

Use of mirror data in detecting irregularities in declared values of intra-Community trade by HS section

Iwona Markowicz¹, Paweł Baran²

Abstract

Official statistics on intra-Community trade in goods depend on data gathered within the Intrastat system. The database consists of declarations made by businesses dispatching goods to other EU member states or receiving goods from them. The database is collected by national statistical offices and then aggregated by Eurostat. Due to the specifics of data it is possible to assess its quality. Mirror data, i.e. data gathered by both sides of transaction, should be reported twice – by the country of the dispatcher and at the same time by the country of the receiver of traded goods. Thus, for every pair of countries the Intrastat system should contain the same values of trade in goods for the same period. Slight differences are permitted for various reasons, such as statistical thresholds or exchange rates. Generally, however, discrepancies in mirror data are due to errors in reporting or concealment of transactions. And it affects the quality of collected data.

The aim of the study was to analyse the quality of data on intra-Community trade by harmonised system (HS) sections in 2017. The data come from the Comext database provided by Eurostat. For each EU-28 country and each HS section, the quality of data was examined. For this purpose, author's own approach was used to determine aggregated indicators of data discrepancies with country-by-country aggregation. The research allowed us to identify those HS sections for which data quality is the lowest in all individual EU member states.

Keywords: *intra-Community trade, mirror data, HS section*

JEL Classification: *F14, C10, C82*

1. Introduction

Analysis of the activities of businesses must be based on data. Such data are often obtained through questionnaires prepared by national statistical offices (in the European Union, often following the model of Eurostat documents). This is the case with collecting data on the value and volume of intra-community trade of Polish companies. (Baran and Markowicz, 2018). The data is obtained on the basis of statistical declarations. Data collection for the Central Statistical Office is made by Intrastat Department of the Tax Administration Chamber in Szczecin, and the aggregate data are transmitted to Eurostat. It should be noted that these are statistical and not fiscal data. And the fact that there is virtually no sanction affects the timeliness and reliability of filling in forms. However, the Intrastat declaration has been the only source of information on intra-Community trade of Polish companies since 2004.

The problem of the quality of statistical data declared by stakeholders is not only a domain of foreign trade statistics. Similarly, sample surveys (DG1) conducted by the Central Statistical Office are currently the main source of information about revenues earned by small companies

¹ Corresponding author: University of Szczecin, Department of Econometrics and Statistics, 64 Mickiewicza St., 71-101 Szczecin, Poland, e-mail: Iwona.Markowicz@usz.edu.pl.

² University of Szczecin, Department of Econometrics and Statistics, 64 Mickiewicza St., 71-101 Szczecin, Poland, e-mail: Pawel.Baran@usz.edu.pl.

(Dehnel and Wawrowski, 2018). The questionnaire concerns the basic characteristics of the company. Likewise, in the study on the duration of companies, data of the Central Statistical Office from the REGON register is being used (Markowicz, 2014). These data are created on the basis of statistical declarations of entrepreneurs. Registers of applications (registration in the Labour Office) are also used as a basis for registered unemployment statistics (Bieszk-Stolorz and Markowicz, 2015).

As far as this type of research is concerned, an extremely important problem is to monitor the quality of the collected data. The way of collecting data in the Intrastat system in the form of mirror data makes it possible to compare the value (and also the quantity) of goods exported from one country and imported to another EU country. These values are declared in two countries by both parties to the transaction, so when referring to international trade data we mean data on particular trade flow as the value reported by the exporter and at the same time as the value reported by the importer – that is the very idea of mirror values (Ten Cate, 2014). Other bilateral data stored as mirror data include figures on direct foreign investment, foreign debt, and international migration. It is worth noting that the problem of discrepancies (often referred to as asymmetries) in mirror data is not new to economic researchers. Parniczky (1980) indicates that such research has been carried out at least since the 1920s, and Tsigas et al. (1992) date it back to the 1880s.

