STATISTIKA

STATISTICS
AND ECONOMY
JOURNAL

VOL. **99** (1) 2019

CZECH
STATISTICAL
OFFICE

EDITOR-IN-CHIEF

EDITORIAL BOARD

EXECUTIVE BOARD

MANAGING EDITOR

Dear Readers,

this year we celebrate the 100th anniversary of the state official statistics in our country (the State Statistical Office was founded in 1919). This event will also be commemorated on the pages of our journal: special papers reminding previous developments in the fields of official statistics as well as essential achievements were and will yet be published.

Statistika: Statistics and Economy Journal of the Czech Statistical Office celebrates also its 99th anniversary this year. Upon this occasion we would like to express our thanks for your lasting interest and favor. The journal in its current form has been issued since 1964 following the tradition of the previous journals of the state official statistics (starting in 1920). Since 2011, it has been published quarterly in English only (both in print and online versions).

Over the years the journal has went through many changes following requirements of its readers, visions of the journal board members, and professional interests of our authors. We are proud of the fact that its citation indices in Scopus increase constantly (*SJR* and *CiteScore*) and except the journal membership in the *ESCI* of the Web of Science and in others databases of scientific journals, *Statistika* was newly included also in Pablikado.

On behalf of the Executive Board we are looking forward to further cooperation with all the authors (and reviewers) bringing results of analyses in the fields of economy, the environment, or social sciences, and reflecting the role of official statistics in supporting decision making processes at all levels. We also believe that the papers published in our journal will continue to be of great value for your everyday work and professional growth.

Marek Rojíček
President of the Czech Statistical Office

Prof. Stanislava Hronová
Editor-in-Chief

CONTENTS

ANALYSES

DISCUSSION

CONSULTATION

100th ANNIVERSARY

INFORMATION

About Statistika

The journal of Statistika has been published by the Czech Statistical Office since 1964. Its aim is to create a platform enabling national statistical and research institutions to present the progress and results of complex analyses in the economic, environmental, and social spheres. Its mission is to promote the official statistics as a tool supporting the decision making at the level of international organizations, central and local authorities, as well as businesses. We contribute to the world debate and efforts in strengthening the bridge between theory and practice of the official statistics. Statistika is professional double-blind peer reviewed open access journal included in the citation database of peer-reviewed literature **Scopus** (since 2015), in the **Web of Science** *Emerging Sources Citation Index* (since 2016), and also in other international databases of scientific journals. Since 2011, Statistika has been published quarterly in English only.

Publisher

The Czech Statistical Office is an official national statistical institution of the Czech Republic. The Office's main goal, as the coordinator of the State Statistical Service, consists in the acquisition of data and the subsequent production of statistical information on social, economic, demographic, and environmental development of the state. Based on the data acquired, the Czech Statistical Office produces a reliable and consistent image of the current society and its developments satisfying various needs of potential users.

Contact us

Journal of Statistika | Czech Statistical Office | Na padesátém 81 | 100 82 Prague 10 | Czech Republic
e-mail: statistika.journal@czso.cz | web: www.czso.cz/statistika_journal

Methodology of Estimating "Financial" Margins and their Capturing in the System of National Accounts

Jana Kramulová [1] | *University of Economics in Prague; Czech Statistical Office, Prague, Czech Republic*
Helena Houžvičková [2] | *Czech Statistical Office, Prague, Czech Republic*
Jakub Vincenc [3] | *Czech Statistical Office, Prague, Czech Republic*

Abstract

Margins on buying and selling transactions, or simply "financial" margins, form one part of the complex system of national accounts. These financial services in acquiring and disposing of financial assets and liabilities on financial markets constitute an important role in output of financial institutions. Up to now, the Czech Statistical Office recorded only small part of these margins in national accounts, but in the next revision which is to be published in June 2020 the system should include all types in the whole time series 1993–2019.

The aim of this paper is to develop the missing methodology of capturing margins and demonstrate all difficulties connected with their estimation.

Our approach is influenced by the fact that in the Czech Republic no suitable database with detailed information about transactions is available. That is why simplifications and assumptions needed to be formulated. The paper contains time series of cross-border margins and domestic margins on transactions with securities, shares, investment fund shares and foreign currencies.[4]

Keywords

National accounts, margins on buying and selling transactions, methodology, Czech Statistical Office, cross-border and domestic margins, tradable financial assets

JEL code

E44, G20, O16

[1] University of Economics in Prague, Faculty of Informatics and Statistics, Department of Economic Statistics, Nám W. Churchilla 4, Prague 3, Czech Republic. Also the Czech Statistical Office, Department of Financial Accounts, Na padesátém 81, Prague 10, Czech Republic. E-mail: jana.kramulova@czso.cz, phone: (+420)274052466.

[2] Czech Statistical Office, Department of Financial Accounts, Na padesátém 81, Prague 10, Czech Republic. E-mail: helena.houzvickova@czso.cz, phone: (+420)274054265.

[3] Czech Statistical Office, Department of Financial Accounts, Na padesátém 81, Prague 10, Czech Republic. E-mail: jakub.vincenc@czso.cz, phone: (+420)274054363.

[4] Earlier version of this paper was presented at the *AMSE 2018* conference held in Kutná Hora, Czech Republic (29th August–2nd September 2018).

INTRODUCTION

This paper deals with margins on buying and selling transactions with financial assets and their capturing in the system of national accounts. The system of national accounts is a complex model of total economy that tries to describe relations among economic entities as well as flows of money. The whole system consists of several various accounts that characterise the economic process from production and intermediate consumption through generation and redistribution of income up to financial account and balance sheets. Each economic activity should be recorded in national accounts on the corresponding accounts with all consequences. All the tables (accounts) are published worldwide by official statistics, i.e. national statistical institutions (hereinafter NSIs). The system is in detail described, explained and standardised in manuals of national accounts – in European System of Accounts ESA 2010 (European Commission and Eurostat, 2013) in the European Union or in System of National Accounts SNA 2008 (European Commission et al., 2009) in other countries.

Although the national accounts cover plenty of crucial information, items and phenomena, the far most important, discussed and cited are probably the gross domestic product (hereinafter GDP) and gross national income (hereinafter GNI). Giovannini (2008, p. 162) even states, that "of all the variables in the national accounts, the most prominent is gross domestic product (GDP)". Both indicators assess the performance of the economy, GDP using the total output of a country produced by both residents as well as non-residents (Kramulová and Musil, 2013), GNI considering also "net primary incomes (mainly dividend, interests and reinvested earnings) with non-residents" (Vltavská and Sixta, 2015, p. 107). Margins on buying and selling transactions affect the final value of GDP as well as of GNI.

ESA 2010 includes more types of margins in national accounts. The main ones are trade and transport margins in the system of input-output tables (for details see also Streicher and Stehrer, 2015), then merchanting margins in services in export and import (for details see also Broussolle, 2015), interest margins (for details see e.g. Fungáčová and Poghosyan, 2011) connected with FISIM (Financial Intermediation Services Indirectly Measured) and, finally, margins for the provision of financial services, called sometimes margins on buying and selling transactions (see e.g. International Monetary Fund, 2009) or "financial" margins or financial services in acquiring and disposing of financial assets and liabilities in financial markets.

Why are these "financial" margins even present in economy? According to our opinion there are two main reasons. The first is the information asymmetry, i.e. the fact that the financial intermediator has always more information about the prices than the customer. The second reason is that the customer is ready to pay the margin within the price of a financial asset, because he knows that the service is included.

Although the margins on buying and selling transactions were defined even before ESA 2010, up to now, the Czech Statistical Office (hereinafter CZSO) has recorded only a small part of "financial" margins in national accounts. As it is quite a complicated topic, Eurostat has recently been very active in investigation about proper recording of the margins in national accounts throughout Europe. This topic is very up-to-date because the CZSO is now developing methodology for estimation of these margins and will include them in benchmark revision in June 2020. Apart from ordinary revisions that are done each year and cover mainly changes due to better data sources, irregularly occur extraordinary revisions that are carried out usually due to changes in methodology or standard of national accounts.

The aim of this paper is to develop missing methodology of capturing the margins and demonstrate all difficulties connected with their estimation. It starts with the decision about national accounts sectors of margin producers, items that are affected or data that you need but cannot obtain.

The paper is organized as follows: Section 1 briefly presents different types of financial services defined in ESA 2010 and provides information about institutional sectors and items that are relevant for margins on buying and selling transactions. Section 2 shows methodological approaches discussed at the beginning of our project as possible sources of methodology. Section 3 introduces our final proposed methodology of estimation of "financial" margins for tradable financial assets. Section 4 is aimed at experience from other

institutions. In Section 5 we discuss difficulties and obstacles faced during our methodology development. Section 6 shows preliminary results of estimates in the Czech national accounts and different approaches that needed to be applied. Section 7 is devoted to capturing of "financial" margins in national accounts. Section 8 brings the discussion about retrapolation of time series of margins dated back to the year 1993.

1 THEORETICAL BACKGROUND

1.1 Financial services

Let us aim now at the theoretical background. Corporations covered by the system of national accounts could be divided into two main groups: "units mainly providing financial services and those mainly providing goods and other services" (United Nations Statistics Division and European Central Bank, 2014, p. 44). Units providing financial services are financial corporations, classified in the sector of financial institutions in the system of national accounts; their production is "the result of financial intermediation, financial risk management, liquidity transformation or auxiliary financial activities" (United Nations Statistics Division and European Central Bank, 2014, p. 44).

Financial corporations are not all the same; there are many varied units (e.g. monetary institution, investment funds, holding companies, insurance companies). In these cases it is not possible to estimate the output using one universal method. Manual ESA 2010 (European Commission and Eurostat, 2013, p. 63) defines three types of financial services:

a) "financial intermediation (including insurance and pension services);
b) services of financial auxiliaries; and
c) other financial services."

"Financial services may be paid for directly or indirectly. Some transactions in financial assets may involve both direct charges and indirect charges. Financial services are provided and charged for in four main ways":

a) "Financial services provided for direct payment" – in form of fees and commissions are usually implicitly included in profit and loss statement.
b) "Financial services paid for through loading interest charges" – marked as FISIM, cannot be directly obtained from data sources; estimation is made under Council Regulation (EC) No 448/98 (Council of the European Union, 2003).
c) "Financial services in acquiring and disposing of financial assets and liabilities in financial markets" represent indirect charges not directly included in business accounting system of corporations. Estimation method is not commonly set up by ESA 2010. Developing of method for estimation in the Czech national accounts is the main topic of this article.
d) "Financial services provided in insurance and pension schemes, where the activity is financed by loading insurance contributions and from the income return on savings" are estimated as model calculation under ESA2010, see chapter 16 (European Commission and Eurostat, 2013, p. 63).

All above mentioned parts of output should be recorded in the system of national accounts. For example if a NSI ignores the margins, the value of output of financial services arising from these transactions would be understated (United Nations Statistics Division and European Central Bank, 2014, p. 105) which subsequently leads to an understatement of GDP and GNI estimates.

1.2 Institutional sectors and financial assets relevant for "financial" margins

Whereas other types of margins are cited in scientific papers, margins on buying and selling transactions are mainly subject of official statistics and are discussed among NSIs or national central banks (hereinafter NCBs). These financial services in acquiring and disposing of financial assets and liabilities in financial markets constitute an important role in output of financial institutions. Financial institutions (labelled S.12)

form one of six main institutional sectors in the system of national accounts, apart from non-financial corporations (S.11), general government (S.13), households (S.14), non-profit institutions serving households (S.15) and the rest of the world (S.2). They are further divided into nine institutional sub-sectors, see Table 1. Some other sectors are also divided, for details see European Commission and Eurostat (2013, p. 31).

Table 1 Nine sub-sectors of financial institutions (S.12)

Sub-sector	Label
Central bank	S.121
Deposit-taking corporations except the central bank	S.122
Money market funds (MMF)	S.123
Non-MMF investment funds	S.124
Other financial intermediaries, except insurance corporations and pension funds	S.125
Financial auxiliaries	S.126
Captive financial institutions and money lenders	S.127
Insurance corporations	S.128
Pension funds	S.129

Source: Adapted from European Commission and Eurostat (2013, p. 511 and following)

One of our first tasks was to select institutional sectors (see Table 1) that are affected by "financial" margins. The Financial Handbook (United Nations Statistics Division and European Central Bank, 2014) suggests in paragraph 3.126 as producers of margins three financial sub-sectors: S.122 (banks), S.125 (security and derivative dealers) and S.126 (foreign exchange bureaux). From user side, volume of the margins should be according to paragraph 3.128 allocated in national accounts into final household consumption expenditures (FHCE) in case of S.14, into intermediate consumption in case of other residential sectors (S.11, S.12, S.13 or S.15) and into export of services for S.2.

