European Conference on Quality in Official Statistics (Q2016)
Madrid, 31 May-3 June 2016

Automatic balancing of the National Accounts

Yingfu Xie[1], Andreas Lennmalm[2], Daniel Lennartsson[3], Annica de Groote[4]

[1] *Statistics Sweden, Stockholm, Sweden; Yingfu.Xie@scb.se*
[2] *Statistics Sweden, Stockholm, Sweden; Andreas.Lennmalm@scb.se*
[3] *Statistics Sweden, Stockholm, Sweden; Daniel.Lennartsson@scb.se*
[4] *Statistics Sweden, Stockholm, Sweden; Annica.Isaksson@scb.se*

Abstract
The uncertainty of the national accounts is of great interest for decision-makers, researchers, and the public. Despite this, the information is most often absent in statistic releases. This is partly due to the complexity of the compilation of the national accounts, which rely on input from a large number of data sources. It is difficult to estimate all the uncertainties of all the inputs. Furthermore, national accounts must comply with the restrictions of the accounting system. When it comes to GDP the supply and use of each product has to be balanced. Estimates of uncertainties of the GDP have to take this balancing into account.
In 1942, Stone, Champernowne, and Meade (hereafter SCM) proposed a weighted least-square approach to balance the economic accounts in a limited scale. This approach has been further extended and applications to large scale national accounts have been reported. The main difficulty has proved to be the estimates of the uncertainties of all data sources.
In our work, the SCM balancing approach is studied in a large scale supply-use framework in the Swedish national accounts. Efforts are made to estimate the uncertainties not only from sampling errors, but also from nonsampling errors. The resulting estimates are used as weights in an automatic balancing procedure. The procedure has great potential compared to procedure used at Statistics Sweden today, which is mainly manual, and highly resource- and time-demanding.

Keywords: National Accounts, Balancing, SCM, Nonsampling error.

1. Introduction

The uncertainty of the national accounts (NA), such as the gross domestic product (GDP), is of great interest for decision-makers, researchers, and the public. Still, the information is often absent in statistic releases. This is partly due to the complexity of the compilations of NA, which rely on input from a large number of data sources. It is difficult to estimate all the uncertainties of all the inputs. In a report from Eurostat (2001), the data sources and possible error sources of the NA compilation are well documented, and the difficulties related to identifying and quantifying the errors (in particular the non-sampling errors) are discussed. It is concluded in the report (paragraph 52) that "given the current state of the art, it is not possible to calculate objective error margins for national accounts aggregates."

Another difficulty is that NA must comply with the restrictions of accounting systems. There are three approaches for calculating the GDP in the NA; the expenditure, the production, and the income approach. Usually, the estimates from the different approaches differ. Therefore, a post-adjustment ("balancing") of the estimates is necessary. The balancing is usually done within the supply-use framework, i.e., the supply and use of different commodities in different NACE (Statistical classification of economic activities in the European Community) industries. The balancing process is a highly resource-demanding activity of great importance for the compilation of NA. The balancing is typically done manually (at least partly), which further complicates any attempts to investigate the uncertainties of the balanced NA aggregates.

In early 1942, Sir Richard Stone and others (Stone, Champernowne, and Meade, 1942, hereafter SCM) proposed a generalized least square (GLS) approach to balance economic accounts in a limited scale. Other authors, in particular Byron (1978, 1996) formalized the approach and associated it with a Lagrange Multiplier approach with a quadratic loss function. A few applications (Van der Ploeg, 1982; Barker, van der Ploeg and Weale 1984) have been reported. More recently, Chen (2006, 2012) at the Bureau of Economic Analysis (BEA) used the SCM method to reconcile the US Industry Accounts and distribute the aggregate statistical discrepancy to industries. These efforts show that the method is feasible and empirically efficient, although it is still difficult to obtain objective estimates of uncertainties.