In (Ten Cate 2014) the discrepancies in reported bilateral statistical data are used to estimate the accuracy of the reporters. In (Ferrantino et al., 2011), the authors detect evading customs declarations with the use of discrepancies analysis. Ferrantino & Wang (2008) describe a measure of data asymmetry similar to the one being the basis of the aggregate index presented later in this paper.

Modeling discrepancies among reported data is a common problem in the compilation of macro-economic statistics (Wroe et al., 1999). In general there is some bookkeeping relation between the reported values which does not hold. According to Ten Cate, the solution consists of estimating the accuracy of the various reports and then finding the optimal adaption of the reported values.

The paper by Lejour et al. (2008) describes CPB's (a government agency in the Netherlands) contribution on bilateral services trade data to version 7 of the GTAP database (OECD Statistics of International Trade in Services). Among others, it uses reliability indices to determine the quality of the data reported by exporting and importing countries and briefly discusses alternative methods to make a choice between two available data being reported.

The aim of the study was to analyse the quality of data on intra-Community trade by harmonised system (HS) sections in 2017. The data come from the Comext database provided by Eurostat. For each EU-28 country and each section, the quality of data was examined. For this purpose, we propose an approach to determine aggregated indicators of data discrepancies with country-by-country aggregation. The research allowed us to identify those HS sections for which data quality is the lowest in all individual EU member states.

2. Methodology

Official statistics on intra-Community trade in goods depend on data gathered within the Intrastat system. The database consists of declarations made by businesses dispatching goods to other EU member states or receiving goods from them. Due to the specifics of data it is possible to assess its quality. Mirror data, i.e. data gathered by both sides of a transaction, should be reported twice – by the country of the dispatcher and at the same time by the country of the receiver of traded goods. Thus, for every pair of countries the Intrastat system should contain the same values of trade in goods for the same period. Slight differences are permitted for various reasons, such as statistical thresholds or exchange rates. Generally, however, discrepancies in mirror data are due to errors in reporting or concealment of transactions. And it affects the quality of collected data.

We used Comext data on intra-Community supplies for 2017 aggregated on country and section level. The goods listed in the database are grouped into 21 HS sections. The survey started with calculating data quality indices for each EU-28 country and each HS goods section.

The quality of data on ICS of a country A by HS section was calculated with the use of aggregated index of data quality:

$${}_Z W_{E^d}^{A,UE} = \frac{\sum_{i=1}^n |E_{A,B_i}^d - I_{B_i,A}^d|}{K} \quad (1)$$

where:

E_{A,B_i}^d – declared value of dispatches (supply) from country A to country B_i ,

$I_{B_i,A}^d$ – declared value of acquisitions by country B_i delivered from country A (mirror data),

$$K = \sum_{i=1}^n \frac{E_{A,B_i}^d - I_{B_i,A}^d}{2}$$

$d = 1, \dots, 21$ – HS section number.

The aggregate index takes values from the range from 0 to 2. The higher the value, the lower the quality of the analysed data is.

On the basis of data quality indices, a ranking of HS sections in each of the EU countries was established and compared. It enabled us to identify the HS sections with the highest and lowest data quality in the EU. The measurement of correlations between indicator vectors determined for individual countries allowed us to identify countries with a similar structure of data quality by section and to distinguish countries with a different structure.

3. Research results – the quality of data

The study used the intra-Community supplies (ICS) values in euros of each EU-28 country and the intra-Community acquisitions (ICA) values mirrored to them. In order to illustrate the scale, main directions and differences in the volume of trade between Member States, the volume of intra-Community trade is presented in Figure 1. (the graph was compiled using the circular library by Gu, (Gu et al., 2014)).

The largest stream of Community goods flows from the Netherlands to Germany. Germany also has the highest turnover (both ICS and ICA). Their main export destinations are France, Great Britain, Belgium and Italy. Spain is another important participant in intra-Community trade in goods. Poland is ranked 8th among 28 EU countries in terms of the volume of trade in goods. The Czech Republic and Austria complete the top ten Member States in terms of intra-EU trade volumes.

