

Article

Symposium 2008:
Data Collection: Challenges, Achievements and New Directions

Coherence Analysis - Improved Quality of National Accounts

by Anna-Greta Erikson

2009



Coherence Analysis - Improved Quality of National Accounts

Anna-Greta Erikson¹

Abstract

In many countries, improved quality of economic statistics is one of the most important goals of the 21st century. First and foremost, the quality of National Accounts is in focus, regarding both annual and quarterly accounts. To achieve this goal, data quality regarding the largest enterprises is of vital importance. To assure that the quality of data for the largest enterprises is good, coherence analysis is an important tool. Coherence means that data from different sources fit together and give a consistent view of the development within these enterprises. Working with coherence analysis in an efficient way is normally a work-intensive task consisting mainly of collecting data from different sources and comparing them in a structured manner. Over the last two years, Statistics Sweden has made great progress in improving the routines for coherence analysis. An IT tool that collects data for the largest enterprises from a large number of sources and presents it in a structured and logical matter has been built, and a systematic approach to analyse data for National Accounts on a quarterly basis has been developed. The paper describes the work in both these areas and gives an overview of the IT tool and the agreed routines.

Key Words: Coherence, Higher quality, Quality assurance.

1. Introduction

In several countries, specific resources have been allocated to treat the largest enterprises, or the key providers of statistics. As a part of a systematic approach of improving the quality of National Accounts, a unit of Large Enterprise Managers was established at Statistics Sweden in 2004. The main task of the unit is to manage the key providers of economic statistics, i.e. the largest enterprises in Sweden, to achieve better coherence and quality in data from these main providers of economic statistics. This paper will focus on the approach of improving the quality of data concerning the largest enterprises, by developing the work with coherence analysis in the scope of the Large Enterprise Management programme at Statistics Sweden.

2. Large Enterprise Management programme

The most important objective of establishing a group of Large Enterprise Managers at Statistics Sweden was to improve the quality of statistics in the system of National Accounts. The managers were given the responsibility to improve the management of the largest enterprises in Sweden, as the quality in the system of National Accounts is dependent on the quality and coherence of data from these key providers.

2.1 The key providers of data

In Sweden, the focus of the Large Enterprise Managers is the 50 largest enterprises. The enterprises are identified by looking at main economic indicators, i.e. turnover/net sales, value added, employment, production value, balance sheet total. In 2005 the 50 largest enterprises accounted for almost 30 per cent of non-financial enterprises contribution to Gross Domestic Product. The objective is to be able to concentrate the work and effort on the most important data providers of the system of economic statistics. Even if there are no current problems in the data collection process of these enterprises, managers will continuously monitor them. If or when, a problem occurs in

¹Anna-Greta Erikson, Statistics Sweden, DFO/SF, SE-701 89 ÖREBRO, Sweden (annagreta.erikson@scb.se)

data collection or data provision from these businesses, there is a risk that there will be a large impact on National Accounts.

2.2 The work of the Large Enterprise Managers

We have chosen a broad scope of the work for the Large Enterprise Managers. With the broad scope of work, the Large Enterprise Managers really get to know the enterprises they are responsible for. They get an overall view and the broad perspective of the enterprises. When it comes to coherence analysis, managers are already familiar with the data collected and the problems from the data collection process. The managers also learn to recognise the patterns and similar phenomena between the enterprises they manage.

2.2.1 Building relationships

One of the most important tasks of the managers is to build a relationship with the largest enterprises and, more specifically, to establish a dialogue with the enterprises in different areas, such as human resources and accounting. In this process, it is important to visit the enterprises and to present the work of the Large Enterprise Managers and to point out the significance of the enterprises' data for the statistics. This has the added consequence of reducing the perceived burden of data provision, enabling the enterprises to better understand the process and putting a "face" to the requests for data.

2.2.2 Single contact point

The Large Enterprise Managers are a single contact point between the statistical office and the enterprise. This involves being both an intermediary and communicating the statistical needs and demands to the specific enterprises and spreading information of changes in the enterprises within the statistical office. It is also essential to identify major problems that may arise.

