in particular GDP real growth are difficult to estimate precisely.

5.3. *Comprehensive and Reduced Compilation*

Typically, national statistics offices produce not one, but three rounds of estimates of the national accounts variables. Suppose we wish to estimate the variables for year t. The first round takes place in the winter or early spring of year $t+1$. Few actual data are then available, so one has to rely heavily on priors. This round is here called "reduced compilation." The next round may take place one year later and is here called "comprehensive compilation." More data and more accurate data are then available. The final round takes place one year after this.

Our analysis so far can be viewed as the second round ("comprehensive compilation"), where we have access to 19 data. In the first round ("reduced compilation") we have only 8 data: [6], [7], [11], [17], [18], [20], [24], and [32]. The only other difference between the two rounds is that, in the absence of a price survey, the inflation rate is assumed to be 0 percent (rather than 2 percent), so that the GDP price deflator [4]/[5] has a prior mean of 1.5983. We assume that its precision is low (L). The resulting estimates for the key variables and the key indicator ratios are presented in Tables 7 and 8, last panel.

About one half of the key indicator ratios are rather well estimated at the "reduced compilation" stage: household disposable income per capita, investment share in GDP, propensity to consume of households, government/household consumption ratio, and household consumption/GDP. These five indicator ratios are not very sensitive to having more data. Two other indicator ratios (per capita GDP and total taxes as a percentage of GDP) are moderately sensitive, while the other indicator ratios, in particular the GDP real growth rate and the export–import gap/GDP, are poorly estimated at the "reduced compilation" stage. All estimates become more precise when more data are available.

6. Concluding Remarks

The Bayesian estimation approach developed in this paper allows us to obtain estimates for the variables and the indicator ratios between the variables, and also reliability intervals of these estimates. The "simultaneity" feature of this approach introduces several new elements in current national accounts estimation practices. First and most importantly, the approach takes full account of all available information together with the assumed prior precision of that information. Secondly, if we have several pieces of information on the same variable or set of variables, then this causes no problems (multiple priors). Thirdly, all variables and indicator ratios are estimated with their corresponding reliability intervals.

This is new, as variable estimates in national accounts are generally point estimates without estimates of the standard errors.

The basics of the approach have been worked out and illustrated above. Its future potential, however, will depend on the further development of the approach for practical application. Some of these potentials are briefly discussed below.

It is important that through the Bayesian approach a direct link is established between the value and reliability of basic data and the value and reliability of the estimates of the national accounts variables. This link was used above to show how improvements or extensions of basic data sources would lead to reduced reliability intervals of posteriors when additional data sources are employed, and thus to improved reliability of the national accounts estimates. The same link could also be used to quantify the additional reliability that would be achieved by integrating basic data sources through the use of the national accounts framework.

The method also facilitates the simultaneous use of indicator ratios in compilation and analysis. This is important as indicator ratios are generally the core of simple analyses (United Nations, 2000, Chapter 4A). Depending on the availability of basic data at the time or the circumstances that national accounts are compiled, current national accounts estimation approaches use fixed values or point estimates of a few selected indicator ratios as assumptions or "priors" in the national accounts compilation, and "posterior" values of the indicator ratios are close to their "prior" values. This practically eliminates the use of these indicator ratios in analysis. In the Bayesian approach the "prior" and "posterior" values of the selected indicator ratios may be different as the "priors" are not point estimates, but defined with help of reliability intervals. As a result, they can still play a role in analysis, together with other indicator ratios that were not used in the compilation.

The example of the Bayesian approach given in Section 5 is a relatively simple one, based on an aggregated national economic accounts framework with limited scope and detail. This was done solely in order to illustrate the essential features of the method. However, the approach can be applied without any difficulty to a much larger number of variables and indicator ratios and thus to a more realistic accounting framework. There are plans to develop the approach further for use in preliminary accounts when a limited set of basic data is available, to comprehensive annual and benchmark economic accounts, and also to satellite accounts. The further development would be closely aligned with the so-called "systems approach" to macro accounts compilation, which has been developed by the UN Statistics Division and implemented in several countries; see United Nations (1999). An intriguing conclusion from our example is that prior values of indicator ratios may have a larger impact on the posterior estimates of the variables than improved or additional information obtained from new data sources, such as household or enterprise surveys. Of course, this tentative conclusion depends on the assumed precisions of the indicator ratios. It would be interesting to verify the conclusion by further elaboration of the approach for more detailed data systems.