Concerning financial assets (for the list of financial assets see European Commission and Eurostat, 2013, p. 521 and following) that should be covered, manual ESA 2010 in paragraph 3.73 proposes to apply margins on securities (AF.3), equities (AF.51), investment fund shares (AF.52) and foreign currencies (AF.21). The Financial Handbook (United Nations Statistics Division and European Central Bank, 2014, p. 105) recommends in Chapter 3 to estimate margins for "foreign exchange, shares, debt securities – such as bills and bonds – financial derivatives and investment fund shares", i.e. only AF.71 in addition to what ESA 2010 proposes.

2 VARIOUS METHODOLOGICAL APPROACHES TO "FINANCIAL" MARGINS

In case of estimation of "financial" margins in national accounts any required approach does not exist. There are only some suggestions how the estimation can be done. Especially ESA 2010 manual is open to any method which will be able to calculate appropriate value of the margin. Based on this situation at the beginning of our research we skipped all the official suggestions and tried to develop our own methodological approach.

2.1 Alternative ways

The very first approach was focused on the Czech stock market index PX and other indexes of relevant financial funds. Using these data on suitable financial instruments we wanted to measure their real

revaluation. Then the idea was to take the differences between stock of each instrument at the beginning and at the end of the period and try to exclude the transactions and the revaluation. The part which remains should be the margin, because no other changes usually happen.

However, we faced variety of obstacles. Firstly, turnover on the Prague Stock Exchange is too low and covers only a few titles. Almost the same problem was with the indexes of financial funds. Moreover, they usually cover mix of different financial instruments and it was impossible to separate them. Due to these reasons the approach does not provide useful data for estimation of the margin.

The aim of the second approach was to estimate the margin using data from VAT returns, as the data are managed by the Czech Statistical Office and are easily available. Although, according to the Czech law the financial services in general are excluded from deduction of VAT that is the reason why most of the margin producers are not even taxpayers. Nevertheless, it seemed to us that there might be some possibility. Especially when we have read the statement of the Ministry of Finance of the Czech Republic (Czech Ministry of Finance, 2007), which mentions the margins in connection with VAT returns in case of exchange bureaux.

As a result, we made analysis of exchange bureaux VAT returns and found out the following information. When the producer of the margin has except the main activity also a secondary activity and one of them is not excluded from deduction of VAT, then it might be more convenient for him to become a taxpayer. If this happens, he has to divide his costs into two parts; one part which is suitable for VAT deduction and the second one which is not. For this separation it is necessary to use a special coefficient you can see in Formula (1).

$$coefficient = \frac{revenues\ for\ taxable\ part}{total\ revenues\ for\ both\ activities} \tag{1}$$

Back to the margin which is hidden in the formula in the part called total revenues. It is a sum of incomes from activities excluded (financial services in general) and not excluded from the deduction of VAT. The statement suggests that the incomes from financial services should be determined as a margin, especially in case of exchange bureaux. It means that when we can easily get the coefficient from VAT returns and the value of total income as well, then we are able to separate the margin from the total income at least for transactions with foreign currencies.

But this approach brings also many problems. The most serious one is that only a few exchange bureaux are VAT payers (to be exact, just 46 out of 126 units in NACE 66.12). In addition, we are not sure who determinates income from financial services as a margin. Considering the above; it is only a suggestion. Thus, in order to obtain the best possible results, we have to create our methodology based on suggestions in manuals.

2.2 What do the manuals say?

At the beginning we went through many handbooks and manuals. The oldest one which probably came up with the margins relating national accounts is the ESA 95 manual (European coal and steel community, European communities, European atomic energy community, 1996). The manual (paragraph 3.64) defines the margins as a production of financial services obtained by customers with no direct payments. And the financial services mean acquiring and disposing securities with financial intermediator.

According to the manual the treatment of this margin should be the same as the treatment of wholesale and retail trade margins. It means that the estimation of the margin is described as a difference between bid (buy) and ask (sell) price. Nevertheless, it includes holding gain which has to be eliminated (for more information about holding gains and revaluation see Rybáček, 2010).

Almost the same description of the margin you can find in the Balance of Payments and IIP manual (BPM6) published in 2006 (International Monetary Fund, 2009), in comparison to ESA 95 it brings practical recommendation for estimation: "The service can also be measured by applying the dealers'

average margin as a percentage to the value of transaction through dealers" (paragraph 10.123). This approach deals with the holding gain problem, however, the dealers' average margin is definitely not easily accessible. Moreover, the BPM6 says that the margins are a part of "…the financial transaction to which they relate" (paragraph 10.123).

It is obvious that the updated manual ESA 2010 had to develop the definition of margins in more details. One of the differences between both manuals is that the producer of the margin is not defined institutionally as a financial intermediator, but by financial market area. Nevertheless, the main difference is hidden in the description of the margin which was divided into two parts. The description depends on whether it is buying or selling transaction and sounds as follows. "When a financial institution offers a security (e.g. bill or bond) for sale, a service charge is levied. The purchase price (the ask price) is equal to the estimated market value of the security plus a margin … when a security is sold, the price offered to the seller (the bid price) being equal to the market value minus a margin" (paragraph 3.73).

Our last source of methodological background was the Financial Production, Flows and Stock in the SNA Handbook (United Nations Statistics Division (UNSD), European Central Bank (ECB), 2014), which proceeds from ESA 2010 manual. Furthermore, it brings useful information about allocation of the production of the margins. According to this manual the production is a part of intermediate consumption or a part of final consumption in case of S.142. The manual does not forget to mention how to care about the cross-border transactions: "Margin should be recorded in exports of goods and services if the financial services are provided to the rest of the world" (paragraph 3.128).

The manual also more deeply develops BPM6 methodology of the estimation of the margins and divides it into two approaches. "Bottom-up" approach is based on "…computing the value of the margin consumed by each sector by applying an average sectoral margin as a percentage of the sectoral value of transaction" (paragraph 3.128). However, the calculation of the value of the margin by sectors is impossible in almost every country due to missing data.

This is the reason why most of the countries will probably use the "top-down" approach, which unlike the bottom-up use the economy-wide average margin. Then the allocation of the economy-wide value of the margin to the appropriate sectors should be done by the volume of the financial transactions in each sector. The manual also admits to use other indicators, which in our point of view could be the stock of relevant assets. Nevertheless, the volume of the financial transactions is the most corresponding one.

2.3 Similarities with FISIM

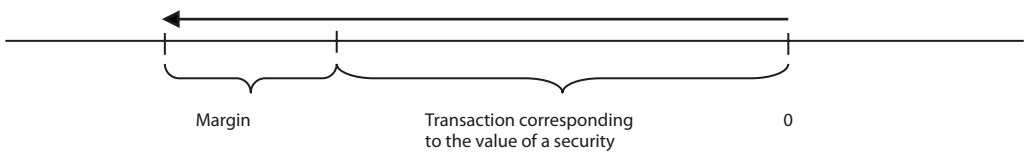
The above information leads to the idea that the margins are similar to the FISIM in many ways. Both of these belong to the group of services which are interesting in the fact, that on one hand their production is indirectly measured and on the other hand they should be allocated as a part of intermediate or final consumption and export of services. The only difference comes from the fact the margins are focused on buying and selling transactions with securities and the FISIM on interests.

It means that if the allocation of the FISIM is by the volume of interests then the margins should be allocated by the financial transactions. And, for the same reason, the correction in case of the margins ought to be done to the financial transaction unlike the FISIM where the correction touches the interests. However, one more difference should be mentioned here. As the FISIM correction is made on both sides of the resources and uses, the margin correction has to be done only in assets. That is because the margins arise from the trading of financial assets only and the FISIM is calculated from borrowings and loans as well.

2.4 Illustration of the margin hidden in financial transaction with securities

To be more accurate let us show you on following figures where the margins are actually hidden, but keep in mind one rule. The arrow always shows the value of the transactions which is currently captured in national accounts of the Czech Republic. And the sense of the margins correction is simply to move a part of these transactions to the production.

Figure 1 Financial transaction with a security when financial intermediator sells it (from his point of view)

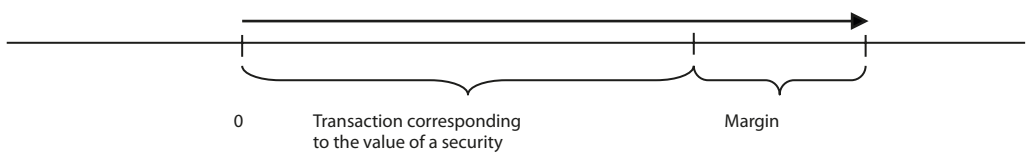


Source: Own elaboration

Figure 1 shows the financial transaction with a security from the financial intermediator point of view, when he sells a security to the customer. The value of the transaction is negative, because it is the opposite of the corresponding money flow. As it is mentioned in introduction, the financial intermediator has more information and that is the reason why the money flow includes also a hidden pay for the services – the margins, which should be part of the production.

In these situations the financial transactions are negative and we have to add the margin to them. Hence, according to the ESA 2010 manual (European Commission and Eurostat, 2013) "Transactions in shares in circulation are recorded at their transaction value" (paragraph 5.156). However, these transactions currently include also a payment for the services, which will be relocated to the production of the financial intermediator.

Figure 2 Financial transaction with a security when common customer buys it (from his point of view)

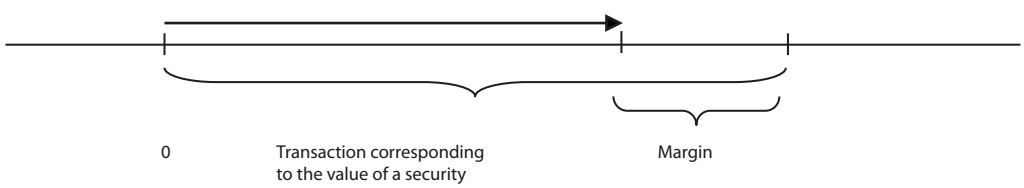


Source: Own elaboration

Figure 2 shows the same situation from the common customer point of view, who buys a security from more experienced financial intermediator. The customer paid more money for a security than he had to pay due to the margin. This issue is also included in the corresponding financial transaction with a security as you can see in Figure 2. In these cases the financial transactions are positive and we have to relocate the margin from them to the intermediate or final consumption.

The financial intermediator is a producer of the margin even if he buys a security, because he is able to reach a better price. It means that the money flow is underestimated in comparison to the value of a security and the corresponding financial transaction is underestimated as well. These situations are illustrated in Figure 3 which shows that according to the margin correction we have to add margin to the financial transactions and the same part to the production of the financial intermediators.

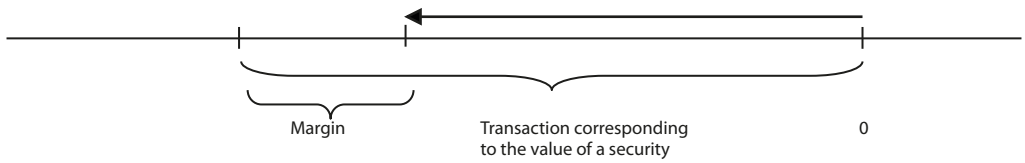
Figure 3 Financial transaction with a security when financial intermediator buys it (from his point of view)



Source: Own elaboration

If we look at this situation from the common customer point of view, we can realize that he receives less money than he has to. Figure 4 shows this fact on the negative financial transaction with a security. The transaction is negative because the customer sells a security. However, it should be even more negative due to the margin as the value of a security is currently based on the underestimated money flow. On the other hand, the margin goes to the intermediate or final consumption.

Figure 4 Financial transaction with a security when common customer sells it (from his point of view)



Source: Own elaboration

To sum up previous paragraphs we can say that in case of the financial intermediators we have to add the margins to the financial transactions with securities and it does not matter whether we speak about buying or selling transaction. Unlike the case of the common customer, the margins have to be deducted from these transactions. Nevertheless, the capturing of the margins will be described in Section 7 in more detail.

3 PROPOSED METHODOLOGY TO ESTIMATION OF "FINANCIAL" MARGINS

Within the effort to find the optimal estimation method of the margins with knowledge of all the background above, we have continued with tendency to find inspiration in computation of the FISIM. Nevertheless, because of partial dissimilarities of both services (the main one is, that adjustment of the FISIM relates to interest from deposits and loans and margins are derived from transactions with the financial assets only), we have faced different obstacles.

When thinking about estimation of the margins, we can mention a parallel with FISIM, where interest is adjusted in order to correspond to real value of money i.e. internal rate of return (IRR). In case of margins, the financial transactions with the assets AF.21, AF.3 and AF.5 should be adjusted to correspond to real value of each asset. It means without added payment for the financial services to the financial intermediator – the margin. Based on these similarities and suggestions of the manuals we have finally decided to develop a spread.

3.1 Treatment of the debt securities, equity and investment fund shares/units

Our spread is an analogy to IRR and expresses an average economy-wide margin as a ratio to each financial transaction. The ratio comes out of the analysis of the margins at the Prague Stock Exchange made by the Czech National Bank (CNB) which assumed that in the Czech Republic the spread equals to 0.03%. Therefore, we can argue that we use the "top-down" approach in the terminology of SNA Handbook.