In our work, an SCM balancing approach is investigated in a large scale supply-use framework in the Swedish national accounts. Efforts are made to estimate the uncertainties, not only from sampling errors, but also from nonsampling errors. The resulting estimates are used as weights in the SCM approach. In Section 2, a framework of GLS and a workable Lagrange multiplier approach are described. The estimation of uncertainties of NA aggregates and test data for the SCM method are given in Section 3. The results are shown in Section 4. In Section 5, we discuss the results and make some suggestions for further work.

2. Framework

As in SCM and Byron (1996), write the estimates in NA to a vector $\boldsymbol{X}$ and the accounting restrictions as $\boldsymbol{AX} = 0$ (or constant C), where the matrix $\boldsymbol{A}$ consists of known constants. Assume that $\boldsymbol{X}^*$ is an initial, unbiased estimate of $\boldsymbol{X}$ with a covariance matrix $\boldsymbol{W}$, but does not satisfy the restriction. For the balanced estimate $\boldsymbol{X}^{**}$, using a quadratic loss function $(\boldsymbol{X}^{**} - \boldsymbol{X}^*)'\boldsymbol{W}^{-1}(\boldsymbol{X}^{**}-\boldsymbol{X}^*)$, the solution by the GLS approach is given by

$$\boldsymbol{X}^{**} = \boldsymbol{X}^* - \boldsymbol{WA}'(\boldsymbol{AWA}')^{-1}\boldsymbol{AX}^* \qquad (1)$$

Under the conditions that $\boldsymbol{W}$ is the correct covariance of $\boldsymbol{X}^*$, and $\boldsymbol{W}$ has full row rank, the variance-covariance of $\boldsymbol{X}^{**}$ is given by

$$Var(\boldsymbol{X}^{**}) = \boldsymbol{W} - \boldsymbol{WA}'(\boldsymbol{AWA}')^{-1}\boldsymbol{AW} \le \boldsymbol{W}. \qquad (2)$$

Note that the matrix $\boldsymbol{X}$ can be huge, possibly containing thousands of elements. The computing capacity required to handle a matrix this size might be one possible reason for the limited use of the SCM model in the past.

In practice, following Byron (1978) and Chen (2006), a more workable formula under the supply-use framework is used in our study. Consider the following notation. For $\mathrm{i} = 1, \cdots, \mathrm{N_i}; \mathrm{k} = 1, \cdots, \mathrm{N_k}; \mathrm{d} = 1, \cdots, 4$, denote

$\mathrm{x_{i,k}}$: The output of product k from industry i.

$\mathrm{z_{i,k}}$: The intermediate consumption of product k for industry i.

$\overline{\mathrm{x}}_\mathrm{i}$: The total output from industry i.

$\overline{\mathrm{z}}_\mathrm{i}$: The total intermediate consumption by industry i.

$y_{k,d}$: dth final use of product k (such as final household consumption, final government expenditure, gross capital formation, change of the inventory).

e_k: Export of product k.

i_k: Import of product k.

$\overline{y_d}$, $\bar{e}$, $\bar{\imath}$: the totals (over products) of final uses, export, and import, respectively.

The corresponding initial estimates are indicated with superscript 0 and w_0 are the corresponding uncertainties. The problem is thus to minimize

$$\min\left(\sum_i\sum_k\frac{\left(x_{i,k}-x_{i,k}^0\right)^2}{w_{x_{i,k}^0}}+\sum_i\sum_k\frac{\left(z_{i,k}-z_{i,k}^0\right)^2}{w_{z_{i,k}^0}}\right.$$

$$+\sum_i\frac{\left(\bar{x_\imath}-\overline{x_\imath^0}\right)^2}{w_{\overline{x_\imath^0}}}+\sum_i\frac{\left(\bar{z_\imath}-\overline{z_\imath^0}\right)^2}{w_{\overline{z_\imath^0}}}+\sum_k\sum_d\frac{\left(y_{k,d}-y_{k,d}^0\right)^2}{w_{y_{k,d}^0}}$$

