This module is part of the

Memobust Handbook

on Methodology of Modern Business Statistics

26 March 2014

Theme: Macro-Integration – Main Module

Contents

General section

1. Summary

Macro-integration is the process to integrate data from different sources on an aggregate level, to enable a coherent analysis of the data. When there are two or even more independent data sources, inconsistencies will inevitably occur. Discrepancies are caused by various kind of errors, like: sampling error, nonresponse error, coverage error, measurement error and processing error (Eurostat, 2009). Macro-integration can be divided in two stages. In the first stage the source data are adapted to comply with the correct definitions. The second part is called data reconciliation. This is the process in which the remaining discrepancies are resolved or at least reduced at the aggregated level. Data reconciliation is often called balancing in the literature. As mentioned in United Nations (1999, page 23) and Rassier et al. (2007) data reconciliation improves the accuracy.

Data reconciliation can be divided into adjusting for major errors and for the remaining (sampling) noise. Large errors are often corrected manually, by using subject matter knowledge. As correction methods for (large) errors are difficult to formalise, these are not covered well in the literature. A reference for such methods is Bloem et al. (2001, Chapter V). Manual integration is discussed in a theme module in this handbook (“Macro-Integration – Manual Integration”).

Working with statistical data based on samples and questionnaires and influenced by non-response etc. means working with margins of error. Even when samples are perfect and response is 100% there will be inconsistencies. The cause is then a statistical one. In such a case balancing could be done automatically. Compared to methods for the correction of large errors, methods for the correction of noise are much more explored in the literature. In this handbook, there are a modules on four of these methods: “Macro-Integration – RAS”, “Macro-Integration – Stone’s Method”, “Macro-Integration – Denton’s Method”, and “Macro-Integration – Chow-Lin Method for Temporal Disaggregation”.

2. General description

2.1 Aim of macro-integration

Macro-integration has two objectives. The first is to facilitate analysis of the interrelationships in data by organising information into an accounting framework. The second is to make more accurate estimates of reality through the reconciliation of the various statistical information contained in a framework of this kind. An accounting framework is defined by a set of variables and a set of relationships between them, so-called accounting rules. These accounting rules are relationships that should hold exactly and that are based on definitions. Examples of accounting rules are:

- Total is equal to the sum of components: For instance: Manufacturing = Food + Textiles + Clothing;
- Commodity balances; For instance, relationship between supply and use of goods and services. Normally, total use has to be the same as total supply;
- Annual alignment: Annual values have to be equal to the sum of four quarters;
- Definition of specific terms: For instance year: Value added = Output – Intermediate consumption

See also Bloem (2001, chapter V).

The data generally come from a wide variety of sources. Some variables may be obtained from external sources, or through sampling, but, where no suitable source exists, variables may be based on model estimates, or ‘expert guesses’.

Usually, the data that are collected by statistical offices are not consistent with the accounting rules. This happens, for example, because data are collected by different methods, using different sample surveys and different data processing methods and because of estimation error in case of missing data.

Macro-integration first ‘translates’ the source data to comply with the correct definitions and then identifies and adjusts for major measurement errors. After the major errors are corrected (the so-called bias), the remaining, usually smaller, discrepancies have to be solved (the so-called noise). These small discrepancies appear more or less by accident, for instance due to sampling errors.

Although macro-integration is often applied to economic data and in particular to the national accounts it may also be applied to other areas, where data are stored in some consistent framework.

National Statistical Institutes (NSIs) have often applied informal methods for macro-integration. These methods heavily depend on mutual agreement of different subject matter experts on the necessary adjustments to the data. Although these informal methods work well in practice there are also some drawbacks. Informal methods are not transparent and therefore irreproducible. Further, the process of achieving consistency is often time-consuming for large data sets.

As an alternative to informal methods, formal methods can be used. In the literature a lot of formal methods are described. Some statistical offices have adopted these formal methods. In this module formal macro-integration methods are discussed, that are aimed at the second step of macro-integration, i.e., for correcting the noise, the small discrepancies that cannot be attributed to measurement errors or other sources of bias. When applying a formal macro-integration method, it is important to realise that “garbage in” means garbage out”, the input data has to be free of large errors, otherwise these errors will remain in the output. In practice it is hard to differentiate between large errors and noise. It is usually necessary to resort to elimination by hand of the largest differences, and to distribute the mass of smaller discrepancies through modelling.