According to Formula (2) we have to use the ratio to multiply the value of the financial transactions with AF.3 and AF.5. Since no suitable data source with detailed information about transactions is available, we get these transactions from a database which is monthly compiled by the CNB as a difference between the stock of assets at the beginning and at the end of the month. Transactions, connected with emission and expiration of the securities, are excluded from the database because there is no margin included.

$$Margins = Transaction\ volume \cdot Spreads\ (as\ a\ ration\ to\ mid - price) \quad (2)$$

Moreover, the transactions in the database are divided per various financial assets. It means that after the multiplication we reach the volume of the margins separately for each asset as you can see in Table 6, which helps us to make the margin correction to the right item of the financial transactions.

Procedure of estimation of cross-border margins will be described in Section 4.4.

3.2 Treatment of the exchanging money services

Production of the margins in the field of money exchanging services takes part only in the sub-sector S.126. We neglect bank exchange bureaux, because they realize only 2–3% of exchanges. The treatment of these margins is slightly different.

The main principal is to charge fees for the services provided by exchange bureau when exchanging foreign currency. These fees are again hidden in offered exchange rate that is not equal to market exchange rate. It can be even dramatically different; we can point out documented cases of exchange with the exchange rate 15 CZK/EUR (iDNES.cz, 2018).

We can say that currency can be treated in the similar way as securities, because it is also traded on the market. But during the estimation process a special approach is needed.

The spread for the money exchanging services is based on the rate published by one of the biggest exchange bureaux in the Czech Republic, Exchange s.r.o. Then, the difference between its monthly average and the average of the market exchange rate (published by the CNB) for the same period is made to reach the spread.

In this case the spread shows the margin which exchange bureau gets for exchanging one Czech crown. We have six different spreads for the most exchanged currencies in the Czech Republic such as CHF, EUR, GBP, HRK, PLN, and USD. For example the spread of EUR/CZK and CZK/EUR transaction was CZK 0.6 in the second quarter 2017.

For the calculation of these margins we use the buy and sell turnovers unlike equation (2). These turnovers are divided by currencies which are available from statistical statement Dev(ČNB) 26-04 entitled "Purchase and sale of foreign currency". This survey is done by the CNB and is obligatory for exchange bureaux with the higher turnover than 20 million per year.

4 EXPERIENCE FROM ELSWHERE

As written in Section 2 almost no source of the best practice exists in the manuals. Moreover, no information from Eurostat or other NSIs exists either. Despite each country should elaborate the GNI inventory (for public versions see Alfresco Content Repository, 2018), i.e. procedure of step-by-step estimation of total GNI. To date in all available documents (12 countries in October 2018) there is no evidence of "financial margins". It is a question if it would be truly useful, because various countries have different conditions at the market, different data available and can be able to use various estimation methods.

We also present some information about estimation procedures throughout Europe from Eurostat questionnaire and the CNB methodology.

4.1 Questionnaire from EUROSTAT

At the beginning of the year 2018 EUROSTAT sent a questionnaire on financial services to all EU countries. One part of it was dedicated to margins from selling and buying transactions. The questionnaire asked for quite detail information about data sources and methods of estimation of margins. In Table 2 you can find summary results of this questionnaire presented by the European Commission and Eurostat.

From figures listed in the Table 2 we can conclude, that there are imperfections in this field almost by all EU countries. As a result most of the EU countries were assigned a task to check their attempt to calculate margins.

Table 2 Overview of the country practices concerning services associated with the acquisition and disposal of financial assets and liabilities (ESA2010 paragraph 3.73)

Question	Yes	Partly	No	N/A	No answer
Output from trading financial assets reported to be included in GNI?	13	6	6	0	3
Output derived from a model using value of transactions and an average percentage margin?	0	2	18	5	3
Margins deemed to be explicitly included in the source item?	16	0	3	6	3
Does the category in the source used to calculate output explicitly refer to "margin" or "spread"?	1	2	13	9	3
Adjustments to the source item made?	0	1	15	8	4
Adjustments made on the use side to reflect the consumption of services associated with trading financial assets?	11	0	7	6	4

Source: Adapted from European Commission and Eurostat (2018)

4.2 Czech National Bank

The CNB has already calculated margins from selling and buying cross-border transactions of securities. These figures have been already involved in balance of payment for the Czech Republic since 2012.

Calculation of margins is based on two main sources (European Commission and Eurostat, 2018):

1) Item portfolio investment is used for the estimation of:
   a) Export of services – purchase or sale of domestic securities by non-residents,
   b) Import of services – purchase or sale of foreign securities by residents.
2) Item foreign exchange reserves held by the CNB (sub-item foreign securities as a part of reserve assets) is used for the estimation of import of services – purchase or sale of foreign securities by residents (i.e. by the CNB).

"Because of the system of data collection on portfolio investment by kind of securities (security by security) and availability of data in CNB database (SBS database), the basis for the estimation procedure is net monthly purchase/sale of each kind of security on a market in a reference period (month)" (European Commission and Eurostat, 2018, p. 21).

"When the margins are calculated each financial market has its own (percentage) spread assigned (i.e. transaction rate – mid rate). The estimation of spread is based on supplement data provided by financial market supervision by the CNB" (European Commission and Eurostat, 2018, p. 21). As far as estimation of spread related to trade in securities in CNB reserve assets is concerned the calculation is analogous and is consulted with CNB dealing centre.

"Estimated spreads are applied to purchases and sales of securities by instrument and territory of an issuer" (European Commission and Eurostat, 2018, p. 21).

5 DIFFICULTIES AND OBSTACLES (DATA)

Apart from obstacles that were up to now mentioned in previous sections (e.g. no best practice in GNI inventories) we faced several other difficulties and obstacles.

Although ESA 2010 defines in paragraph 3.73 (European Commission and Eurostat, 2013, p. 64) financial services consisting of acquiring and disposing of financial assets and liabilities in financial markets, there is no more specific description, how to estimate these margins.

The best way under the handbook of United Nations Statistics Division and European Central Bank (2014, p. 105) is "to develop a securities database capturing each transaction in the financial asset together with the bid and offer prices so that it can calculate the margin for each transaction". But in practise

such a database does not exist and it is impossible to create it, too costly and also badly feasible for NSI. This is the case of the transactions with securities which represent the difference between the stock of each security at the beginning and at the end of the month, because the real turnovers are not available. Therefore, there is an assumption that these securities are not traded too frequently during a month. If we are wrong and the real turnovers are much higher, this assumption underestimates the value of the margin from trading securities.

The other possible obstacle is unavailability of required data. This problem can have more forms. The data are not available in a desirable time series, as often as it is needed, in the suitable form or no one provides us all the needful data. Let us show an example. The task (now, in the year 2018) is to estimate "financial" margins of exchange bureaux in the Czech Republic for the whole time series Czech national accounts are published (i.e. since 1993), if possible in quarterly periodicity. If the necessary assumptions are formulated, there could be the chance to estimate let's say all quarters of the year 2017. But what is the situation like with the estimation of previous years?

Using Japanese case (Takeda, 2013) we can formulate previously mentioned Formula (2) for calculation of margins. It means that we need for estimation of margins transactions and spreads, the assumptions relate to both parts of the equation. Firstly, we have some requirements on data about buy price; sell price and also mid-price. Mid-price is for simplicity defined as the middle between buy and sell price (or alternatively bid and ask) as in Formula (3), because usually the "real" mid-price is not known. The mid-price is then used for calculation of the spread according to Formula (4).

$$Mid - price = \frac{bid + ask}{2} \tag{3}$$

$$Spread = \frac{ask - mid}{mid} \tag{4}$$

But, secondly, we have other requirements also on data about transactions. Not in all cases we even have the transactions and sometimes we have to replace them by another indicator, e.g. transaction as closing stock less opening stock.

Another important requirement in this field is a necessity of a narrow cooperation between NSI and NCB. The difficulty may occur when these two institutions do not communicate; fortunately this is not the case of the Czech Republic. CZSO and CNB cooperate on a long-term basis in both financial accounts and balance of payment / international investment position areas. Margins are then the result of both sides according to a joint methodology.

6 PRELIMINARY RESULTS – FIRST ESTIMATES IN THE CZECH NATIONAL ACCOUNTS

Due to data limitations and various types of financial assets different approaches had to be used, as we introduced in previous sections. Since we are in the middle of our grant, we have still preliminary results for margins that are subject to further possible changes. All results would be impossible without perfect cooperation with our colleagues from CNB. We can divide the results into three categories.

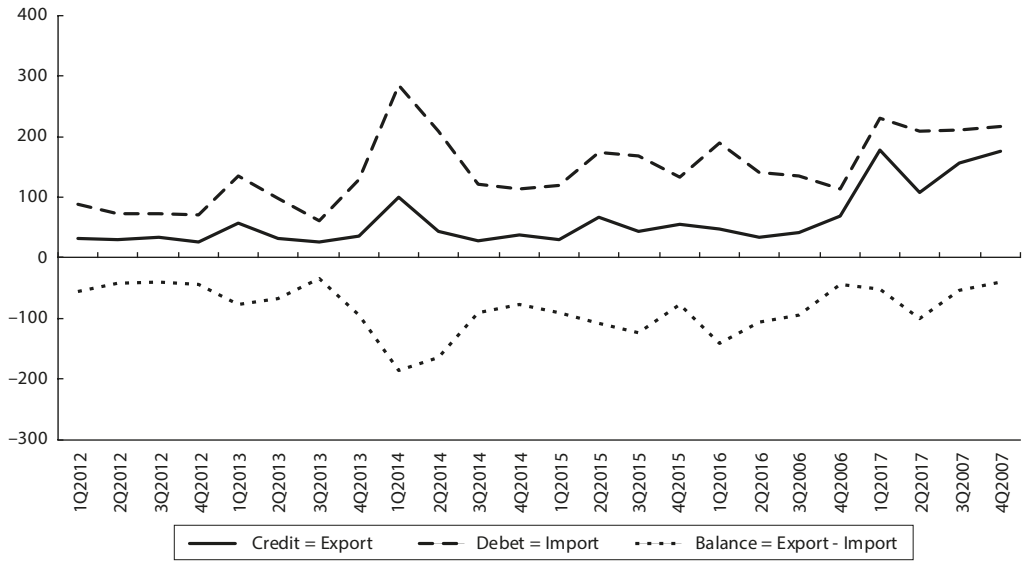
Firstly, we can present export and import of margins on securities, shares and investment fund shares. This part is included in the Czech national accounts already for a longer time. Secondly, new estimates of domestic margins on securities and shares are newly available. And finally, the first estimates of margins on foreign currencies are at our disposal. Let us show the results.

6.1 Export and import of margins on securities, shares and investment fund shares

The methodology of estimation of cross-border margins is introduced in Section 4.4. Figure 5 shows export and import of margins on securities, shares and investment fund shares in quarterly time series

(since the first quarter of 2012 to the last quarter of 2017). These are aggregated upon the whole world. These data are also available on the level of other, smaller, supranational units (e.g. EURO-area 19, EU-28 countries and other). It is evident, that in the whole period the balance remains quite stable around CZK –100 mil. Export and import have usually similar course during the whole examined period.

Figure 5 Export, import and their balance of margins on securities, shares and investment fund shares – whole world (in CZK mil.)



Source: Own elaboration from CNB data

6.2 Estimates of domestic margins on securities and shares

These margins are again estimated in association with the CNB. The methodology of computation is based on the same principle and assumptions as in case of cross-border margins (see Section 4.4). We have obtained the preliminary estimates for 4Q2017 and 1Q2018 and the provisional numbers can be subject to changes and investigation. Table 3 points out the aggregated results for financial assets AF.3, AF.511, AF.512 and AF.52.

Table 3 Estimates of domestic margins on securities and shares (in CZK mil.)

Code	Financial asset	4Q2017	1Q2018
AF.3	Debt securities	39.65	28.09
AF.511	Listed shares	1.47	0.81
AF.512	Unlisted shares	6.94	1.63
AF.52	Investment fund shares	0.96	0.56
	Total	49.01	31.10

Source: Own computation from CNB data

6.3 Margins on foreign currencies

First of all we analysed exchange bureaux market. Table 4 summarises results of this analysis. We can see that only almost 50% of exchange bureaux are classified in S.126 (for institutional sectors see Table 1), but their share on purchases as well as sales raises up to 78% or 80%. From these data we could formulate some assumption, e.g. that the producers' sector will be limited only to S.126.

