Working Party on National Accounts

UNCERTAINTY AND AUTOMATIC BALANCING OF NATIONAL ACCOUNTS

To be held on 27-28 October 2016
OECD Conference Centre
Beginning at 4:10 pm on the first day

This document has been prepared by Yingfu Xie, Andreas Lennmalm, Daniel Lennartsso and Annica de Groote – Statistics Sweden and will be presented under item 13 of the draft agenda

JT03403313

Complete document available on OLIS in its original format
This document and any map included herein are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.
Uncertainty and Automatic Balancing of National Accounts

Yingfu Xie¹, Andreas Lennmalm², Daniel Lennartsson³, Annica de Groote⁴

Abstract

The uncertainty of the national accounts (NA) such like the gross domestic product (GDP), is of great interest for decision-makers, researchers, and the public. This information is nevertheless often absent in statistic releases. It is partly due to the complexity of the compilation of NA with its enormous data sources. It is difficult to estimate all uncertainties of the initial estimates. Furthermore, national accounts must comply with the restrictions of accounting systems, for example the supply and use of the different products in the economy have to be balanced. The estimate of uncertainties of NA has to take the 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 difficulty 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 from not only sampling errors but also non-sampling errors. Those estimates are used as the weights in the balancing. An automatic balancing procedure is thus made possible, compared with the current, manually, manual balancing procedure which is both time-demanding and requires expertise of NA. Our study illustrates also a lot of challenges and possible future works in this area.

Key words: National Accounts, Balancing, SCM method, Non-sampling error

1. Introduction

The uncertainty of the national accounts (NA) such like the gross domestic product (GDP), is of great interest for decision-makers, researchers, and the public. This information is nevertheless often absent in statistic releases. It is partly due to the complexity of the compilation of NA with its enormous data sources. It is difficult to estimate all uncertainties of the initial estimates. In a report from Eurostat (2001) the data sources and possible error sources of NA compilation are well documented, and the difficulties related to identifying and quantifying the errors, in particular the non-sampling errors are discussed. The report (paragraph 52) concludes that “given the current state of the art, it is not possible to calculate objective error margins for national accounts aggregates.”

¹ Statistics Sweden, Senior Methodologist, Stockholm, Sweden, Email: Yingfu.Xie@scb.se
² Statistics Sweden, Senior Advisor, Stockholm, Sweden, Email: Andreas.Lennmalm@scb.se
³ Statistics Sweden, Expert, Stockholm, Sweden, Email: Daniel.Lennartsson@scb.se
⁴ Statistics Sweden, Senior Methodologist, Stockholm, Sweden, Email: Annica.Isaksson@scb.se
Furthermore, NA must comply with the restrictions of accounting systems. There are three approaches for calculating GDP in the NA: the expenditure, the production, and the income approach. Usually the estimates from the different approaches differ. Therefore, a post-adjustment (“balance”) of those 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 important for the compilation of NA. However this process is not only highly demanding on expertise and time-consuming, but also in most cases manual, which complicates further eventual attempts to investigate the uncertainties of the balanced NA aggregates.

In 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 to Lagrange Multiplier approach with a quadratic loss function. A few of applications (Van der Ploeg, 1982; Barker, van der Ploeg and Weale 1984 among others) have been reported. Recently Chen (2006, 2012) used the SCM method to reconcile the US Industry Accounts and distribute the aggregate statistical discrepancy to industries in the Bureau of Economic Analysis (BEA). Those works show that the method is feasible and empirically efficient, although it is still difficult to obtain objective estimates to 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 from not only sampling errors but also non-sampling errors. Those estimates are used as the 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 will be shown in Section 4. Discussions to the results and future work are in Section 5.

2. The framework

Following the description of Stone, Champernowne, and Meade (1942) and Byron (1996), write the estimates in NA to a vector $X$ and the accounting restrictions as $AX = 0$ (or constant $C$), where the matrix $A$ consists of known constants. Assume $X^*$ is an initial, unbiased estimate of $X$ with a covariance matrix $W$, but does not satisfy the restriction. For the balanced estimate $X^{**}$, using a quadratic loss function $(X^{**} - X^*)'W^{-1}(X^{**} - X^*)$, the solution by GLS approach is given by

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

(1)

Furthermore, under the conditions that $W$ is the correct covariance of $X^*$, $W$ has full row rank, variance-covariance of $X^{**}$ can be induced as

$$\text{Var}(X^{**}) = W - WA'(AWA')^{-1}AW \leq W.$$  

(2)

Note that the matrix $X$ 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 (SU) framework is used in our study. Consider the following notation. For $i = 1, \cdots, N_i; k = 1, \cdots, N_k; d = 1, \cdots, 4$, denote

- $x_{i,k}$: The output of product $k$ from industry $i$.
- $z_{i,k}$: The intermediate consumption of product $k$ for industry $i$.
- $\bar{x}_i$: The total output from industry $i$.  