The direct estimation of indicator ratios together with the underlying values of the variables can also be extended to the integration of estimation approaches used in accounting and modeling. In the current approach, indicator ratios are largely defined between variables within the same period. However, the approach could also be applied to accounting covering several periods. This would involve inter-temporal indicator ratios such as growth rates and capital output ratios, which are defined between variables of different periods. Using indicator ratios in this sense would establish a first link between the Bayesian approach and simple modeling based on the use of indicator ratios defined within and between periods. Finally, once the inter-temporal problem has been worked out, the approach could be extended to a more complex link between the Bayesian approach to macro accounts compilation and parameter estimation methods used in econometrics.

7. Appendix: Proof of Theorem 1

The proof proceeds in three steps of increasing generality on the assumed prior distribution. In Lemma A1 we assume that prior information is available on *all* latent variables.

Lemma A1. Let x be an $n\times 1$ vector of latent variables and let d be a $p\times 1$ data vector such that $d|x\sim N_p(Dx,\Sigma)$, where the $p\times n$ matrix D has full row-rank and Σ is positive definite. Suppose that prior information is available in the form $x\sim N_n(q,Q)$, where Q is positive semidefinite (possibly singular). Then the posterior distribution of x is given by $x|d\sim N_n(\mu,V)$, with

$$(32)\quad V=Q-QD'(\Sigma+DQD')^{-1}DQ,\qquad \mu=q-QD'(\Sigma+DQD')^{-1}(Dq-d).$$

Proof. It is obvious and well-known that the posterior of x is normally distributed. Assume first that Q is nonsingular. Then we find the moments μ and V by completing squares:

$$(33)\quad (d-Dx)'\Sigma^{-1}(d-Dx)+(x-q)'Q^{-1}(x-q)=(x-\mu)'V^{-1}(x-\mu)+R,$$

where R does not depend on x. This gives

$$(34)\quad V^{-1}=D'\Sigma^{-1}D+Q^{-1},\qquad V^{-1}\mu=D'\Sigma^{-1}d+Q^{-1}q,$$

and hence the expressions in the lemma. If Q is singular, the expressions remain valid, because $\Sigma+DQD'$ remains nonsingular. ||

The next step contains the crux of the proof. Here we allow some of the priors to be non-informative (have infinite variance).

Lemma A2. Assume that the conditions of Lemma A1 hold. Assume further that

$$(35)\quad Q=Q_0+(1/\lambda^2)LL',$$

where Q_0 is positive semidefinite, L has full column-rank ≥ 1, and the identifiability condition $r(DL)=r(L)$ is satisfied. Then, as $\lambda^2\to 0$, the posterior distribution of x is given by $x|d\sim N_n(\mu,V)$, with

$$(36)\quad V=Q_0-Q_0D'\Sigma_0^{-1}DQ_0+CKC'$$

and

(37) $$\mu = q - (Q_0 + CK)D'\Sigma_0^{-1}(Dq - d),$$

where

(38) $$\Sigma_0 = \Sigma + DQ_0D', \qquad C = I - Q_0D'\Sigma_0^{-1}D, \qquad K = L(L'D'\Sigma_0^{-1}DL)^{-1}L'.$$

Proof. We apply the results of Lemma A1. Letting $R = \Sigma_0^{-1/2}DL$, we have

(39) $$\Sigma + DQD' = \Sigma_0 + (1/\lambda^2)DLL'D' = \Sigma_0^{1/2}(I + (1/\lambda^2)RR')\Sigma_0^{1/2}.$$

The identifiability condition implies that R has full column-rank and hence that $R'R + \lambda^2 I$ has full rank, also at $\lambda^2 = 0$. Now,

(40) $$(\Sigma + DQD')^{-1} = \Sigma_0^{-1/2}(I - R(R'R + \lambda^2 I)^{-1}R')\Sigma_0^{-1/2}$$

and hence

(41) $$QD'(\Sigma + DQD')^{-1} = Q_0D'\Sigma_0^{-1/2}(I - R(R'R + \lambda^2 I)^{-1}R')\Sigma_0^{-1/2} + L(R'R + \lambda^2 I)^{-1}R'\Sigma_0^{-1/2}.$$

This gives

(42) $$\begin{aligned} V &= Q - QD'(\Sigma + DQD')^{-1}DQ \\ &= Q_0 - Q_0D'\Sigma_0^{-1}DQ_0 - CL(R'R + \lambda^2 I)^{-1}R'\Sigma_0^{-1/2}DQ_0 \\ &\quad + (1/\lambda^2)CL(I - (R'R + \lambda^2 I)^{-1}R'R)L' \\ &= Q_0 - Q_0D'\Sigma_0^{-1}DQ_0 + CL(R'R + \lambda^2 I)^{-1}L'C' \end{aligned}$$

and

(43) $$\begin{aligned} \mu &= q - QD'(\Sigma + DQD')^{-1}DQ \\ &= q - (Q_0 + L(R'R + \lambda^2 I)^{-1}L' \\ &\quad - Q_0D'\Sigma_0^{-1/2}R(R'R + \lambda^2 I)^{-1}L')D'\Sigma_0^{-1}(Dq - d) \\ &= q - (Q_0 + CL(R'R + \lambda^2 I)^{-1}L')D'\Sigma_0^{-1}(Dq - d). \end{aligned}$$

Letting $\lambda^2 \to 0$ gives the desired results. ||

Based on Lemmas A1 and A2 we can now prove Theorem 1.