Table 4 Shares of exchange bureaux, purchases and sales according to sector classification in 3Q2017 (in %)

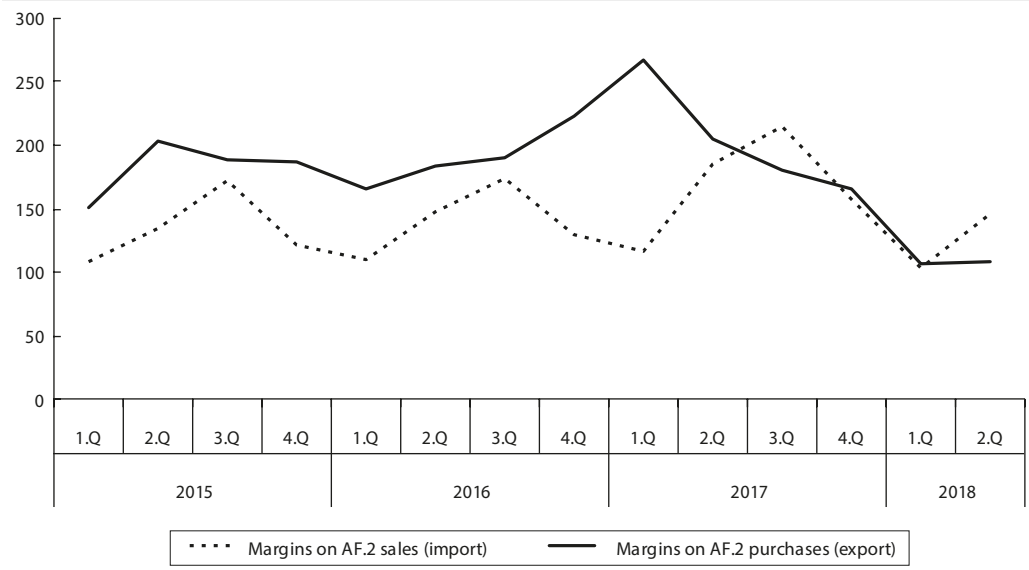
Sector	Exchange bureaux	Purchases	Sales
S.11	35.96	14.48	17.20
S.125	0.44	0.06	0.03
S.126	47.37	78.30	80.17
S.141	7.46	2.15	0.87
S.142	8.77	5.01	1.73

Source: Own computation from CNB and CZSO data

The data for margins on foreign currencies are surveyed by CNB using statistical statement Dev(ČNB) 26-04 entitled "Purchase and sale of foreign currency". By this statement approximately 200 big exchange bureaux are surveyed quarterly. They cover approximately 90% of all cash transactions with foreign currencies.

The preliminary results for quarterly time series (since 1Q2015 to 2Q2018) are in Figure 6 below. They are divided to margins from purchases and margins from sales. Into the computation six main foreign currencies are taken into account: EUR, USD, GBP, CHF, PLN and HRK (as already stated in Section 3.2).

Figure 6 Margins on foreign currencies (in CZK mil.)



Source: Own computation from CNB data

6.4 Importance of margins in the Czech national accounts

From the preliminary results we can assume that the total value of margins in comparison to GDP is not significant in the Czech Republic because it does not even reach 0.1% of GDP in the year 2017. However, under ESA 2010 manual margins form an important part of financial services and must be estimated in national accounts for each country.

7 CAPTURING OF "FINANCIAL" MARGINS IN NATIONAL ACCOUNTS

We have already discussed that on the resources side there are three sectors in which the margins form the part of their output. These are S.122 (trade with residents and non-residents), S.125 (trade with residents) and S.126 (foreign exchange bureaux). We decided to impute margins into the item adjustment for FISIM (P.119) or import of services (P.72).

On the uses side we took our inspiration from the United Nations Statistics Division (UNSD), European Central Bank (ECB) (2014), paragraph 3.128: "The margins should be recorded as the intermediate consumption of corporations, general government or NPISHs if the financial services are provided to these sectors. If they are provided to households, these should be recorded as their final consumption expenditure unless the financial service is provided to an unincorporated enterprise. The margins which are recorded as the intermediate consumption of general government and NPISHs are also recorded in their output and final consumption expenditure, since the institutional units in these two sectors are non-market producers whose output is calculated as the sum of costs. On the other hand, the margins should be recorded in exports of goods and services if the financial services are provided to the rest of the world." It means that the margins are recorded into items intermediate consumption (P.2), individual consumption expenditure (P.31), collective consumption expenditure (P.32) and export of services (P.62).

7.1 Export and import of margins on securities, shares and investment fund shares

When we focus on export and import of margins on tradable financial assets, Table 5 shows resources and uses in the relevant sectors of national accounts.

Table 5 Export, import of margins on securities, shares and investment fund shares in 2017 – in CZK mil.

	Resources	Uses	Value
Export of services	Output (S.122)	Export of services (S.2)	617.239
Import of services	Import of services (S.2)	All relevant sectors according to financial assets	865.294

Source: Balance of payment (CNB)

Uses are divided into relevant sectors according to transactions with financial assets into items intermediate consumption (P.2), individual consumption expenditure (P.31) or collective consumption expenditure (P.32). Apart from changes shown in Table 5 we can emphasize also changes in financial account. Export of margins is added into transaction of S.122 (producer) and import of margins is deducted from transaction value in all other sectors.

7.2 Domestic margins on securities and shares

If we provisionally calculate yearly domestic margins as simply four times 4Q2017, we can again analyse resources and uses (see Table 6).

Table 6 Domestic margins on securities and share in 2017 (in CZK mil.)

	Resources	Uses	Value
AF.3	Output (S.122 and S.125)	All relevant sectors according to financial assets	158.60
AF.511	Output (S.122 and S.125)	All relevant sectors according to financial assets	5.88
AF.512	Output (S.122 and S.125)	All relevant sectors according to financial assets	27.76
AF.52	Output (S.122 and S.125)	All relevant sectors according to financial assets	3.84
Total	Output (S.122 and S.125)	All relevant sectors according to financial assets	196.04

Source: Own computation from CNB data

Resources are divided into S.122 and S.125 according to volume of transactions, the same approach is applied in case of uses into items intermediate consumption (P.2), individual consumption expenditure (P.31) or collective consumption expenditure (P.32). Additionally, in financial account margins are added to transaction volumes in case of margin producers (S.122 and S.125) and excluded from transaction volumes in case of other sectors on uses side.

7.3 Margins on foreign currencies

The only thing which has to be solved is allocation of margins produced in S.126. We have no information about who exchanges. But we suppose that foreign currencies are bought only by S.14, so the whole production coming from these transactions is allocated in their final consumption (FHCE). On the other side we suppose that the foreign exchange bureaux buy foreign currencies only from S.2 and in this case the produced margin is exported. We omit other sectors such as S.11, because we expect, that they are able to reach better exchange rate than S.14 with almost no margin included.

If we take an example with values from our experimental results for 2017, we can see the proper capturing in Table 7.

Table 7 Margins on foreign currencies in 2017 (in CZK mil.)

	Resources	Uses	Value
Purchases of foreign currency	Output (S.126)	Export of services (S.2)	818.6
Sales of foreign currency	Output (S.126)	Individual consumption expenditure (S.14)	672.7

Source: Own computation from CNB data

From the Table 7 we can sum up, that the output of S.126 raises of about +1 492 mil. CZK (819 + 673), export of services in S.2 of about +819 mil. CZK and individual consumption expenditure in S.14 of about +673 mil. CZK. Additionally, changes are made also in financial account; there is an increase in transaction value in currency in S.126 of about +1 492 mil. CZK and decrease in transaction value in currency in S.14 of about –673 mil. CZK and –819 mil. CZK in S.2.

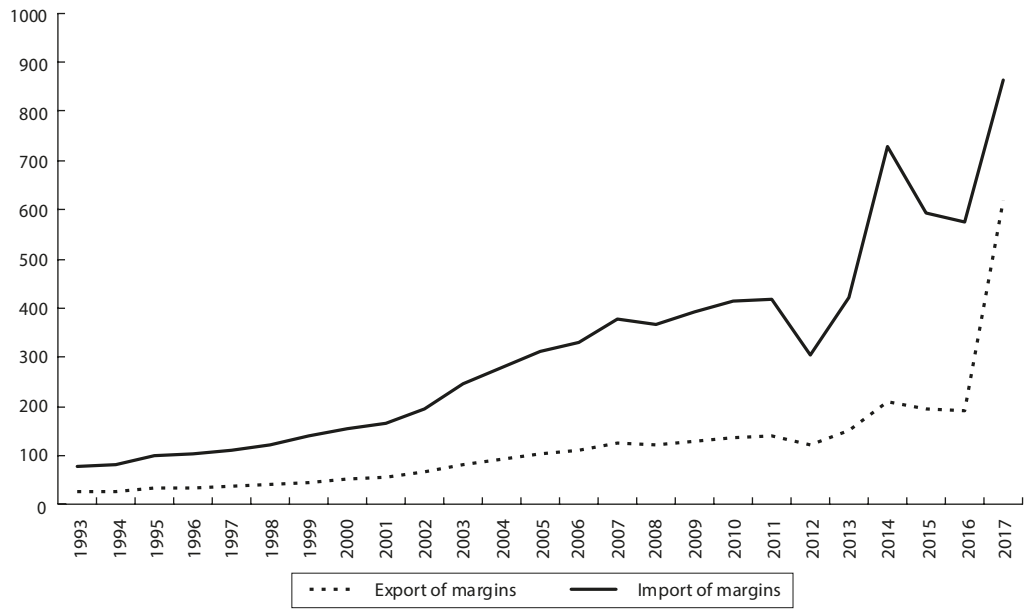
8 DISCUSSION ABOUT RETRAPOLATION OF TIME SERIES BACK TO 1993

At present time (October 2018) estimates for export and import of margins quarterly from 2012 onwards, for domestic margins on securities and shares since 4Q2017 onwards and for margins on foreign currencies from 2015 onwards are available.

There are more possibilities how to make estimates back to the year 1993. We assessed all three introduced types of margins and tried to choose the best method for each of them. This part of our paper is still subject to discussion and future possible changes. Final data will be available in June 2020.

Concerning export and import of margins, the current state of our research prefers to apply average share of export or import margins (from available years) either on transactions or stocks, and to retrapolate the values back to the year 1993. The transactions are more volatile whereas stocks are successively rising in the whole time series. An experimental estimate of export and import of margins based on stock is shown in Figure 7.

Figure 7 Experimental estimate of export and import of margins based on stocks (in CZK mil.)

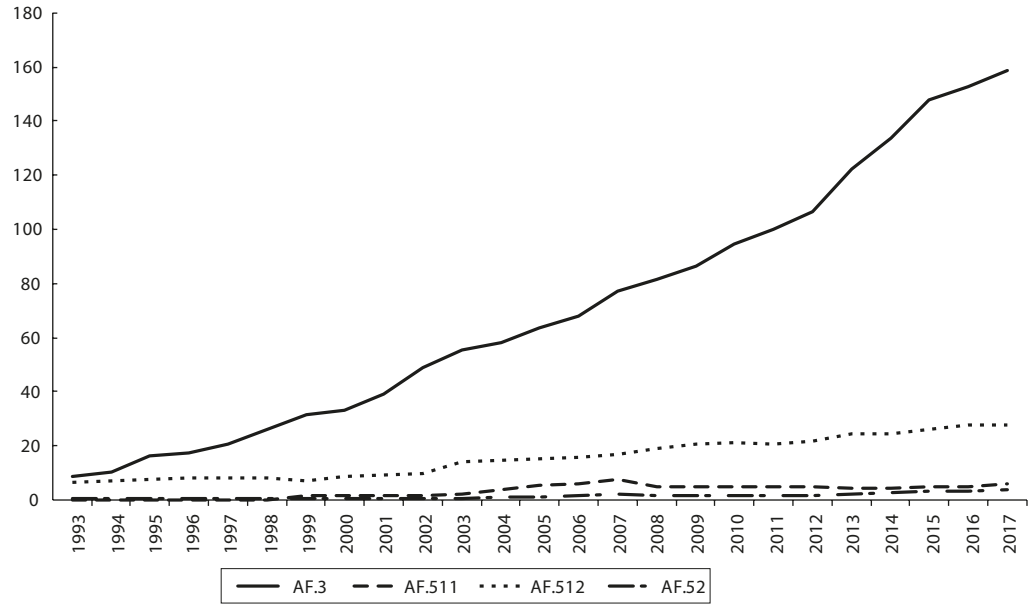


Source: Own computation from CNB data

If we turn to domestic margins on securities and shares, we are willing to apply the same procedure as in case of export and import of margins. We have again two possibilities – to retrapolate the values according to stocks or transactions. Analogously, transactions are much more volatile than the stocks. In this case there is one additional problem, some transactions (especially with shares) are negative, i.e. they would produce negative margins. Figure 8 shows experimental estimate of time series of domestic margins on securities and shares divided into the individual items based on stocks.

As for margins on foreign currencies, there will be no available information for previous years as the computation is done using exchange rates in a reference foreign exchange bureau. We decided to use an average of available data and retrapolate them back to 1993. Some inflation coefficient from CPI statistics will probably be used.

Figure 8 Experimental estimate of domestic margins on securities and shares based on stocks (in CZK mil.)