The total intermediate consumption by industry $i$. 

$y_{kd}$: 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 $k$.

$\bar{y}_d$, $\bar{e}$, $\bar{t}$: 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

$$
\min \left( \sum_i \sum_k \left( \frac{x_{i,k} - x_{i,k}^0}{w_{x_{i,k}}^0} \right)^2 + \sum_i \sum_k \left( \frac{z_{i,k} - z_{i,k}^0}{w_{z_{i,k}}^0} \right)^2 
\right) + \sum_i \left( \frac{\bar{y}_i - \bar{y}_i^0}{w_{\bar{y}_i}} \right)^2 + \sum_i \left( \frac{\bar{z}_i - \bar{z}_i^0}{w_{\bar{z}_i}} \right)^2 + \sum_k \sum_d \left( \frac{y_{kd} - y_{kd}^0}{w_{y_{kd}}^0} \right)^2
$$

$$
+ \sum_d \left( \frac{y_d - y_d^0}{w_{y_d}} \right)^2 + \sum_k \left( \frac{e_k - e_k^0}{w_{e_k}} \right)^2 + \sum_k \left( \frac{i_k - i_k^0}{w_{i_k}} \right)^2 + \left( \frac{t_k - t_k^0}{w_{t_k}} \right)^2 + \left( \frac{\bar{t}_k - \bar{t}_k^0}{w_{\bar{t}_k}} \right)^2
$$

(3)

over $x_{i,k}$, $z_{i,k}$, $\bar{y}_i$, $\bar{z}_i$, $y_{kd}$, $e_k$, $i_k$, $\bar{y}_d$, $\bar{e}$, $\bar{t}$, 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,

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

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

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

$$\sum_k y_{kd} = \bar{y}_d. \quad \sum_k e_k = \bar{e}, \quad \sum_k i_k = \bar{t}. \quad (6)$$

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

$$(1 + t_k^0) \cdot \sum_i x_{i,k} + i_k + \alpha_k^0 + m_{k}^0 = e_k + \sum_d y_{kd} + \sum_i z_{i,k}. \quad (7)$$

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 $t_k^0$ 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), \( o_k^0 \), and trade margins (including the third-part trading), \( m_k^0 \), they are included in Eq. (7) for the sake of the accounting, but are kept unchanged.

3. The application of the SCM balancing approach

3.1. The initial estimates \( X^* \)

The SCM balancing approach is tested on the Swedish annual NA for year 2012 under a 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 (\( N_l \) 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. The estimation of the uncertainties \( W \)

As aforementioned, it is a huge challenge to estimate the uncertainties of \( X^* \). In our work, attempts are made not only to obtain the sampling error, but also to estimate the non-sampling errors, and combine them. Table 1 below illustrates our idea with an example of Gross Productions. In Table 1, the sampling error is the standard deviation from the Structure Business Survey, expressed as percentage to the corresponding estimate of Gross productions for different industries. The non-sampling errors, whose classification follows Biemer et al (2014), are judgments made jointly by methodologist and subject expertise. The non-sampling 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 Gross Productions, in percentage to the corresponding estimates of Gross Productions.

<table>
<thead>
<tr>
<th>NACE industry</th>
<th>Sampling error</th>
<th>Specification error</th>
<th>Frame error</th>
<th>Non-response error</th>
<th>Measurement error</th>
<th>Data processing error</th>
<th>Model error</th>
</tr>
</thead>
<tbody>
<tr>
<td>NACE_A01</td>
<td>0.6</td>
<td>0.1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.1</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NACE_G45</td>
<td>0.3</td>
<td>0.1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.5</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NACE_T</td>
<td>3.0</td>
<td>0.1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.1</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Total Gross Production</td>
<td>0.1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.1</td>
</tr>
</tbody>
</table>

In our study, several different kinds of weights in Eq. (3) are tested, including Constant weights (all \( w \circ = 1 \)), neutral weights (using the square of an initial estimate as its weight, i.e., \( w \circ = (x_0)^2 \)), the sampling error, and the total uncertainties as discussed above.

3.3. The optimizing program

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. The 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, whereof for manufactory products it is -31,984 MSEK and for service products 42,413 MSEK.

4.1. The 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 to service and manufactory products respectively. Table 3 shows the estimates of GDP before balancing, with 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. Observe nevertheless that the manual balancing is not really comparable with the SCM approach since the start positions for the balancing procedure are not really the same. No conclusion can be drawn about which approach outperforms others.