2.2 Related process steps

In this module much attention is paid to macro-integration methods for the reconciliation of data that cover more than one time period and that comprise time periods of different length. An example of this is the integration of quarterly and annual data, at an aggregated level. This particular macro-integration problems plays an important role in the production process of statistical offices, since national accounts are published, both on a quarterly basis and on an annual basis.

As mentioned in Bloem et al. (2001) “To avoid confusion about interpreting economic developments, it is imperative that quarterly national accounts are consistent with annual accounts. Differences in growth rates quarterly and annual accounts would perplex users and cause uncertainty about the actual situation. Concerning the level of the data, this means that the sums of the estimates for the four quarters of the year should be equal to the annual estimates. In a situation where the annual accounts are built up from the quarterly accounts, this is more or less self-evident. However, more commonly, the annual accounts are based on different sources than the quarterly estimates, and if that is the case, differences could develop. To avoid this, the quarterly national account data should be aligned with the annual data”.

The low frequent data are often called annual data, while the high frequent data are often referred to as sub-annual data. These two terms will also be used henceforth.

In the area of macro-integration problems with a temporal component two problems are often mentioned: benchmarking and temporal disaggregation. Benchmarking is the process to achieve consistency between sub-annual and annual data and temporal disaggregation is the process of deriving sub-annual data (for instance quarterly data) from annual data, possibly by using indicators of the sub-annual data (i.e., related time series).

Although benchmarking and temporal disaggregation differ from a conceptual point of view, both process steps are closely related from a methodological point of view. The same methods can be used for both problems. The differences between benchmarking and temporal aggregation are:

1) Benchmarking assumes the same definitions for the annual and the sub-annual data, while these definitions may differ for temporal disaggregation;
2) In the problem of benchmarking the sub-annual data are already available, but for temporal disaggregation only indicators may (or may not) be available;
3) For temporal disaggregation more than one indicator time series may be used for the disaggregation of one time series, while in case of benchmarking always one sub-annual time series is aligned to one sub-annual time series.

Temporal disaggregation can be further divided into

- *Temporal distribution* which deals with flow variables (variables that are measured over an interval of time)
- *Interpolation*, which is the estimation of missing values of a stock variable (variables that are measured at one point of time).

2.3 Related methods

There is a clear distinction between manual and mathematical methods. The first class of methods are described in the module “Macro-Integration – Manual Integration”. The latter class of methods are described in four modules: “Macro-Integration – RAS”, “Macro-Integration – Stone’s Method”, “Macro-Integration – Denton’s Method” and “Macro-Integration – Chow-Lin Method for Temporal Disaggregation”. The motivation of this choice is that these four methods are well-known, computationally easy, and suitable for large-scale application. References will be given to other macro-integration methods.

Manual integration

In balancing the detection of inconsistencies and implausibilities on the one hand and finding the causes on the other is most important part of the job. Having this knowledge, finding a way to resolve the inconsistency is then mostly straightforward.

Inconsistencies can be detected by observing violations of the identities of the accounting framework. For instance: a discrepancy between the total use and total supply of some good. In addition, less strict relations between the variables of the system can be used. For example, the volume change of the production of cheese should not differ too much from the volume change of the (intermediate) input of milk.

Once the inconsistencies are identified, the next step is to reveal its cause. This can be done by using independent secondary information. Two examples of this:

- The sales of motorcars can be confronted with the number of newly registered number plates.
- The consumption of liquor can be compared with tax revenues on liquor from the government administration.

Additionally, expert knowledge can be used on the reliability of the data sources. An example of this is the known underestimation of self-reported alcohol consumption.

In the last step the data are corrected for errors. This has to be done such that the data obey the consistency constraints of the accounting framework. For more details we refer to the theme module "Macro-Integration – Manual Integration".

Model based integration applied to data of one time period

The literature refers to various formal macro-integration methods, each with its own origins. There is a correspondingly great variety in applicability, interpretability and generality.

The simplest methods were devised at a time before powerful computers were widely available. An example is the *RAS method*, a numerical method that allows the entries of a rectangular matrix to be aligned with a set of row and column totals. For this specific problem other methods may also be applied, for an overview we refer to Lahr and De Mesnard (2004) and Lenzen et al. (2009).

The RAS method is the easiest method to apply, small problems can even be solved without a computer. However its field of application is more narrow than for the other methods.