2.2.3 Getting to know the enterprises

It is fundamental in the work to really get to know the enterprises and their enterprise group. This means becoming familiar with both the legal and operational organisation and being continuously updated on changes in the organisation. It is also important to know the enterprises' data capacity and other unique characteristics, such as accounting systems and other administrative systems.

2.2.4 Updating of the Business Register and profiling of statistical units

Another task of the Large Enterprise Managers is to continuously ensure that the Business Register is updated. Furthermore, through monitoring of legal and statistical units, they can decide on relevant and possible reporting units (enterprise units, kind-of-activity units etc.). As the managers work only with a few enterprise groups, it is possible to use all available sources in the process of updating the Business Register and the profiling of statistical units. They do not have to rely exclusively on the general sources of the Business Register.

2.2.5 Data collection and editing

The Large Enterprise Managers work on a daily basis with data concerning their enterprises. In some specific annual surveys, this means direct work and regular contacts with data providers. For short-term statistics, and other annual surveys, the managers are prepared to be an advisor and a resource to help solve problems arising in the statistical production process. This practical work in the collection process ensures a better understanding of the system of surveys. The task involves continuously monitoring the ongoing data collection process and using this as a tool to identify forthcoming problems both in the data provision and the data collection process. If there are problems in either process, managers may decide to set up customized solutions concerning data provision for specific enterprises.

2.2.6 Coherence in data

The most important objective of establishing a group of Large Enterprise Managers at Statistics Sweden is to improve the quality of statistics in the system of National Accounts. The managers have been given the responsibility to ensure quality and coherence in data from the largest enterprises.

In the long run, it is supposed that the need for data editing, as a result of coherence analysis, will decrease. Working closely together with the enterprises, and when necessary establishing specific solutions for each enterprise, will mean that the enterprises provide high quality data in their normal data provision process and that the need for post-collection editing will decrease.

3. Coherence analysis

In Sweden the large enterprise managers have gradually been forming the practical work. Coherence analysis, as one of the most important tasks of the managers, has been on the agenda since the start in 2004. The work started with a focus on the coherence of data in annual statistics, a work integrated in the normal editing process of three annual surveys. These surveys are main sources of National Accounts and also form the basis of the system of coherence analysis.

3.1 Systematic approach - Extending the coherence analysis

Since the start there have been plans to extend the coherence analysis and to involve additional surveys and fields of data. It has been important to also integrate short term statistics in coherence analysis. In 2005 a systematic approach started, the foundation of the approach was that the surveys and registers are all part of a system - the System of Economic Statistics. This system is very complex and contains of surveys and registers either directly used in National Accounts or related to surveys or registers used in National Accounts. In this system you can identify different types of relationships between the surveys and registers.

3.2 Mapping the connections

The work started by mapping the connections between the surveys and registers in the system, in terms of variables. When integrating both annual and short term statistics and also a broader field of data into the coherence analysis, there was also a need to structure the large sets of data in some way. We started to group the variables into different fields of data, based on the idea that the indicators were related to each other. We created seven coherence areas, suitable for the purpose.

3.3 Coherence areas and relationships

The coherence areas identified were: Production, Foreign trade, Intermediate consumption, Wages and salaries, Employment, Investments and Inventories etc. In some of the coherence areas we also defined several relationships. The purpose of breaking down the coherence areas into different combinations of variables was to be able to give a clearer overview, and a more effective analysis of a large set of variables. The relationships between surveys are either direct; meaning that surveys contain identical or nearly identical indicators, or indirect; meaning that the indicators in the two surveys are not identical but still related to each other in some way.

4. The tool for coherence analysis

The structured approach resulted in a customized tool for coherence analysis. One of the main reasons to build a customized tool was the decentralized data storage at Statistics Sweden. This means that data concerning business statistics was, and still is, located on different physical servers. Data is not stored in a standardized way – the

structures in different databases are not harmonized. In all cases, even if data storage would have been central – it would have been necessary to structure data, and to transform data to comparable aggregates and periods.