$$\left.+\sum_d\frac{\left(\overline{y_d}-\overline{y_d^0}\right)^2}{w_{\overline{y_d^0}}}+\sum_k\frac{\left(e_k-e_k^0\right)^2}{w_{e_k^0}}+\sum_k\frac{\left(i_k-i_k^0\right)^2}{w_{i_k^0}}+\frac{\left(\bar{e}-\bar{e}^0\right)^2}{w_{\bar{e}^0}}+\frac{\left(\bar{\imath}-\bar{\imath}^0\right)^2}{w_{\bar{\imath}^0}}\right) \quad (3)$$

over $x_{i,k}$, $z_{i,k}$, $\bar{x_\imath}$, $\bar{z_\imath}$, $y_{k,d}$, e_k, i_k, $\overline{y_d}$, $\bar{e}$, $\bar{\imath}$, under the restriction that, for each industry i, the total output from the industry shall be equal to the sums of outputs for all products from this industry,

$$\sum_k x_{i,k}=\bar{x_\imath} \text{ for each industry i.} \quad (4)$$

Similar restrictions apply to the intermediate consumption, final uses, export and import, respectively:

$$\sum_k z_{i,k}=\bar{z_\imath} \text{ for each industry i}, \quad (5)$$

$$\sum_k y_{k,d}=\overline{y_d},\ \sum_k e_k=\bar{e},\ \sum_k i_k=\bar{\imath}. \quad (6)$$

An additional restriction is that the total supply is equal to the total use for each product k,

$$\left(1+\mathrm{t}_{\mathrm{k}}^{0}\right)\cdot\sum_{\mathrm{i}}\mathrm{x}_{\mathrm{i,k}}+\mathrm{i}_{\mathrm{k}}+\mathrm{o}_{\mathrm{k}}{}^{0}+\mathrm{m}_{\mathrm{k}}{}^{0}=\mathrm{e}_{\mathrm{k}}+\sum_{\mathrm{d}}\mathrm{y}_{\mathrm{k,d}}+\sum_{\mathrm{i}}\mathrm{z}_{\mathrm{i,k}}. \quad (7)$$

Remark 1: The initial estimates and their uncertainties in Eq. (3)-(7) can be expressed in both current and constant prices. In the case of constant prices, the uncertainties are induced from uncertainties in the values in current price and the deflating price indices.

Remark 2: It is not surprising to find that the greatest challenge we faced so far is to collect and calculate all the uncertainties of the initial estimates. There are still many subjective judgments. It can be of interest to study other weights in Eq. (3), such as all constant weights or the corresponding (squared) initial estimates.

Remark 3: Besides restrictions (4)-(7), it is possible to impose more restrictions, such as known value-added for each industry.

Remark 4: The coefficients $\mathrm{t}_{\mathrm{k}}^{0}$ in Eq. (7) are included to account for the taxes (such as customs, value-added taxes, minus subventions). They are assumed to be proportional to the total output for each product. Due to the complex definition and compilations of output from governments (incl. NPISH), $\mathrm{o}_{\mathrm{k}}{}^{0}$, and trade margins (including the third-part trading), $\mathrm{m}_{\mathrm{k}}{}^{0}$, they are included in Eq. (7) for the sake of the accounting, but are kept unchanged.

3. Application of the SCM balancing approach

3.1 The initial estimates $\boldsymbol{X}^*$

The SCM balancing approach is tested on the Swedish annual NA for year 2012 under a supply-use (SU) framework. The SU tables consist of the supplies and uses for each product group in each NACE 2-digits industries; see Eq. (7). There are 66 industries (N_i in Eq. 3) and 65 products (N_k) in our study. The initial estimates are obtained from a fixed earlier time point in the compilation. The total discrepancy between the supply and use for all products at this stage, 10,429 MSEK, is to be balanced.