Source: Own computation from CNB data

CONCLUSION

Margins form an essential part of national accounts. As in the Czech Republic, so far, they are not recorded completely, this paper brings a methodological insight in this task. It summarises available theory and practical issues connected with estimation of margins as part of financial services, which are hidden in transactions with financial assets. At the beginning the role of these margins in the system of national accounts is introduced. We present various methodological approaches that we discussed and final methodology that will be applied on estimation of margins.

Apart from methodological questions we discussed also the difficulties and obstacles connected mainly with the data. One of the most important is the non-existence of overall guideline for computation of this kind of margins that is why we try to find the best methodological way for the Czech Republic. Some examples of practice from other countries and institutions are included as well.

Despite the aforementioned limitations, the paper contains the first ever preliminary results estimated for the Czech Republic as well as assumptions that should be formulated and taken into account. Section 6 is divided into three independent parts; each of them dealing with different types of margins in financial services and showing our attempt to their calculation. The preliminary results could be subject to a change and further obtained knowledge will be incorporated. Section 7 shows the practical capturing of "financial" margins in the whole system of national accounts, i.e. which items will be adjusted. Section 8 introduces our very first attempts of retrapolation back to the year 1993. The final results will be included in national accounts by June 2020 for period of 1993–2019 as a result of benchmark revision.

Last but not least, this paper is prepared under the support of EUROSTAT and is limited by rules of official statistics. Some research can naturally be done, but estimates must correspond to manuals that are crucial for NSIs.

ACKNOWLEDGMENTS

The paper has been prepared under the European Commission EUROSTAT grant; project No. 131703 Action 3 "Improvement of GNI data" (Agreement No. 04121.2017.001-2017.255). Authors also thank the Czech Statistical Office for financial support.

References

ALFRESCO CONTENT REPOSITORY. *Directory listing for GNI Inventories – public versions* [online]. 2018. [cit. 24.10.2018]. <https://circabc.europa.eu/webdav/CircaBC/ESTAT/GNI%20OR/Library/ESA%202010%20GNI%20Inventories%20-%20public%20versions>.

BROUSSOLLE, D. The Emerging New Concept of Services in the System of National Accounts and the Balance of Payments. *Review of Income & Wealth*, 2015, 3, pp. 573–586.

CZECH MINISTRY OF FINANCE. *Zápis z jednání Koordinačního výboru s Komorou daňových poradců ČR konaného dne 27. 11. 2007* [online] (in Czech). Prague: Czech Ministry of Finance, 2007. [cit. 03.10.2017]. <http://www.financnisprava.cz/assets/cs/prilohy/d-prispevky-kv-kdp/KV27112007.pdf>.

EUROPEAN COAL AND STEEL COMMUNITY, EUROPEAN COMMUNITIES, EUROPEAN ATOMIC ENERGY COMMUNITY. *European system of accounts ESA 1995.* Luxembourg: Office for Official Publications of the European Communities. 1996.

EUROPEAN COMMISSION, EUROSTAT. *Final report on the cross-country comparison. Financial services (including FISIM).* Internal materials for 36th Meeting of the GNI Committee held 25–26 April 2018 in Luxembourg, 2018.

EUROPEAN COMMISSION, EUROSTAT. *European system of accounts ESA 2010*. Luxembourg: Publications Office of the European Union. 2013.

EUROPEAN COMMISSION, INTERNATIONAL MONETARY FUND, ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT, UNITED NATIONS, WORLD BANK. *System of National Accounts 2008*. New York: United Nations. 2009.

FUNGÁČOVÁ, Z. AND POGHOSYAN, T. Determinants of bank interest margins in Russia: Does bank ownership matter? *Economic Systems*, 2011, 4, pp. 481–495.

GIOVANNINI, E. *Understanding economic statistics: an OECD perspective*. Paris: Organisation for Economic Cooperation and Development. 2008.

IDNES.CZ. *Euro za 15 korun? Směnárna Chequepoint po odnětí povolení od ČNB končí* [online] (in Czech). 2018. [cit. 31.10.2018]. < https://ekonomika.idnes.cz/smenarna-chequepoint-praha-cnb-pokuta-zakaz-cinnosti-pdj-/ekonomika.aspx?c=A180611_094915_ekonomika_mato>.

INTERNATIONAL MONETARY FUND. *Balance of Payments and International Investment Position Manual. Sixth Edition (BPM6)* [online]. Washington, D.C.: International Monetary Fund, 2009. [cit. 26.10.2018]. <https://www.imf.org external/pubs/ft/bop/2007/pdf/bpm6.pdf>.

KRAMULOVÁ, J. AND MUSIL, P. Experimentální odhad složek výdajové metody regionálního HDP v ČR. (Experimental Estimate of Components of Expenditure Approach to Regional GDP in the Czech Republic. With English summary.). *Politická ekonomie*, 2013, 6, pp. 814–833.

RYBÁČEK, V. Theory and practice of holding gains and losses, is importance of revaluation reflected in the national accounts? *Statistika*, 2010, 6, pp. 506–516.

STREICHER, G. AND STEHRER, R. Whither Panama? Constructing a Consistent and Balanced World SUT System Including International Trade and Transport Margins. *Economic Systems Research*, 2015, 2, pp. 213–237.

TAKEDA, H. *Margins on buying and selling transactions – Japan's case* [online]. 2013. [cit. 18.04.2018] <https://tinyurl.com/ycfuueeg>.

UNITED NATIONS STATISTICS DIVISION (UNSD), EUROPEAN CENTRAL BANK (ECB). *Financial Production, Flows and Stocks in the System of National Accounts.* New York: United Nations, 2014.

VLTAVSKÁ, K. AND SIXTA, J. A Historical View on the Development of Czech Economy from 1970. *Prague Economic Papers*, 2015, 1, pp. 105–122.

A New Clipping Approach for Robust ACF Estimation

Samuel Flimmel [1] | *University of Economics in Prague, Prague, Czech Republic*
Ivana Malá | *University of Economics in Prague, Prague, Czech Republic*
Jiří Procházka | *University of Economics in Prague, Prague, Czech Republic*
Jan Fojtík | *University of Economics in Prague, Prague, Czech Republic*

Abstract

The importance of working with sufficiently robust methods has been rising in recent years. This growth is related to the extensive usage of highly frequent data, which we currently encounter in many fields including finance. Since with an increasing number of observations, the probability of outlier presence also rises. Moreover, as it is known, standard methods are not able to work correctly with outliers and, consequently, standard estimates are often biased. We focus on estimators of autocorrelation function for univariate time series, for which we propose a method based on clipping an original time series and working with a binary time series instead. The clipping helps to deal with outliers and the estimation is not affected as much as with standard methods. We also derive an asymptotical distribution of the estimator, what gives our method a major advantage in comparison with other robust methods, which are often presented without this. Furthermore, knowing the distribution of the estimator allows us to perform statistical inference.

Keywords

ACF, robust estimation, clipping, confidence interval, time series

JEL code

C10, C22

INTRODUCTION

The autocorrelation function (ACF) expresses the correlation among observations of the time series. It plays an important role in time series theory because it partially describes the relationship among the observations of the series. Furthermore, it gives us an overview of the time series, and we can use it to investigate or to model the time series.

Estimation of the ACF can be negatively affected by many factors. One of them is an outlier presence, very relevant nowadays, when we face many problems related to the extensive usage of big data. There are several robust methods for ACF estimation designed to be able to take account of outliers. These methods should, naturally, be less sensitive to outliers and should lead to better results in general (Chan and Wei, 1992).

There has been a plethora of approaches proposed by many authors. Chakhchoukh (2010) presented a median approach, where he suggested to use the median instead of the mean in the standard estimator. Ma and Genton (2000) proposed a Gnanadesikan-Kettenring approach based on the special relationship

[1] Department of Statistics and Probability, Faculty of Informatics and Statistics, University of Economics in Prague, Nám W. Churchilla 4, Prague 3, Czech Republic. Corresponding author: e-mail: samuel.flimmel@vse.cz.

for an autocovariance function (Gnanadesikan and Kettenring, 1972), while Maronna et al. (2006) proposed a robust filtering approach. Dürre et al. (2015) presented a useful overview of robust ACF estimators, including those we do not mention specifically. However, the above mentioned estimators are almost always presented without specifying the distribution. Therefore, we are not able to test the significance of the ACF order, or to test hypotheses related to the ACF.

In this paper, we present a new approach for ACF estimation. This approach is based on clipping, i.e. the process, when we replace original observations by zeros when they are below a given threshold, resp. by ones when they are above. We do not only construct the ACF estimator, but we also derive the asymptotical distribution of the ACF estimator. Using the distribution of the estimator, we suggest an analogous approximation to Bartlett's approximation (Bartlett, 1946), which is used to determine the ACF order significance.

We apply the proposed clipping approach in a simulation study in order to compare its results with a standard sample estimator. Finally, both approaches are used in a study with real world data, where we investigate the behavior of the 1-year (1Y) historical volatility of Bitcoin logarithm of daily returns.

The methodology of the clipping approach is presented in Section 1. The distribution of the estimator is derived in Section 2. The simulation study is presented in Section 3. The real data study is presented in Section 4. The last section includes conclusions of our study.

1 METHODOLOGY

Let $\{Z_n, n \in N_0\}$ be a stationary time series. Then we can define an autocovariance function of the lag k, $k \in Z$, $R(k)$ as

$$R(k) = E(Z_k - \mu)(Z_0 - \mu), \tag{1}$$

where μ is the expected value of the process.

We define an autocorrelation function (ACF) of the lag k, $k \in Z$, $\rho(k)$ of the stationary process $\{Z_n, n \in N_0\}$ as

$$\rho(k) = \frac{R(k)}{\sigma^2}, \tag{2}$$

where σ^2 is the variance of the time series.

We define a sample autocorrelation function of the lag k, $k \in Z$, $\hat{\rho}(k)$, of the data $Z_0, ..., Z_m$ as

$$\hat{\rho}(k) = \frac{\sum_{n=k}^{m} (Z_n - \bar{Z})(Z_{n-k} - \bar{Z})}{\sum_{n=0}^{m} (Z_n - \bar{Z})^2}, \tag{3}$$

where $\bar{Z}$ is the mean of the data.

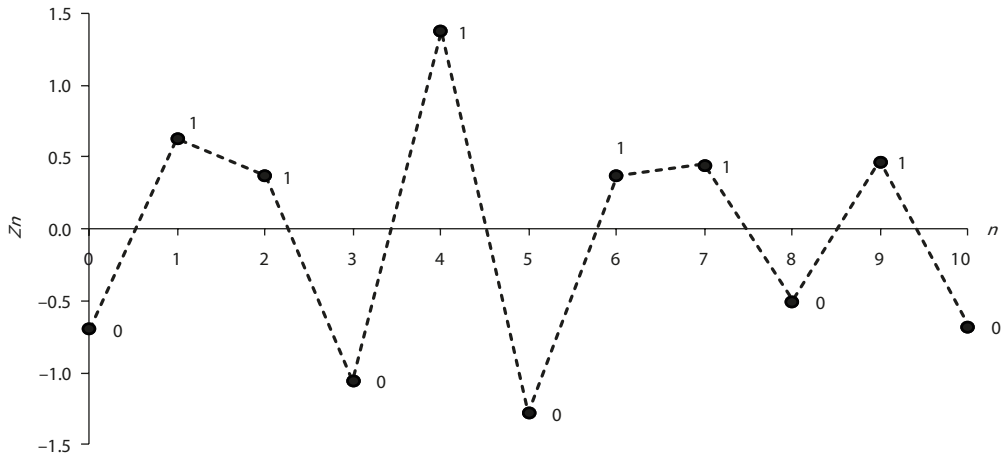
Let $\{Z_n, n \in N_0\}$ be a strictly stationary time series with autocorrelation function $\rho_Z(k)$. For a fixed h, $h \in R$, a so called threshold, we define $\{X_n, n \in N_0\}$ followingly:

$$X_{n,h} = \begin{cases} 1, Z_n \geq h, \\ 0, Z_n < h. \end{cases} \tag{4}$$

A time series $\{Z_n, n \in N_0\}$ is called original and $\{X_{n,h}, n \in N_0\}$ is known as clipped or hard-limited. Let us denote an autocorrelation function of this clipped time series by $\rho_X(k)$.

In Figure 1, we can see an illustrative example of clipping the original time series $\{Z_n, n \in N_0\}$ at the threshold $h = 0$.