4.1 Functionality

The customized tool was built to gather requested data from different databases and physical servers, and to gather data with an up to date status from different kinds of surveys and registers in the actual moment.

4.1.1 Periodicity

The customized tool handles the problem with different periodicity in data, i.e. aggregates short-term-data to annual data, and also monthly data to quarterly data. This handling of different periodicity in data makes it easier to compare data.

4.1.2 Aggregation

The tool also aggregates data collected for different statistical units, to make it comparable. Data on Local unit level is aggregated to Kind-of-Activity unit level, and data collected for Kind-of-Activity unit level is aggregated to Enterprise unit level. From Enterprise unit level, data is aggregated to a, for the coherence analysis unique type of unit, the Coherence unit.

4.1.3 Transformation

The customized tool also transforms gathered data into variables more suitable for the coherence analysis, i.e. the tool aggregates variables on a more detailed level to be more comparable to other variables. Quarterly averages are calculated from monthly data, and annual averages from quarterly data. The aim is to be able to compare data that is not possible just to aggregate.

4.1.4 Presentation

The tool also presents underlying data on demand. By expanding the presentation, you can see the original data gathered from the original data bases for each survey and register in the presentation. This means that you can easily see the data before the transformation and aggregation. It is also easy to copy data into Excel, and process data further on in Excel.

4.2 Proceedings

When setting up a coherence analysis in the customized tool, you start by doing some fundamental settings.

4.2.1 Choose statistical unit

You start your analyse by choosing the most suitable statistical unit for your purpose. Depending on which type of unit you choose to gather data for, the number of surveys or registers possible to gather data from varies. There are five different units available in the system:

- Coherence unit
- Enterprise unit
- Legal unit
- Kind-of-Activity unit
- Local unit

The most common unit to gather data for is the Enterprise unit. The Coherence unit is a unit created only for the purpose of analysing time series in the coherence analysis. It makes it possible to take into account the effects of

organizational changes over time, by creating a unit which is more comparable over time. A Coherence unit consists of one or several enterprise units.

If you choose to gather data on a more aggregated level, i.e. the Coherence unit, you can always drill down to data concerning the underlying units; the Enterprise units, the Legal units, Kind-of-Activity units or Local units. If you choose the Enterprise unit to gather data for, it gives you the possibility to drill down to, and analyse data on the Legal units, Kind-of Activity units or the Local units of the enterprise.

4.2.2 Choose the period

In the next step you choose the period you want to analyse. You choose which year you would like to gather data for. You also choose if you would like to analyse data on a yearly basis, or if you want to analyse data on a quarterly or monthly basis. If you choose to analyse quarterly or monthly data; you can choose to gather data for one single quarter or month, or for all quarters or months during a year.

In the tool it is also possible to gather data in the form of time series. Data is available from 2003 and onwards.

4.2.3 Choose the field of data

In the third step, you decide which field of data you want to analyse. In the tool for coherence analysis, there are eight Coherence areas:

- Production
- Production and Foreign Trade
- Intermediate consumption
- Intermediate consumption and Foreign Trade
- Employment
- Wages and Salaries
- Investments
- Inventories etc.

In the customized tool, we have chosen to combine one original Coherence area, Foreign Trade, with both Production and Intermediate consumption areas. The reason to gather data concerning both Production and Exports is that you get a more complete picture of the flows within the unit. In combination with data on production, you can see flows of re-exports in the data on exports. And it is the same concerning data on Intermediate consumption gathered in combination with data concerning Foreign trade.

For a certain coherence area there are one or several defined relationships. For example the Coherence area concerning Production data consists of four defined relationships: The first relationship gathers data concerning Net sales and Reported deliveries of goods in different sources. The second one gathers data concerning the value of Production of goods and Net sales generated from manufacturing activities reported in different sources. The third relationship gathers data concerning Net sales generated from trading activities collected in different sources. The fourth gathers data concerning Net sales generated from other activities in different sources.