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

<table>
<thead>
<tr>
<th></th>
<th>Intermed. Cons.</th>
<th>House-holds final Expenditure</th>
<th>Gross Capital Formation</th>
<th>Change of Inventory</th>
<th>Export</th>
<th>Gross production</th>
<th>Import</th>
<th>Taxes minus subsidies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant weights</td>
<td>M -19802</td>
<td>-4809</td>
<td>72</td>
<td>-306</td>
<td>245</td>
<td>5818</td>
<td>373</td>
<td>1192</td>
</tr>
<tr>
<td></td>
<td>S 21481</td>
<td>5038</td>
<td>-53</td>
<td>330</td>
<td>1067</td>
<td>-13061</td>
<td>-400</td>
<td>-1089</td>
</tr>
<tr>
<td>Neutral weights</td>
<td>M -1010</td>
<td>-3491</td>
<td>-2860</td>
<td>-4</td>
<td>-8684</td>
<td>13607</td>
<td>970</td>
<td>1358</td>
</tr>
<tr>
<td></td>
<td>S 3489</td>
<td>8479</td>
<td>1689</td>
<td>4</td>
<td>3556</td>
<td>-21738</td>
<td>-2682</td>
<td>-775</td>
</tr>
<tr>
<td>Sampling error</td>
<td>M -15380</td>
<td>-1347</td>
<td>-11341</td>
<td>-3313</td>
<td>0</td>
<td>351</td>
<td>14</td>
<td>239</td>
</tr>
<tr>
<td></td>
<td>S 15781</td>
<td>6876</td>
<td>5982</td>
<td>2515</td>
<td>2352</td>
<td>-6587</td>
<td>-1674</td>
<td>-646</td>
</tr>
<tr>
<td>Total uncertainty</td>
<td>M -4914</td>
<td>-10414</td>
<td>-11871</td>
<td>-4466</td>
<td>-525</td>
<td>-312</td>
<td>-50</td>
<td>157</td>
</tr>
<tr>
<td>Manual balancing</td>
<td>S 7021</td>
<td>14328</td>
<td>6347</td>
<td>4889</td>
<td>10704</td>
<td>2354</td>
<td>-1469</td>
<td>-9</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Expenditure Approach</th>
<th>Production Approach</th>
<th>Manual Balancing (Released)</th>
<th>Constant Weights</th>
<th>Neutral Weights</th>
<th>The Sampling Error</th>
<th>Total Uncertainty</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 681 852</td>
<td>3 692 281</td>
<td>3 684 800</td>
<td>3 683 462</td>
<td>3 682 254</td>
<td>3 685 236</td>
<td>3 692 363</td>
</tr>
</tbody>
</table>
4.2. The estimate 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, these can also be used to judge the data quality during the manual balancing and trade off the possible data sources whose quality have to be improved.

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

<table>
<thead>
<tr>
<th>Product</th>
<th>Intermed. cons. by industries, %</th>
<th>Intermed. cons. by govern., %</th>
<th>Households final expenditure, %</th>
<th>Govern. final cons., %</th>
<th>Gross capital formation, %</th>
<th>Change of inventory, %</th>
<th>Export, %</th>
<th>Total Use, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPA_A 01-03</td>
<td>5.0</td>
<td>18.8</td>
<td>0.6</td>
<td>0.0</td>
<td>12.7</td>
<td>1642</td>
<td>0.9</td>
<td>2.8</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPA_G 45_47</td>
<td>8.0</td>
<td>20.1</td>
<td>10.8</td>
<td>1.2</td>
<td>0.0</td>
<td>0</td>
<td>16.8</td>
<td>6.1</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPA_R9 0T98</td>
<td>6.4</td>
<td>10.0</td>
<td>7.5</td>
<td>1.9</td>
<td>12.0</td>
<td>0</td>
<td>26.5</td>
<td>3.3</td>
</tr>
<tr>
<td>Total product s(^5)</td>
<td>0.1</td>
<td>1.0</td>
<td>0.6</td>
<td>0.3</td>
<td>2.1</td>
<td>8260</td>
<td>0.4</td>
<td>0.3</td>
</tr>
</tbody>
</table>

5. Discussions and final remarks

It is shown in our study that an automatic SCM balancing approach is possible for the compilation of NA. With the SCM approach, the balancing is not only more objective, but also fully replicable. Furthermore, the manual balancing procedure existing today requires much more resources and expertise.

In able to implement the SCM approach in the official statistics production, more experiments have to be carried out and evaluated since there is no obvious evaluation criterion, in particular compare with the manual procedure. It is a big challenge how to estimate the uncertainties in the initial NA estimates. It is an open question if the non-sampling errors should be included. There is no work available in the literature concerning the quantification of the non-sampling errors and how to combine them with the sampling error. In our study, the application is done on the current price. It is not trivial to extend the approach to the constant price. The possibility to compute the uncertainty of the balanced aggregates, such like the GDP,

---

\(^5\) The estimate of the uncertainty for the total of products is carried separately to that for individual products since better data source is available for the total.
and the theoretical property (2) are of great interest. However the conditions are very difficult to check. All those are possible topics for our future work.

References