Proof of Theorem 1. If $m = n$, Theorem 1 follows from Lemma A1 by letting $q = A^{-1}h$ and $Q = A^{-1}HA^{-1\prime}$. If $m < n$, we have less than n "informative" priors. To the m informative priors Ax we now add $n - m$ non-informative priors $L'x$. Since $(A', L)^{-1} = (A^+, L)'$, one verifies that the two statements

(44) $$\begin{pmatrix} Ax \\ L'x \end{pmatrix} \sim N\left[\begin{pmatrix} h \\ 0 \end{pmatrix}, \begin{pmatrix} H & 0 \\ 0 & (1/\lambda^2)I \end{pmatrix}\right]$$

and

(45) $$x \sim N(A^+h, A^+HA^{+\prime} + (1/\lambda^2)LL')$$

are equivalent. Hence, prior information $Ax \sim N(h, H)$ is equivalent to prior information

$$x \sim N(q, Q_0 + (1/\lambda^2)LL'), \qquad q = A^{+}h, \qquad Q_0 = A^{+}HA^{+\prime}, \tag{46}$$

when λ^2 approaches 0. Direct application of Lemma A2 now yields the required results. ||

References

Barker, T., F. van der Ploeg, and M. Weale, A Balanced System of National Accounts for the United Kingdom, *The Review of Income and Wealth*, 30, 461–85, 1984.

Byron, R. P., The Estimation of Large Social Account Matrices, *Journal of The Royal Statistical Society, Series A*, 141, 359–67, 1978.

Clemen, R. T. and R. L. Winkler, Combining Probability Distributions from Experts in Risk Analysis, *Risk Analysis*, 19, 187–203, 1999.

Genest, C. and J.V. Zidek, Combining Probability Distributions: A Critique and Annotated Bibliography, *Statistical Science*, 1, 114–48, 1986.

Goldstein, M., Adjusting Belief Structures, *Journal of the Royal Statistical Society, Series B*, 50, 133–54, 1988.

Magnus, J. R. and H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Econometrics*, revised edition, John Wiley, Chichester/New York, 1999.

Morgenstern, O., *On the Accuracy of Economic Observations*, second edition, Princeton University Press, Princeton, 1963.

O'Hagan, A., *Bayesian Inference*, Kendall's Advanced Theory of Statistics, Volume 2B, Edward Arnold, London, 1994.

Rao, C. R., *Linear Statistical Inference and Its Applications*, 2nd edition, John Wiley, New York, 1973.

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

United Nations Development Program, *Human Development Report*, Oxford University Press, New York/Oxford, 1996.

United Nations, Eurostat, IMF, OECD, and World Bank, *A System of National Accounts*, New York, Brussels/Luxembourg, Paris, Washington, DC, 1993.

United Nations, Statistical Office, *A System of National Accounts*, Studies in Methods, Series F, Volume 2, No. 3, United Nations, New York, 1968.

United Nations, Statistics Division, *Handbook of National Accounting: A Systems Approach to National Accounts Compilation*, Studies in Methods, Series F, No. 77, United Nations, New York, 1999.

———, *Handbook of National Accounting: Use of Macro Accounts in Policy Analysis*, Studies Methods, United Nations, New York, 2000, forthcoming.

Van der Ploeg, F., Reliability and the Adjustment of Sequences of Large Economic Accounting Matrices (with discussion), *Journal of the Royal Statistical Society, Series A*, 145, 169–94, 1982.

———, Economic Policy Rules for Risk-Sensitive Decision Making, *Zeitschrift für Nationalökonomie*, 44, 207–35, 1984.

———, FIML Estimation of Dynamic Econometric Systems from Inconsistent Data, *International Journal of Systems Science*, 16, 1–29, 1985.

Van Tongeren, J. W., Development of an Algorithm for the Compilation of National Accounts and Related Systems of Statistics, *The Review of Income and Wealth*, 31, 25–47, 1985.

Wiper, M. P. and S. French, Combining Experts' Opinions Using a Normal-Wishart Model, *Journal of Forecasting*, 14, 25–34, 1995.

World Bank, *World Development Report, Investing in Health*, Oxford University Press, New York/Oxford, 1993.