In the coherence area concerning Production data and Foreign Trade data, there are both direct and indirect relationships: The first relationship, Total exports, gathers data concerning the total exports of goods and services from different sources. The second, Total exports and Production, gathers data concerning production and both exports and imports. The third relationship, Production of goods and Exports of goods, gathers data concerning production of goods and exports and imports of goods, grouped by desired level of commodity code. The fourth, Production of services and Foreign trade in services, gathers data concerning production of services and both exports and imports of services, grouped by predefined services groupings.

4.3 Implementation

When we started to use the tool in the ongoing work of the Large Enterprise Managers it became clear that the tool for was useful, not only for detecting lack of coherence in the system at an early stage, but also for monitoring the data collection process. The managers have the possibility to monitor the process of both short term statistics and annual statistics - i.e. to monitor whether changes, due to restructuring or reorganization, appear as expected.

4.3.1 Annual coherence analysis

The customized tool makes it possible to involve additional surveys in the regular coherence analysis done on annual basis; data from both short-term statistics and annual statistics is now easy to access. As the tool also provides the managers a broader field of data, the implementation of routines for involving these areas in the regular analysis has started, but is not fully implemented.

4.3.2 Quarterly coherence analysis

Simultaneously as new routines for annual coherence analysis are implemented, routines for coherence analysis on quarterly basis are also gradually implemented in the management of the largest enterprises. This development is done in co-operation with personnel from National Accounts. As a result of this co-operation, quarterly coherence reports are delivered for a subset of high priority enterprises. The priority enterprises are mainly enterprises that completely dominate a certain product or NACE group in National Accounts, or enterprises operating in areas where there are often inconsistency problems.

There are two purposes of the quarterly coherence reports; the first is to get awareness of inconsistencies at an early stage and to be able to temporarily take care of the inconsistencies in National Accounts, the second is to be able to initiate further investigation. On quarterly basis there is usually too short time to manage to solve inconsistency problems involving several statistical surveys, but as National Accounts personnel are aware of the lack of coherence in a certain area, they can take it in to account on macroeconomic level. In the meanwhile Large Enterprise Managers initiates further analysis and contacts with the enterprise and the different subject matter personnel. In an optimal situation, the next quarter coherence analysis is finalized and data concerning the specific enterprise is coherent.

4.3.3 Quarterly coherence reports

The quarterly coherence analysis reports present data in a standardized. The report consists of four main data sheets presenting time series in different coherence areas: Production area, Foreign trade area, Wages and Salaries area and the area covering data on Employment and hours worked. The time series present data both quarterly and annually. In addition to the presented time series, there are also index series, where data is compared to the same period last year. The index series are presented for a subset of variables. There are also some ratios presented, i.e. Wages and salaries per hour worked, Wages and salaries per employee and an approximation of the development in productivity.

5. The future

Even though we have not fully implemented the new customized tool in the regular work of the Large Enterprise Managers; we have identified some areas for future development. We would like to have automatic signalling of suspected incoherence, with large impact on statistical estimates in general and most important with impact on National Accounts aggregates. The aim is to be able to be more effective in the process, and to shorten the time for analysis. In an optimal situation, the managers would like to be able to “predict” the future questions of National Accounts personnel and get going immediately with the further analysis and contacts.

Another step for the future could be to get the facility to simulate edited values in a certain area, and to see the effects on coherence in a related area. The aim is not to fabricate coherence – but, again, to streamline the process of coherence analysis and to be able to be more effective and to contact the enterprises “once and for all”.

Coherence analysis is a complex area of work, especially concerning large and complex enterprises. Automating the work in this area can not compensate for the knowledge and the understanding of the specific units you analyse. But you can always ensure a common level and decrease the reliance on a certain person.

References

Erikson, A. (2007). Large Enterprise Management - Higher Quality and Lower Burden through Better Relations, *Proceedings of the International Conference on Establishment Surveys (ICES-III)*, American Statistical Association, 825-829.

Erikson, J. (2002). Coherence Analysis as a Tool for Questionnaire Evaluation in Enterprise Statistics, *International Conference on Questionnaire Design Evaluation and Testing (QDET)*