The survey was conducted for product groups divided by HS section (listed in Table 1).

For each country, the values of the ICS data quality indicators for each of the HS sections have been determined according to formula (1). The ICS sections in each country were then ranked according to the value of the index. Table 2 presents a summary of the sections with the highest and lowest data quality in intra-Community trade for each of the 28 Member States.

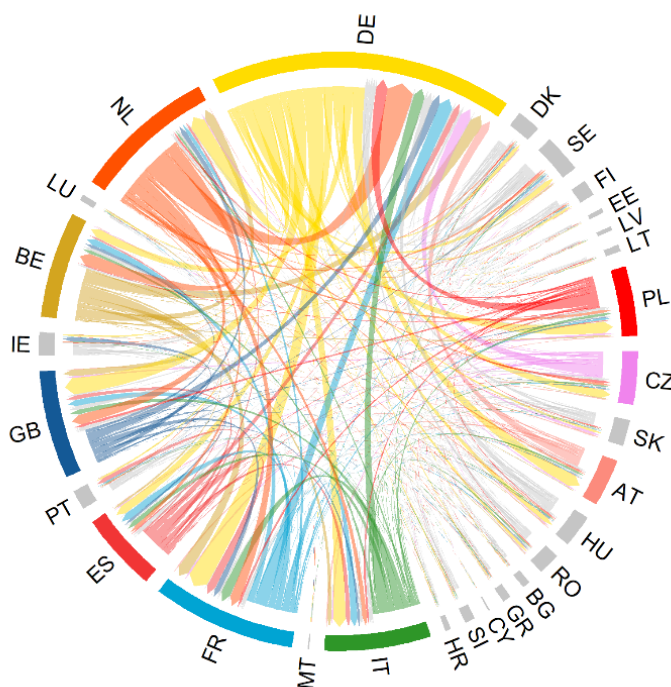


Fig. 1. Intra-Community trade – ICS and ICA by EU-28 countries

Some general schemes can be observed here. Among the sections with the lowest mirror data quality, the most common are sections XIX – weapons, XXI – works of art and special positions, and XIV – pearls, precious stones and metals. On the one hand, these are the sections with relatively small total turnover and therefore transactions within these product groups do not often exceed the statistical threshold. On the other hand, these sections are inherently susceptible to errors, misrepresentations and data gaps (e.g. arms trade can be included among special items, which affects total amounts in two sections).

Among the best-documented sections in terms of turnover reported by both sides are those in which there is a significant exchange of goods between large companies that take seriously the obligation to report intra-Community supplies. Sections I – animals and animal products,

IV – prepared foodstuffs and beverages and VI – chemical industry products – are the most frequently observed in the first three sections according to data quality in different countries. High data quality is also recorded in sections II – vegetable products and VII – plastics, and to a lesser extent in sections XV – base metals and XVI – machinery, appliances, and equipment.

Table 1. HS sections and their descriptions

Section	Description
I	Live animals; animal products
II	Vegetable products
III	Animal or vegetable fats
IV	Prepared foodstuffs; beverages, spirits and vinegar; tobacco and manufactured tobacco substitutes
V	Mineral products
VI	Products of the chemical or allied industries
VII	Plastics and articles thereof; rubber and articles thereof
VIII	Raw hides and skins, leather, furskins and articles thereof; saddlery and harness; travel goods, handbags and similar containers; articles of animal gut (other than silkworm gut)
IX	Wood and articles of wood; wood charcoal; cork and articles of cork; manufactures of straw; basketware and wickerwork
X	Pulp of wood or of other fibrous cellulosic material; waste and scrap of paper or paperboard; paper and paperboard and articles thereof
XI	Textiles and textile articles
XII	Footwear, headgear, umbrellas, sun umbrellas, walking-sticks, seat-sticks, whips, riding-crops and parts thereof; prepared feathers and articles made therewith; artificial flowers
XIII	Articles of stone, plaster, cement, asbestos, mica or similar materials; ceramic products; glass and glassware
XIV	Natural or cultured pearls, precious or semi-precious stones, precious metals, metals clad with precious metal and articles thereof; imitation jewellery; coin
XV	Base metals and articles of base metal
XVI	Machinery and mechanical appliances; electrical equipment; parts thereof; sound recorders and reproducers, television image and sound recorders and reproducers, and parts and accessories of such articles
XVII	Vehicles, aircraft, vessels and associated transport equipment