There are more general methods, with a better statistical foundation and a broader scope of applicability, that estimate reconciled results from source data, while complying with certain constraints, in accordance with a specific procedure. The best statistical estimate corresponds with the optimum value of some objective function. Different additional assumptions give rise to specific model variants.

Many of the common methods can be classified as *generalised least-squares methods*, which belong to the class of quadratic optimisation methods.

Quadratic optimisation is the most simple form of optimisation, after linear optimisation. This kind of problem has been extensively studied and many computational efficient algorithms are available. These algorithms were developed in a time when computers were less powerful than today. Hence, quadratic optimisation might have been the only choice for practical applications. Nowadays, more powerful computers are available and alternatives for quadratic optimisation can be used. Quadratic optimisation is however still an attractive choice, due to its relative simplicity and the availability of efficient solving methods that can be applied to large problems.

The assumption of quadratic optimisation with linear, equality constraints that should hold exactly (without an error term) leads to the method of *Stone*, which is one of the older (1942) and most rudimentary of the least-squares methods.

In comparison with RAS, the Stone method is more general: it does not require input data that can be represented in one matrix and it allows for more general type of constraints than the alignment to pre-specified row and column totals.

The basic version of Stone's method can be expanded by

- non-linear constraints, e.g., ratio constraints, that say that the proportion between two variables should be equal to some pre-specified value;
- nonbinding (or soft) constraints, i.e., constraints that should not hold exactly, but only by approximation;
- inequality constraints, of which the non-negativity constraint is the most well-known.

These extensions are worked out by Magnus et al. (2000) for a Bayesian approach reconciliation model. By using the above mentioned features elaborate model constructions are possible. Amongst others, expert knowledge may be incorporated into a model. The expert knowledge may be based on observations on the past, related variables or other expert knowledge. Examples of expert knowledge are:

- For some perishable goods the value of the change of stocks, summed over the four quarters of one year, should not differ much from zero.
- The production of certain industries should not deviate too much from the production of last year.
- The output per employee should be between some lower and upper bound.

These requirements can be translated into soft constraints in a reconciliation model. An advantage of incorporating expert knowledge into a reconciliation model is that maximum use is made of all information available. A disadvantage is that the subject matter knowledge cannot be used to monitor the results, when that information has already been used in the reconciliation model.

Although reconciliation models with a quadratic objective function are often used, other forms of objective functions can be used as well. For instance, the cross-entropy approach of Robinson et al. (2000) that minimises the logarithmic Kullback-Leibler cross-entropy measure. The cross entropy method leads to different results than the quadratic optimisation procedures.

Model based integration with a temporal component: Benchmarking

For an overview of benchmarking methods in the literature we refer to Bloem et al. (2001) and Dagum and Cholette (2006). Here, we describe a well-known method for benchmarking: the *Denton method*.

In comparison with the Stone method, the Denton method is more specific. The Denton method is meant for achieving consistency between data of different frequencies (for instance quarterly data that has to comply with annual data), while the Stone method does not include a time component.

In achieving consistency, the Denton method assumes that the annual data sources are of high precision and provide reliable information on the overall levels. On the other hand, the sub-annual data are less precise, but they provide the only information on the short-term movements. The Denton method combines the (assumed) strengths of both types of data. This method adjusts the sub-annual time series to align with the annual time series, while preserving as much as possible the trend of the sub-annual data. Originally, the Denton method is described for univariate data, Denton (1971). However, there are extensions in the literature, for multivariate data.

Model based integration with a temporal component: Temporal disaggregation

An overview of temporal disaggregation methods is given by Chen (2007). A distinction can be made for methods that require the availability of indicator time series on the sub-annual level and methods that can be applied in absence of sub-annual indicator series.

When sub-annual data are not available smoothing methods can be used, for instance Cubic Splines, or the Boot, Feibes and Lisman smoothing method, see for instance Boot et al. (1967) and Wei and Stram (1990). For the case when indicator series on the sub-annual are available, the Chow-Lin regression method and its variants are often used.

The Chow-Lin method derives high-frequency data from low-frequency data, by using low-frequency indicators. These indicators are time series that are related to the target time series and thus measure a different topic than the time series to be estimated. Chow and Lin do not discuss the nature of the related indicators. Presumably, the indicators should be socio-economic variables deemed to behave like the target variable. In absence of such variables, they suggest using functions of time (Dagum and Cholette, 2006).

3. Design issues

4. Available software tools

5. Decision tree of methods

6. Glossary

For definitions of terms used in this module, please refer to the separate "Glossary" provided as part of the handbook.