3.2 Estimation of the uncertainties W

As mentioned before, it is a huge challenge to estimate the uncertainties of $\mathbf{X}^*$. In our work, we attempt to estimate both sampling and nonsampling errors and combine them. Table 1 illustrates our idea with an example of Gross Productions. In the table, the sampling error is the standard deviation from the Structure Business Survey, expressed as a percentage of the corresponding estimate of Gross productions for different industries. The nonsampling errors, as classified in Biemer et al (2014), are judgments made jointly by methodologist and subject-matter experts. The nonsampling errors are then added to the sampling error (see Calzaroni and Puggioni, 1998) to obtain the total uncertainties. In a similar manner the uncertainties are obtained for all necessary items in Eq. (3).

Tab. 1: The estimate of uncertainties in Gross Productions in different NACE industries and the total, in percentage to the corresponding estimates of Gross Productions.

NACE industry	**Sampling error**	**Specification error**	**Frame error**	**Non-response error**	**Measurement error**	**Data processing error**	**Model error**
NACE_A01	0.6	0.1	0.0	0.0	0.1	0.0	0.0
…							
NACE_G45	0.3	0.1	0.0	0.0	0.5	0.0	0.0
…							
NACE_T	3.0	0.1	0.0	0.0	0.1	0.0	0.0
Total	0.1	**0.0**	0.0	0.0	0.0	0.0	0.1

In our study, several different kinds of weights in Eq. (3) are tested, including constant weights (all $w_{\ \ 0} = 1$), neutral weights (using the square of an initial estimate as its weight, i.e., $w_{\ \ 0} = (\ \ ^{0})^2$), the sampling error, and the total uncertainties as discussed above.

3.3 Optimizing program

The Mathematical Programming Package SAS/OR® is used to perform the optimization (Eq. (3)–(7)). There are around 4,600 variables and the run time of the program is less than one second.

4. Results

Some summarized results are reported below and others are available upon request to the authors. Note again that for the initial estimates, the total discrepancy for all products (the total supply minus the total use) is 10,429 MSEK. The discrepancy for manufactory products is -31,984 MSEK and for service products 42,413 MSEK.

4.1 Balanced NA estimates X^{**}

With different kinds of weights, the adjustments between the balanced and the initial estimates (i.e., $X^{**} - X^{*}$) are shown in Table 2, which are divided between service and manufactory products. Table 3 shows the estimates of GDP before balancing, with the SCM approach, and the manual balancing. It is easy to see that different weights yield quite different results. Results from the manual balancing are included as a benchmark. The manual balancing is however not really comparable with the SCM approach since the start positions for the balancing procedure are not really the same. Therefore, no conclusion can be drawn about which approach outperforms the other.

Tab. 2: Adjustments done by the SCM approach with different kinds of weights to the initial NA aggregates, divided by Service (S) and Manufactory (M) products, and the manually balanced results.

		Intermed. Cons.	House-holds final Expenditure	Gross Capital Formation	Change of Inventory	Export	Gross production	Import	Taxes minus subventions
Constant weights	M	-19802	-4809	72	-306	245	5818	373	1192
	S	21481	5038	-53	330	1067	-13061	-400	-1089
Neutral Weights	M	-1010	-3491	-2860	-4	-8684	13607	970	1358
	S	3489	8479	1689	4	3556	-21738	-2682	-775
Sampling error	M	-15380	-1347	-11341	-3313	0	351	14	239
	S	15781	6876	5982	2515	2352	-6587	-1674	-646
Total Uncertainty	M	-4914	-10414	-11871	-4466	-525	-312	-50	157
	S	7021	14328	6347	4889	10704	2354	-1469	-9
Manual balancing	M	-32267	740	1060	-2163	-1524	-6895	51	4674
	S	41268	3658	-1831	444	3407	1362	792	2379

Tab 3: The estimates of GDP (MSEK) before balancing, from SCM balancing with different kinds of weights, and manual balancing.