Section	Description
XVIII	Optical, photographic, cinematographic, measuring, checking, precision, medical or surgical instruments and apparatus; clocks and watches; musical instruments; parts and accessories thereof
XIX	Arms and ammunition; parts and accessories thereof
XX	Miscellaneous manufactured articles
XXI	Works of art, collectors pieces and antiques

In order to indicate which countries have a similar data quality structure by section, the correlation between the vectors of the data quality indicators defined for all sections in the countries was measured. As many as 20 out of 28 countries were found to have a fairly clear correspondence of this structure, while the others differed from the latter and from each other.

Examples of pairs of countries with the highest correlation of data quality structure are Poland and Hungary ($r_{xy} = 0.9670$), Czech Rep. and Latvia ($r_{xy} = 0.9663$) or Estonia and Germany ($r_{xy} = 0.9545$). Figure 2 shows relations and connections between those pairs of countries for which the correlation coefficient between the vectors of the data quality indices in the sections was not less than 0.8. As can be seen, Cyprus (which differs most from the rest of EU-28 in terms of recorded data quality, as confirmed by other surveys), the United Kingdom, Luxembourg, Romania and Malta remain unrelated to other EU countries in this approach. In addition, the group of countries with an individual data quality structure includes Denmark, Ireland and Finland, each with 1, 2 and 3 links in the graph respectively.

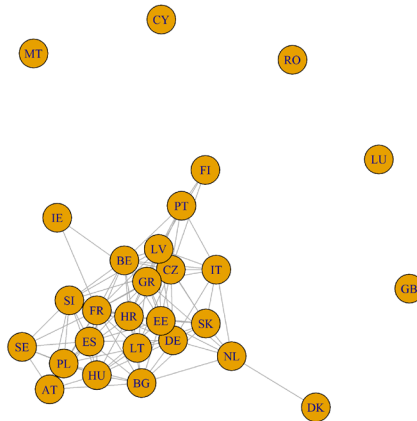


Fig. 2. Data quality correlations between EU-28 countries (only values exceeding 0.8 visible as connections)

Table 2. HS sections with the highest and lowest data quality – by country

Country	Sections with high data quality			Sections with low data quality		
	Rank 1	Rank 2	Rank 3	Rank 19	Rank 20	Rank 21
AT	I	IV	XV	V	XXI	XIX
BE	VI	IV	I	XII	XIX	XXI
BG	XII	XV	XIII	XIV	XXI	XIX
CY	I	VI	II	XII	IX	XXI
CZ	VI	I	XVII	XIX	XIV	XXI
DE	I	IV	X	XIX	XIV	XXI
DK	I	IV	XVIII	VI	XIX	XIV
EE	XX	III	XVI	XIV	XIX	XXI
ES	VIII	III	IV	V	XXI	XIX
FI	VII	IX	XII	III	XVII	XXI
FR	IV	VII	I	XIV	XIX	XXI
GB	XV	VI	IV	XXI	IX	XIX
GR	I	III	IV	XVII	XIX	XXI
HR	XIII	II	I	VIII	XIX	XXI
HU	XVI	IV	IX	XIV	XXI	XIX
IE	I	VI	III	XVII	XXI	XIX
IT	VI	III	VII	XX	XIV	XXI
LT	VI	I	IV	XIV	XIX	XXI
LU	XVIII	VII	XVI	V	XIX	III
LV	IX	IV	I	XIV	V	XXI
MT	XX	XVIII	VII	XXI	XIX*	III*
NL	VII	I	IV	XIX	XXI	XIV
PL	XVI	VI	II	III	XXI	XIX
PT	VI	V	XVI	XVII	XI	XXI
RO	I	II	VI	XIV	XIX*	XXI*
SE	III	IX	XV	XII	XXI	XIX
SI	VI	XIV	XV	VIII	XIX	XXI
SK	XII	XV	II	XIX	XIV	XXI