7. References

Bloem, A. M., Dippelsman, R. J., and Maehle, O. N. (2001), *Quarterly National Accounts Manual: Concepts, Data Sources, and Compilation*. International Monetary Fund, Washington, DC. http://www.imf.org/external/pubs/ft/qna/2000/Textbook/index.htm

Boot, J. C. G., Feibes, W., and Lisman, J. H. C. (1967), Further methods of derivation of quarterly figures from annual data. *Cahiers Economiques de Bruxelles* **36**, 539–546.

Chen, B. (2007), An Empirical Comparison of Methods for Temporal Distribution and Interpolation at the National Accounts. Bureau of Economic Analysis.

Chow, G. and Lin, A. L. (1971), Best linear unbiased interpolation, distribution and extrapolation of time series by related series. *The Review of Economics and Statistics* **53**, 372–375.

Cholette, P. A. and Chhab, N. B. (1991), Converting Aggregates of Weekly Data into Monthly Values. *Journal of the Royal Statistical Society, Series C* **40**, 411–422.

Dagum, E. B. and Cholette, P. A. (2006), *Benchmarking, Temporal Distribution and Reconciliation Methods for Time Series*. Springer, New York.

Denton, F. T. (1971), Adjustment of monthly or quarterly series to annual totals: an approach based on quadratic minimization. *Journal of the American Statistical Association* **66**, 99–102.

Eurostat (2009), *ESS Handbook for Quality Reports*. Eurostat Methodologies and Working Papers, Luxemburg.

Lahr, M. L. and de Mesnard, L. (2004), Biproportional Techniques in Input-Output Analysis: Table Updating and Structural Analysis. *Economic Systems Research* **16**, 115–134.

Lenzen, M., Gallego, B., and Wood, R. (2009), Matrix balancing under conflicting information. *Economic Systems Research* **21**, 23–44.

Magnus, J. R., van Tongeren, J. W., and de Vos, A. F. (2000), National Accounts Estimation Using Indicators Analysis. *Review of Income and Wealth* **46**, 329–350.

Rassier, D. G., Howells, T. F., Morgan, E. T., Empey, N. R., and Roesch, C. E. (2007), Implementing a Reconciliation and Balancing Model in the U.S. Industry Accounts.

Robinson, S., Cattaneo, A., and El-Said, M. (2000), Updating and Estimating a Social Accounting Matrix Using Cross Entropy Methods. IFPRI, Discussion Paper 64.

United Nations (1999), *Handbook of National Accounting: a systems approach to national accounts compilation*. Technical Report, ST/ESA/STAT/SER.F/77, United Nations Statistics Division, New York. http://unstats.un.org/unsd/publication/SeriesF/SeriesF_77E.pdf

Wei, W. W. S. and Stram, D. O. (1990), Disaggregation of time series models. *Journal of the Royal Statistical Society* **52**, 453–467.

Interconnections with other modules

8. Related themes described in other modules

1. Macro-Integration – Manual Integration

9. Methods explicitly referred to in this module

1. Macro-Integration – RAS
2. Macro-Integration – Stone's Method
3. Macro-Integration – Denton's Method
4. Macro-Integration – Chow-Lin Method for Temporal Disaggregation

10. Mathematical techniques explicitly referred to in this module

1. Quadratic optimisation under linear constraints
2. Interpolation
3. Extrapolation
4. Linear regression

11. GSBPM phases explicitly referred to in this module

1. GSBPM Phase 6.2

12. Tools explicitly referred to in this module

1.

13. Process steps explicitly referred to in this module

1. Data Reconciliation
2. Benchmarking
3. Temporal Disaggregation
4. Temporal Distribution

Administrative section

14. Module code

Macro-Integration-T-Main Module

15. Version history

Version	Date	Description of changes	Author	Institute
0.1	31-03-2011	first version	Jacco Daalmans	CBS
0.2	20-01-2012	second version	Jacco Daalmans	CBS
0.3	21-06-2013	third version	Jacco Daalmans	CBS
0.3.1	06-09-2013	preliminary release		
0.3.2	09-09-2013	page numbering adjusted		
0.4	12-03-2014	minor revisions	Jacco Daalmans	CBS
1.0	26-03-2014	final version within the Memobust project		

16. Template version and print date

Template version used	1.0 p 4 d.d. 22-11-2012
Print date	26-3-2014 13:24