Figure 1 Illustrative example of clipping at the threshold h = 0



Source: Own construction

In Kedem (1980) it was proved, under the assumption of a zero threshold ($h = 0$) and a zero mean ($\mu = 0$) strictly stationary Gaussian original time series $\{Z_n, n \in N_0\}$, that

$$\rho_X(k) = \frac{2}{\pi} \arcsin \rho_Z(k), k \in Z. \tag{5}$$

Easily, it can be rewritten into

$$\rho_Z(k) = \sin\left(\frac{\pi}{2} \rho_X(k)\right), k \in Z. \tag{6}$$

Using (6) we can construct a new robust ACF estimator $\tilde{\boldsymbol{\rho}}_{\boldsymbol{Z}}(\boldsymbol{k})$. The construction can be divided into the following steps:

1. Derive a clipped ($\mathbf{h = 0}$) time series $\{\boldsymbol{X}_{n,0}, \boldsymbol{n} \in \boldsymbol{N}_0\}$ from the original time series $\{\boldsymbol{Z}_n, \boldsymbol{n} \in \boldsymbol{N}_0\}$ which is of our interest.
2. Calculate a sample $\hat{\boldsymbol{\rho}}_{\boldsymbol{X}}(\boldsymbol{k})$ from the clipped time series $\{\boldsymbol{X}_{n,0}, \boldsymbol{n} \in \boldsymbol{N}_0\}$.
3. Calculate an estimation $\tilde{\boldsymbol{\rho}}_{\boldsymbol{Z}}(\boldsymbol{k})$ from the $\hat{\boldsymbol{\rho}}_{\boldsymbol{X}}(\boldsymbol{k})$ using Formula (6).

The clipping in step 1 helps to face outliers. It is similar to widely used trimming methods, however, the loss of information has a different nature.

2 DISTRIBUTION OF THE ESTIMATOR $\tilde{\boldsymbol{\rho}}_{\boldsymbol{Z}}(\boldsymbol{k})$

Bartlett's approximation (Bartlett, 1946) is frequently used for determination of the ACF order significance.

It can be formulated as

$$\hat{\rho}_Z(k) \overset{as.}{\sim} N\left(\rho_Z(k), \frac{v_k^Z}{m}\right), \tag{7}$$

where:

$$v_k^Z = \sum_{i=1}^{\infty} \left(\rho_Z(i+k) + \rho_Z(i-k) - 2\rho_Z(i)\rho_Z(k) \right)^2 \tag{8}$$

and m is the number of observations.

Analogously, we can formulate a similar approximation of our estimator

$$\tilde{\rho}_Z(k) \overset{as.}{\sim} N\left(\rho_Z(k), \frac{\pi^2}{4m} \cos^2\left(\frac{\pi}{2} \rho_X(k) \right) v_k^X \right), \tag{9}$$

where:

$$v_k^X = \sum_{i=1}^{\infty} \left(\rho_X(i+k) + \rho_X(i-k) - 2\rho_X(i)\rho_X(k) \right)^2 \tag{10}$$

and m is the number of observations.

We can prove Formula (9) using Bartlett's approximation and Delta method (Greene, 2003). Bartlett's approximation for the clipped time series $\{X_{n,0}, n \in N_0\}$ yields the following:

$$\hat{\rho}_X(k) \overset{as.}{\sim} N\left(\rho_X(k), \frac{v_k^X}{m} \right). \tag{11}$$

Delta method is a result concerning the asymptotic distribution of the transformed random variable in a specific situation. If there is a time series $\{Z_n, n \in N_0\}$ satisfying:

$$\sqrt{n}(Z_n - \theta) \xrightarrow{D} N(0, \sigma^2), \tag{12}$$

where θ and σ^2 are finite valued constants and $\xrightarrow{D}$ denotes convergence in distribution, then

$$\sqrt{n}\left(g(Z_n) - g(\theta) \right) \xrightarrow{D} N\left(0, \sigma^2 \left(\frac{d}{dx} g(\theta) \right)^2 \right), \tag{13}$$

for any function $g(x)$ satisfying that $\frac{d}{dx} g(\theta)$ exists and is non-zero valued.

Finally, if we set $g(x) = \sin\left(\frac{\pi}{2} x \right)$, then Delta method gives us approximation (9), because

$$\frac{d}{dx} g(x) = \frac{d}{dx} \sin\left(\frac{\pi}{2} x \right) = \frac{\pi}{2} \cos\left(\frac{\pi}{2} x \right), \tag{14}$$

$$\left(\frac{d}{dx} g\left(\rho_X(k) \right) \right)^2 = \frac{\pi^2}{4} \cos^2\left(\frac{\pi}{2} \rho_X(k) \right). \tag{15}$$

Easily, we can show equivalence (16), which we use later to prove statement (18),

$$\rho_X(k) = 0 \Leftrightarrow \rho_Z(k) = 0. \tag{16}$$

Since $\rho_Z(k) = 0$ only if $\rho_Z(k) = 0$, and $\sin\left(\frac{\pi}{2}\rho_X(k)\right) = 0$ only if . So it comes from Formulas (5) and (6).

Similarly, we use a special case of equation (10) for $\rho_X(k) = 0$ for $k > k_0$. We have:

$$v_k^X = \sum_{i=1}^{\infty}\left(\rho_X(i+k)+\rho_X(i-k)-2\rho_X(i)\rho_X(k)\right)^2 = \sum_{i=1}^{\infty}{\rho_X}^2(i-k) = \sum_{i=k-k_0}^{k+k_0}{\rho_X}^2(i-k) = \sum_{j=-k_0}^{k_0}{\rho_X}^2(j). \quad (17)$$

To determine the significant order of an ACF by our estimator, we use the following statement, which is proved by Formulas (16) and (17). If $\rho_Z(k) = 0$ for $k > k_0$, then

$$\tilde{\rho}_Z(k) \overset{as.}{\sim} N\left(0, \frac{\pi^2}{4m}\sum_{i=-k_0}^{k_0}{\rho_X}^2(i)\right), k > k_0. \quad (18)$$

So, we would look for k_0 that holds

$$\left|\tilde{\rho}_Z(k)\right| \geq u_{1-\alpha}\sqrt{\frac{\pi^2}{4m}\sum_{i=-k_0}^{k_0}\rho_X(i)}, k > k_0, \quad (19)$$

where $u_{1-\alpha}$ is a $(1 - \alpha)$% quantile of the standard Gaussian distribution and α is a significance level.

3 SIMULATION STUDY

In the presented simulation study, we compare our estimator with a standard sample estimator. We use MA(q) time series $\{Z_n, n \in N_0\}$:

$$Z_n = \varepsilon_n + \theta_1\varepsilon_{n-1} + \theta_2\varepsilon_{n-2} + \cdots + \theta_q\varepsilon_{n-q}, \quad (20)$$

where $\{\varepsilon_n, n \in N_0\}$ and $\theta_1, \theta_2, \ldots, \theta_q$ are parameters of the time series.

The simulation study was designed in the R software (R Core Team, 2015).

We run 10 000 simulations with 1 000 observations, which we contaminate with additive outliers (Fox, 1972).

In the additive outlier (AO) model, we assume that we do not observe the process of interest $\{Z_n, n \in N_0\}$ but, actually, we observe a process $\{Y_n, n \in N_0\}$ defined as

$$Y_n = Z_n + O_n, \quad (21)$$

where processes $\{Z_n, n \in N_0\}$ and $\{O_n, n \in N_0\}$ are assumed to be independent of one another.

Let $\{O_n, n \in N_0\}$ be a process with independent and identically distributed (i.i.d.) random variables that have a normal mixture distribution with a degenerate central component:

$$O_n \sim (1-\varepsilon)\delta_0 + \varepsilon N\left(\mu_O, \sigma_O^2\right), \quad (22)$$

where δ_0 is the point mass distribution located at zero, and we assume that the normal component $N\left(\mu_O, \sigma_O^2\right)$ has a variance significantly higher than the process $\{Z_n, n \in N_0\}$, $\sigma_O^2 \gg \sigma_Z^2$.

The probability of outlier occurrence is represented by ε, which is usually small. Consequently, the probability of occurrence of two outliers in a row is a much smaller ε^2, which means that the AO model generates mostly isolated outliers.

The percentage of outliers present in a single simulation is chosen randomly with a uniform distribution, i.e. $\varepsilon \sim U\left([0.00, 0.05]\right)$. We use the outlier standard deviation $\sigma_O = 10$.

The absolute values of the parameters of the MA(q) are generated randomly with a uniform distribution, i.e. $\theta_i \sim U([0.2,1.0])$, $i = 1,2, \ldots, q$. Values of θ_i being close to zero are not taken into account because they are difficult to observe. The sign of the parameters is generated randomly with Bernoulli's distribution with the probability of a success $p = 0.5$.

We divide our simulation study into 2 parts. In the first part, we work with MA(1), MA(2) and MA(3) times series, where we estimate ACF of the series and compare the methods using mean average error (MAE) criterion.

In the second part, we work with MA(3), which have theoretically significant ACF to 3rd order and we estimate the highest significant order according to the methods. In both parts we use $h = 0$ for the clipping method.

For the first part, we run 10 000 simulations for every model, so together 30 000 simulations. We obtain results summarized in Table 1.

Table 1 Comparison of the standard sample estimator and the clipping approach estimator in point estimation in the simulation study

MA(q)	ρ(k)	Standard approach	Clipping approach
MA(1)	$\rho(1)$	0.2258	0.0361
MA(2)	$\rho(1)$	0.1674	0.0368
	$\rho(2)$	0.1679	0.0396
MA(3)	$\rho(1)$	0.1468	0.0373
	$\rho(2)$	0.1365	0.0399
	$\rho(3)$	0.1277	0.0414

Source: Own construction

In Table 1, we can see that our clipping method gives better results for every model and every order of autocorrelation function. It is caused by the bias of the standard method (Maronna et al., 2006).

For the second part we use a significant level $\alpha = 0.05$ and we obtain results summarized in Table 2.

Table 2 Comparison of the standard sample estimator and the clipping approach estimator in significant order in the simulation study.

Significant order	Standard approach	Clipping approach
1	2.94%	0.22%
2	10.01%	5.48%
3	70.26%	77.48%
4	3.83%	3.65%
5	3.98%	4.11%
6	4.38%	4.13%
7 and more	4.60%	4.93%

Source: Own construction

In Table 2, we can see how many simulations have the highest significant order of ACF at presented values 1, 2, ..., 6, 7 and more . We see that the standard approach tends to underestimate the significant order more than the robust estimator based on clipping. It is caused by outliers which weaken correlation between neighbor observations (Maronna et al., 2006), i.e. autocorrelation function.

4 REAL DATA APPLICATION

We investigate a daily time series of Bitcoin (the most used crypto currency) log returns. We have data (close prices) from *Yahoo Finance* from the period from 16/07/2010 to 18/08/2018 (2 956 observations). Usually, log returns are investigated instead of the original rates, because of a non-stationarity and a high autocorrelation, which could be misleading. Firstly, we define a log return:

$$r_n = \log\left(\frac{p_n}{p_{n-1}}\right), n > 1, \tag{23}$$

where p_n is a price for n-th observation.

Then we calculate the 1-year historical volatility from the log returns:

$$\sigma_n^{1Y} = \sqrt{\frac{1}{364}\sum_{i=n-364}^{n}\left(r_i - \overline{r_n}\right)^2}, n > 365, \tag{24}$$

where:

$$\overline{r_n} = \frac{1}{365}\sum_{i=n-365}^{n} r_i, n > 365 \tag{25}$$

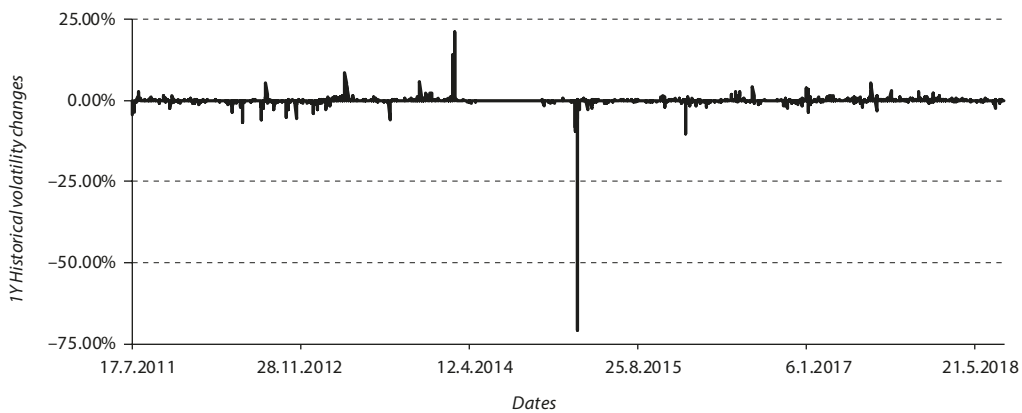
and 365 is the number of days within a year.

The point of our interest is the change (delta) of volatility, so we have to define the logarithmic change of the 1Y historical volatility:

$$\Delta_n^{1Y} = \log\left(\frac{\sigma_n^{1Y}}{\sigma_{n-1}^{1Y}}\right), n > 366. \tag{26}$$

Our logarithmic changes of the 1Y historical volatility are displayed in Figure 2.