* both ranked on positions 20–21

Conclusions

HS sections with high data quality in each country and those sections where mirror data is reported asymmetrically were identified. Among the best reported commodity groups, those with large turnover and traded by large companies predominate. Low data quality is more often observed in the case of sections with lower turnover and those where counterparties on one side do not exceed the national statistical threshold. There are also those goods which are not classified unequivocally.

Both literature and EU documents underline the need to ensure the quality of statistical data on intra-Community trade in goods. The survey indicates, on the one hand, the need to improve the quality of data on specific product groups and, on the other hand, it has distinguished a group of countries for which the quality of data differs significantly from the others. There is no doubt that further work is needed on harmonising the procedure for collecting data and improving the quality of data across the EU.

References

- Baran, P., & Markowicz, I. (2018a). Analysis of intra-Community supply of goods shipped from Poland. In M. Papież and S. Śmiech (Eds.), *The 12th Professor Aleksander Zelias International Conference on Modelling and Forecasting of Socio-Economic Phenomena. Conference Proceedings*. Cracow: Foundation of the Cracow University of Economics, 12–21.
- Bieszk-Stolorz, B., Markowicz, I. (2015). Influence of Unemployment Benefit on the Duration of Registered Unemployment Spells. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, 10(3), 167–183.
- Cate ten A. (2014). The Identification of Reporting Accuracies from Mirror Data. *Jahrbücher für Nationalökonomie und Statistik*, 234(1).
- Dehnel, G., & Wawrowski, Ł. (2018). Robust estimation of revenues of Polish small companies by NACE section and province. In M. Papież and S. Śmiech (Eds.), *The 12th Professor Aleksander Zelias International Conference on Modelling and Forecasting of Socio-Economic Phenomena. Conference Proceedings*. Cracow: Foundation of the Cracow University of Economics, 110–119.
- Ferrantino, M.J., Liu, X., & Wang, Z., (2012). Evasion behaviors of exporters and importers: Evidence from the U.S.-China trade data discrepancy. *Journal of International Economics*, 86(1), 141–157.
- Ferrantino, M.J., & Wang, Z. (2008). Accounting for discrepancies in bilateral trade: The case of China, Hong Kong, and the United States. *China Economic Review*, 19(3), 505–520.
- Gu, Z., Gu, L., Eils, R., Schlesner, M., & Brors B, (2014). Circize implements and enhances circular visualization in R. *Bioinformatics*, 30(19), 2811–2812.
- Lejour, A., van Leeuwen, N., & ten Cate, A. (2008). *The quality of bilateral services trade data: contribution to GTAP7 database*. CPB Memorandum, CPB Netherlands Bureau

for Economic Policy Analysis, 212. https://www.researchgate.net/publication/23968410_The_quality_of_bilateral_services_trade_data_contribution_to_GTAP7_database (accessed on 22.01.2019).

- Markowicz, I. (2014). Business Demography – Statistical Analysis of Firm Duration. *Transformations in Business & Economics*, 13(2B), 801–817.
- Parniczky, G. (1980). On the Inconsistency of World Trade Statistics. *International Statistical Review*, 48(1).
- Tsigas, M.E., Hertel, T.W., & Binkley, J. K. (1992). Estimates of systematic reporting biases in trade statistics. *Economic Systems Research*, 4(4).
- Wroe, D., Kenny, P., Rizki, U., & Weerakoddy, I. (1999). *Reliability and quality indicators for National Accounts aggregates*. <http://epp.eurostat.ec.europa.eu>, Office for National Statistics (UK).