GDP before balancing		Manual Balancing (Released)	Constant Weights	Neutral Weights	The Sampling Error	Total Uncertainty
Expenditure Approach	Production Approach					
3 681 852	3 692 281	3 684 800	3 683 462	3 682 254	3 685 236	3 692 363

4.2 Estimates of uncertainties

Table 4 shows the estimates of the total uncertainties in the Use table for different CPA (Statistical Classification of Products by Activity) products. The total of all products is estimated independently. Such estimates are not only used as input to the SCM approach, they can also be used to judge the data quality during the manual balancing and identify data sources whose quality needs to be improved.

Tab 4: An example of estimates of the total uncertainties in different NA aggregates (and the total) in the Use table, for different products (and the total)

Product	**Intermed. cons. by business %**	**Intermed. cons. by govern. %**	**Households final expenditure %**	**Govern. final cons. %**	**Gross capital formation %**	**Change of inventory**	**Export %**	**Total %**
CPA_A01-03	5.0	18.8	0.6	0.0	12.7	1 642	0.9	2.8
…								
CPA_G45_47	8.0	20.1	10.8	1.2	0.0	0	16.8	6.1
…								
CPA_R90T98	6.4	10.0	7.5	1.9	12.0	0	26.5	3.3
Total*	**0.1**	**1.0**	**0.6**	**0.3**	**2.1**	**8 260**	**0.4**	**0.3**

5. Discussion and final remarks

We have shown that an automatic SCM balancing approach is possible for the compilation of the Swedish NA. To fully replace a manual balancing process with an automatic balancing would probably not be desirable since the supply and use tables are a powerful tool to detect inconsistencies in source data. However, by integrating the SCM approach in the process, the balancing would not only become more objective, but also fully replicable. It is also likely to require much less resources and expertise than the manual balancing procedure in use today.

Several important challenges remain to be addressed before the SCM approach can be implemented in actual statistics production. One problem is that there is no obvious criterion for evaluating the approach, and for comparing it with the present procedure. Also, the estimation of the uncertainties in the initial NA estimates needs further investigation. For instance, it is still an open question if the nonsampling errors should be included. The literature does not, to our knowledge, provide much guidance in this case. Finally, in our study, we restrict our attention to the GDP expressed in current price. It is not trivial to extend the approach to the constant price. All these topics require further work.

6. References

Biemer, P. Trewin, D. Bergdahl, H. and Japec, L. (2014), A System for Managing the Quality of Official Statistics, *Journal of Official Statistics*, Vol. 30 (3), pp.381-415.

Byron, R.P. (1978), The Estimation of Large Social Account Matrices, *Journal of Royal Statistical Society, Series A*, Vol. 141 (3), pp.359-367.

Byron, R.P. (1996), Diagnostic Testing and Sensitivity Analysis in the Construction of Social Accounting Matrices, *Journal of Royal Statistical Society, Series A*, Vol. 159 (1), pp.133-148.

Calzaroni, M. and Puggioni, A. (1998), Evaluation and Analysis of the Quality of the National Accounts Aggregates, Report to the Task Force on Accuracy Assessment of National Accounts, Eurostat.

Chen, B. (2006), A Balanced System of Industry Accounts and Structural Distribution of Aggregate Statistical Discrepancy, Working paper WP2006-8, Bureau of Economic Analysis, Washington, DC.

Chen, B. (2012), A Balanced System of U.S. Industry Accounts and Distribution of Aggregate Statistical Discrepancy by Industry, *Journal of Business and Economic Statistics*, Vol. 30 (2), pp.202-211.

Eurostat (2001), Task Force on Accuracy Assessment of National accounts Statistics-Final Report. Eurostat/B1/CPNB/296.

Stone, R., Champernowne, D.G. and Meade, J.E. (1942), The precision of national income estimates, *The Review of Economic Studies,* Vol. 9, pp.111 – 135.

Van der Ploeg, F. (1982), Reliability and the Adjustment of Sequences of Large Systems and Tables of National Accounts Matrices, *Journal of Royal Statistical Society, Series A*, Vol. 145 (2), pp.169-194.