Figure 2 Logarithmic changes of the yearly historical volatility of Bitcoin log returns

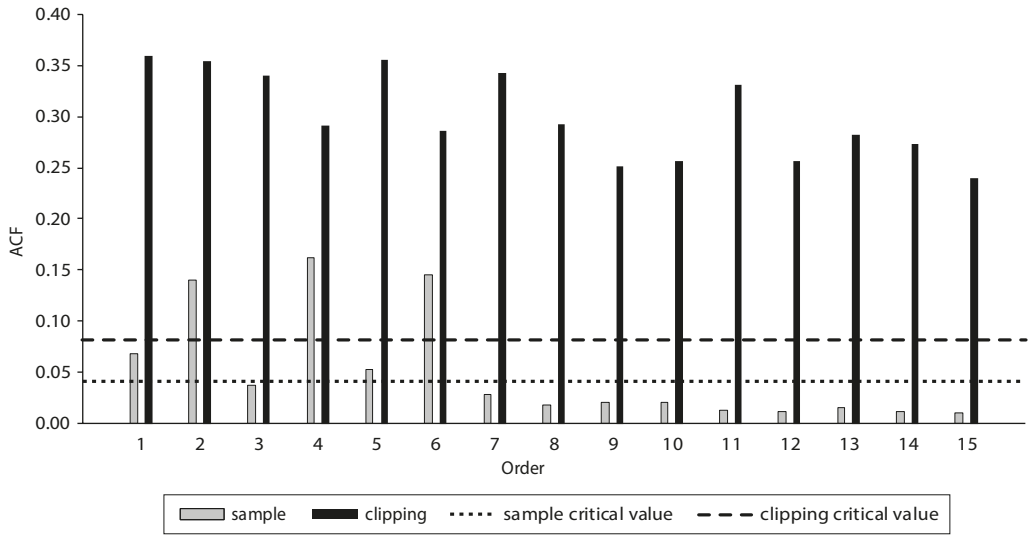


Source: Own construction

We can see a few isolated outliers in Figure 2. The biggest one is from 26/02/2015 and it is caused by losing a high log return from 26/02/2014, which was caused by price change from $135.78 to $593.14. The presence of these outliers could lead to a weaker autocorrelation.

Autocorrelation functions for standard sample estimator and clipping approach estimator are shown in Figure 3. It contains also critical values for a significant level $\alpha = 0.05$ for both estimators.

Figure 3 Comparison of the standard sample ACF estimation and the clipping approach ACF estimation with critical values



Source: Own construction

We can see the significance to 6th order of ACF for standard sample method. On the other hand, the clipping approach shows much higher correlations (with maximum 0.35 and average 0.3 over first 15 orders) and it is significant to the last presented order. The standard sample method could mislead us to a lower order of ACF and a weaker autocorrelation (with maximum 0.15 and average 0.05 over first 15 orders), but in reality, we should consider higher lags, or, alternatively, we could try to model the time series using AR(p) and have satisfying results.

DISCUSSION AND CONCLUSION

We have constructed a new robust ACF estimator based on clipping. Furthermore, we presented the asymptotical distribution of the estimator. We consider the knowledge of the distribution as a major advantage in comparison with other robust estimators, since it allows us to investigate the significance of the ACF orders or to test relevant hypotheses that occur when solving particular problems.

In Section 3, we have designed a simulation study, where we have compared the clipping approach estimator with the standard sample estimator using data contaminated with additive outliers. Firstly, we have compared the point estimates of the methods and our proposed clipping approach method has given better results for all cases. Secondly, we have exploited the knowledge of the clipping approach estimator and the standard sample estimator, the distribution of which is well known. The simulation study has shown us the underestimation of the standard approach in comparison with the clipping approach. This has confirmed our expectations, since additive outliers should weaken the relationship between neighboring observations.

Finally, we have compared both approaches in a study with real world data, where we have investigated the logarithmic change of 1Y historical volatility of Bitcoin log returns. The standard approach has suggested a weaker autocorrelation. We tend to trust more the clipping approach, since it has shown a stronger autocorrelation, which could lead us even to AR(p) time series.

In conclusion, we suggest to use robust methods in the case of additive outliers. On the other hand, for innovative outliers, it may be better to use the standard approach (Flimmel et al., 2017). If there is a need to know the distribution of the estimator, we definitely recommend the clipping approach estimator.

ACKNOWLEDGMENTS

This study was supported by the grant F4/17/2017 (Robustnosť v úmyselnom useknutí časového radu), which has been provided by the Internal Grant Agency of the University of Economics in Prague.

References

BARTLETT, M. S. On the Theoretical Specification and Sampling Properties of Autocorrelated Time-Series. *Supplement to the Journal of the Royal Statistical Society*, 1946, 8(1), pp. 27–41.

DÜRRE, A., FRIED, R., LIBOSCHIK, T. Robust estimation of (partial) autocorrelation. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2015, 7(3), pp. 205–222.

FLIMMEL, S., ČAMAJ, M., MALÁ, I., PROCHÁZKA, J. Comparison of Robust Methods for ARMA Order Estimation. In: *Applications of Mathematics and Statistics in Economics (AMSE 2017)*, 2017, pp. 133–144.

FOX, A. J. Outliers in time series. *Journal of the Royal Society*, 1972, 34(3), pp. 350–363.

GNANADESIKAN, R. AND KETTENRING, J. R. Robust estimates, residuals, and outlier detection with multiresponse data. *Biometrics*, 1972, 28, pp. 81–124.

GREENE, W. H. *Econometric Analysis, 5th Edition*. Prentice-Hall, New Jersey, 2003.

CHAKHCHOUKH, Y. A new robust estimation method for ARMA models. *IEEE Transactions on Signal Processing*, 2010, 58(7), pp. 3512–3522.

CHAN, W.-S. AND WEI, W. W. A comparison of some estimators of time series autocorrelations. *Computational statistics & data analysis*, 1992, 14(2), pp. 149–163.

KEDEM, B. *Binary Time Series. Lecture Notes in Pure and Applied Mathematics. Vol. 52.* New York and Basel: Marcel Dekker, 1980.

MA, Y. AND GENTON, M. Highly robust estimation of the autocovariance function, *Journal of time series analysis*, 2000, 21(6), pp. 663–684.

MARONNA, R., MARTIN, D., YOHAI, V. *Robust Statistics: Theory and Methods*. Chichester: John Wiley & Sons, 2006.

R CORE TEAM. *R: A Language and Environment for Statistical Computing* [online]. R Foundation for Statistical Computing, Vienna, Austria, 2015. <https://www.r-project.org>.

Website Hosting Data and Analysis

Petr Ilgner [1] | *Brno University of Technology, Brno, Czech Republic*
Dan Komosný [2] | *Brno University of Technology, Brno, Czech Republic*
Saeed Ur Rehman [3] | *Auckland University of Technology, Auckland, New Zealand*

Abstract

We have collected a large dataset – more than 21 000 websites – through web-crawling the public resources of the Czech Internet. The proposed method for website hosting detection along with their geographic location and software were applied on the collected data to extend basic statistical information about the Czech websites published by the national domain registrar CZ.NIC. For analysis, we divided the data into nine categories to show differences between them, for example, between the public and private sector. The procedures used in this paper may also be applied for an extended analysis of websites in other countries, for example, for verification of fulfillment of legal directives to be implemented by public sector.

Keywords

Internet, web content, hosting, geographical location, Czech Republic, CZ.NIC

JEL code

C63, C80, L86

INTRODUCTION

Statistical data about the Internet in a country are used for various purposes. An example may be verification of fulfillment of legal directives issued by a government to be implemented by governmental institutions, public sector entities, and Internet Service Providers (ISP). From the user's point of view, the data may be also used for checking the shared resource plans as published by web hosting service providers.

In various countries, there are national domain registrars that publish statistical data about the national Internet. These data may provide information about domains, DNSSEC, DNS traffic, IPv6, registries, etc. The data are typically published as 'open access'.

Czech registrar CZ.NIC (2017, CZ Domain Hosting Statistics) presents data about domains remotely hosted by particular hosting providers (further referred to as 'hosted'). The data are divided into three sets: i) domains at the organization premises (further referenced as 'self-hosted'), ii) domains hosted at other organization, and iii) domains with unknown hosting status. If a hosting provider is not listed, the owners can report their data to the registrar. The hosting data are specifically categorized to mail hosting, nameserver hosting, and web hosting. The web hosting is of primary importance as it reflects the situation for end-users accessing the web content.

[1] Department of Telecommunications, Faculty of Electrical Engineering and Communication, Brno University of Technology, Technická 12, 612 00 Brno, Czech Republic. E-mail: petr.ilgner@vut.cz, phone: (+420)541146927.
[2] Department of Telecommunications, Faculty of Electrical Engineering and Communication, Brno University of Technology, Technická 12, 612 00 Brno, Czech Republic. E-mail: komosny@feec.vutbr.cz, phone: (+420)541146973.
[3] Department of Electrical and Electronic Engineering, Auckland University of Technology, 34 St. Paul St., Auckland, 1010, New Zealand. E-mail: saeed.rehman@aut.ac.nz.

In this paper, we present these statistics about websites: i) website hosting data, ii) specific data for the defined website categories by their content, iii) geographical related data, and iv) data about security implementations and used software.

Our data come from more than 21 000 websites that we have collected by crawling public resources of the Czech Internet, including the lists of web addresses in public company directories at www.firmy.cz, www.sreality.cz, www.toplist.cz.

The details about the data presented in this paper are the following:

i) Web hosting detection is a complex process as there is no direct (straightforward) approach to identify a website to be self-hosted or hosted. Therefore, we propose a method consisting of four particular tests. These tests aim to identify the hosting status based on information 'clues' that can be obtained from public resources. We consider these information pieces: 'reverse domain lookup', 'database of web hosting providers', 'network owner name', and 'network administrator email'. Each of this information is assigned a weight to calculate the final hosting status.
ii) The data available from CZ.NIC (2017, CZ Domain Hosting Statistics) are global numbers with no particular information about the websites and the entities they represent in their content. Therefore, we define nine entity categories as follows: banks, e-shops, hospitals, insurance companies, real estate agencies, craftsmen, government institutions, secondary schools, and universities. For each category we detect whether the entity website is self-hosted or hosted. Additionally, we show the share (or popularity) of web hosting providers across the categories.
iii) The base data do not cover the geographical distribution of the servers. Therefore, we relate the data to the location of web servers. We show the numbers for the Czech regions, cities and we list countries hosting websites with Czech content.
iv) Finally, we include the data about security implementations in each of the defined categories and give statistic about the software used for running the web.

The procedures used in the paper could be used for various purposes, including market analysis or motivation for websites hosting improvements in terms of reducing load on Internet resources. The latter one is of particular importance as previous research showed that communication latency in company web pages access has a correlation with the revenues (Sigla et al., 2014). Therefore, the decision of self-hosting or hosting, including selection of the hosting provider and its geographical location may be important. Also, the shared web hosting plans can be verified using the described method with large input data (number of websites).

The paper is structured as follows: Section 1 indicates how CZ.NIC obtains the hosting data and discusses the results they publish. Their results are compared with other sources. Related papers to this work are described, mainly considering the hosting status check. In Section 2 we describe in detail our approach to detect the web hosting status. The examples are given for each particular test, including the source of the input data. We also show how we obtained the geographical data and other related information. Description of the implementation, including web crawling, and the detailed numbers about the collected websites are given in Section 3. Section 4 discusses the results and it is divided into particular subsections according to the data type.

1 RELATED WORK AND DATA

Wang et al. (2011) proposed a method for IP geolocation that included identification of website hosting status. They assumed that the same IP address is used for a set of websites (domain names) hosted at the same provider (possible in the order of hundreds). On the other hand, if a specific IP address is used for a single website then it is concluded that such site is very likely to be self-hosted. They detected the web hosting status by accessing a website by its domain name and by its IP address. The returned homepage was checked according to these three suggested options i) its content, ii) head information

(<head></head>), and iii) title information (<title></title>). If this information was equal (based on the selected option), they will conclude the IP address represented a single website, i.e. the site was self-hosted. The different information may be a blank page or error message. They also stated a problem with this method when the first request is redirected. In this case, they sent an additional request for the targeted page.

Tsou et Lusher (2015) grouped the websites into twelve categories, such as 'News', 'Entertainment', 'Forum', and 'Non-profit organizations'. They compared the geographical information obtained from the pages of categorized websites to the location estimated for the IP address of the server running a site. The geographical information of web pages was taken by a manual inspection of the page content when looked for text such as 'Contact' or 'Privacy Policy/Terms of Service'. The postal address of the content creator was used. If the postal information was not found on the pages, the external information source was used for the looked-up company, such as Wikipedia. The location of the web server (given by its IP address) was obtained from the location database (Maxmind, 2017; GeoIP2 Databases). The categorized websites were compared in order to see the difference between the postal address obtained for the content creator and the location obtained for the IP address of the server running the site. The threshold for similar location identification was 50 miles as a range of a city. The most geographically accurate (smaller difference) were websites in the categories 'Educational', 'Social Media', and 'Governmental'. The least accurate websites were in the categories 'Blog', 'Special Interest Group', and 'Non-profit organizations'.

The primary data about web hosting in the Czech Republic are provided by CZ.NIC (2017, CZ Domain Hosting Statistics). There are 72 web hosting providers listed in total, see details in the Annex. Selected data are shown in Table 1. The table shows the first ten Czech web hosting providers. The row 'Unknown' shows that about 40% of websites could not be determined (self-hosted or hosted). The last row shows the number of self-hosted websites.

Table 1 Share of Czech web hosting providers

Rank	Czech hosting provider	Share [%]
1	WEDOS Internet, a.s.	12.27
2	FORPSI	7.71
3	ACTIVE 24	6.14
4	ZONER software, a.s.	2.80
5	Cesky hosting*	2.57
6	Web4U, s.r.o	2.24
7	Gransy, s.r.o	2.02
8	Ignum s.r.o	1.57
9	ONEsolution, s.r.o.	1.50
10	Stable.cz	1.49
–	Unknown	40.85
–	Without hosting	4.97

Note: * – (THINline interactive, s.r.o.).
Source: CZ.NIC, November 2017

Other data about web hosting are available from BuiltWith (2017). The global (world) numbers are given along with the numbers for particular countries. The Czech Republic is also listed, and the results are shown in Table 2. The table again shows the top ten hosting providers. Some hosting providers listed were

not the same as with CZ.NIC. It is probably due to the used BuiltWith methodology based on comparing data from the IP address allocation database of RIPE NCC. Most Czech web hosting providers have not assigned their own IP address block and use the block of other ICT companies, such as Casablanca INT, which is included in the BuiltWith statistics.

Table 2 Share of Czech web hosting providers

Rank	Czech hosting provider	Share [%]
1	Casablanca INT	20.6
2	WEDOS Internet	17.7
3	SuperNetwork	16.8
4	VSHosting	15.4
5	Internet Cz	8.0
6	Zoner	6.3
7	Active 24	5.2
8	Ignum	4.0
9	Dial Telecom 2	2.0
10	CESNET	2.0

Source: BuiltWith, November 2017

Table 3 compares the values for the specific hosting providers listed in both sources included – WEDOS, Active 24, Zoner, and Ignum. The column 'Difference–Share' shows values ranging from 1 to 5%. The percentages are of small values and therefore, only a small difference changes the rank. The last column 'Difference-Rank' shows the change in the relative order.

Table 3 Data about Czech web hosting providers

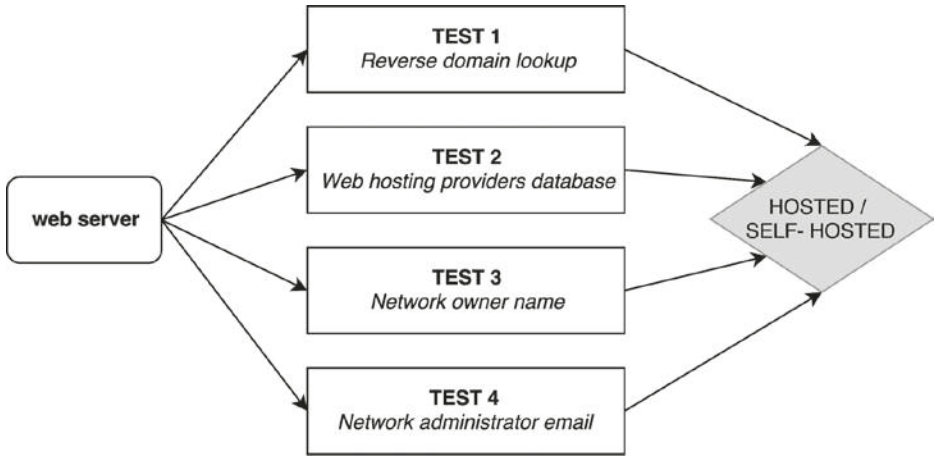
Web hoster	CZ.NIC	BuiltWith	Difference-Share	Difference-Rank
WEDOS	12.27	17.70	5.43	1 (1–2)
Active 24	6.14	5.20	0.94	4 (3–7)
Zoner	2.80	6.30	3.50	2 (4–6)
Ignum	1.57	4.00	2.43	0 (8–8)

Source: Own construction

2 METHOD FOR DETECTION OF WEBSITE HOSTING

Currently, there is no simple (straightforward) method known to detect whether a website is present at the owner's premises (self-hosted) or remotely hosted at a hosting provider (hosted). Related work refers to hosting identification by a single source of information. In our method, we use a set of different information that we process in a form of tests. For each test, we empirically assign a weight. The particular tests (source of information) are 'reverse domain lookup', 'web hosting provider database', 'network owner name', and 'network administrator email', as shown in Figure 1. We use the hosting status results for further data processing described in Section 4.

Figure 1 Web hosting status identification



Source: Own construction

In the following subsections we describe each test in detail, list input data (when used), and present a use-case example.

Test 1 – ***Reverse domain lookup***

For self-hosted sites owned and managed by an organization, it is expected that an administrator sets up the PTR records in the DNS system for each server owned. The PTR (pointer to a canonical name) records associate IP addresses with domain names and this information is used for reverse lookups (IP-to-domain). It shows if an IP address is used in a specific domain.

The test 'reverse domain lookup' compares a website domain name with the domain name of the IP address obtained from a DNS reverse query. Firstly, the IP address of the server is found by a DNS A query. Using the following DNS reverse query, the domain name for that IP is obtained. If the domain name for the tested web server matches the found PTR record, the website is self-hosted as shown in Listing 1.

This test is skipped if the PTR record is not set for an IP address. It may also falsely indicate hosting if the domain name of ISP is used instead of the organization's name.

Note: Our method considers Virtual Private Servers (VPS) as 'self-hosted'. They use not-shared IP addresses and a virtual server is maintained by the website owner.

Listing 1 Example of reverse domain lookup comparison test result executed in dig

```
$ dig www.mendelu.cz AAA +short
valar.mendelu.cz
195.178.72.2
$ dig +short -x 195.178.72.2
valar.mendelu.cz
```

Source: Own construction

Test 2 – ***Web hosting provider database***

Web hosting providers typically set up a DNS PTR record for the hosted websites to point to their domain name. For example, a Czech hosting company WEDOS assigns the domain names for their sites in the following pattern `xxx.wedos.net`. If the second-level domain name `wedos.net` is included in a list of domains of known web hosting providers, the website is very likely to be hosted. For the purpose of this test, we have collected domain names of the Czech web hosting providers from DNS reverse lookups and manual verification at the provider web page. The created list is shown in Table 4.

Table 4 List of known web hosting providers as input data for test 2. Some providers use multiple domain names, such as 'Cesky hosting' and FORPSI

Trademark	Domain name	Trademark	Domain name
ACTIVE24	active24.cz	Neomezeny hosting	neomezeny-hosting.cz
AeroHosting	aerohosting.cz	Neomezeny webhosting	neomezeny-webhosting.cz
Ahosting	ahosting.cz	ONEbit.cz	onebit.cz
Angel hosting	angel-hosting.cz	Otoman	otoman.cz
aspone.cz	aspone.cz	oXyShop	oxyonline.cz
ATTIVO	attivo.eu	Pipni.cz	pipni.cz
Banan.cz	banan.cz	Profitux	profitux.cz
Bezobav.cz	bezobav.cz	Quantasoft Hosting	qhs.eu
BlueBoard	BlueBoard	Rosti.cz	rosti.cz
Cesky hosting	ceskyhosting.cz, thinline.cz	Savana	savana.cz
Datahousing	datahousing.cz	Stable.cz	stable.cz
domeny.as	domeny.as	Station webhosting	station.cz
Ebola	ebola.cz	SvetHostingu.cz	svethostingu.cz
eBRANA	ebrana.cz	Sweb	sweb.cz
Endora	endora.cz	Thosting	thosting.cz
Eshop-rychle	eshop-rychle.cz	Tojeono.cz	tojeono.cz
Exo hosting	exohosting.cz	Web areal	webareal.cz
FORPSI	forpsi(.com, .net)	Web zdarma	webzdarma.cz
Gigaserver	gigaserver.cz	Web4ce	web4ce.cz
Gigaweb	gigaweb.cz	Web4U	web4u.cz
HexaGeek	hexageek(.com, .cz)	WebDum.com	webdum.com
Hosting 90	hosting90.cz	Webhosting C4	webhosting-c4.cz, skok.cz
Hosting Blueboard.cz	blueboard.cz	WebHosting.FM	webhosting.fm
HostingSolutions.cz	hostingsolutions.cz	Webnode	webnode.com, rubicus.com
HostingZdarma.cz	hosting(-)zdarma.cz	Webprostor.eu	webprostor.eu
Hukot.cz	hukot(.cz, .net)	Websupport	websupport(.cz, .sk)
IGNUM	ignum.cz	WEDOS	wedos(.cz, .net)
iSOL.cz	isol.cz	ZONER	zarea.net

Source: Own construction

The test may falsely indicate self-hosting if the web hosting organization is not included in our list. It may also falsely indicate hosting if the website of the hosting organization itself is tested. Table 5 shows an example of a positive evaluation of this test.

Test 3 – ***Network owner name***

For both the domain name and the IP address of a web server, it is possible to get the holder name from the relevant registers. National domain names can be looked-up in the WHOIS database managed by a national registrar. Regarding the IP address, this information can be acquired from he international WHOIS database managed by RIPE NCC. If the names from both sources are the same, the website is likely self-hosted.

Table 5 Example of 'web hosting providers database' test

Web address	www.uzis.cz
Server IP address	178.238.37.157
Domain name for IP address	yivo.onebit.cz
Found hosting provider	onebit.cz

Source: Own construction

This test may return a false hosting result if the holder of IP address is an ISP and not an end-organization. Also, the found names may not be exactly the same. For example, the company names stored in the CZ.NIC registrar are typically listed as the name plus some suffix according to the legal form of the institution, such as 's.r.o.' In such cases, it may also indicate false hosting result. In our implementation we calculate the similarity factor of organization names obtained from the registrar and RIPE NNC to eliminate the impact of same name variants.

An example of a positive test result is shown in Listing 2.

Test 4 – ***Network administrator email***

Large organizations typically have their own IP address space. Therefore, the relevant WHOIS registry should also contain an email to contact the holder of that IP space, in case of abuse etc. If this email address is identical with domain name of the tested website, the site is evaluated as self-hosted as being run in the organization address space. Table 6 shows an example of a positive evaluation of this test.

This test may falsely indicate hosting if the email has a different domain name from the web server.

Table 6 Example of 'network administrator email' test

Web address	czso.cz
Server IP address	194.48.241.132
Assigned IP addresses	194.48.241.0–194.48.241.255
Administrator email	jiri.lejnar@czso.cz

Source: Own construction

Listing 2 Example of testing conformity of domain and network holder name (listing is shorted)

```
$ whois www.cvut.cz
contact: SB:R1S-CES-8079-FA
org: Ceske vysoke uceni technicke v Praze
name: Ceske vysoke uceni technicke v Praze
address: Zikova 4
address: Praha 6
address: 16636
address: CZ
e-mail: neuman@vc.cvut.cz

$ dig www.cvut.cz AAA +short
cvut.cz.
147.32.3.202

$ whois 147.32.3.202
organisation : ORG -CVUT1 - RIPE
org - name : Ceske vysoke uceni technicke v Praze
address : Ceske vysoke uceni technicke v Praze
address : Zikova 1903/4
address : Praha 6
address : 166 36
address : The Czech Republic
abuse - mailbox : abuse@cvut.cz
(listings shortened)
```

Source: Own construction

2.1 Final hosting result

As described above, the particular tests could indicate the hosting status, but they may also fail in some cases. Therefore, we empirically assign each test result a weight as shown in Table 7.

We assign a value of –0.5 when the test 'reverse domain lookup' shows that the website is self-hosted. We assign this value as we believe its accuracy is in-between the accuracies of the last two tests. We assign the last two tests values of –0.4 (lowest) and –0.6 (highest) respectively towards self-hosting. The second test 'hosting provider database' is very firm and therefore we assign it a value of +1.

The weight of every test is included in the final score. If any test fails due to technical reasons (e.g. DNS query fails), we exclude its weight. The final hosting result is given by a sum of the weights. If the final score for a website is zero or negative value, we evaluate it as self-hosted (otherwise hosted).

Table 7 Test results and assigned weights – negative value indicates self-hosting

Test	Result weight	
	Hosted	Self-hosted
Reverse domain lookup	0	–0.5
Hosting providers database	1	0
Network owner name	0	–0.4
Network administrator email	0	–0.6

Source: Own construction

3 IMPLEMENTATION, DATA SOURCES, AND COLLECTED DATA

For the purpose of this work, we developed an application consisting of several modules that we categorize as source of data, data processing, and data storage. The module for source data covers crawling the Web, getting data from Internet registries, getting geographical data, and getting other related data (such as security implementations). The module for data processing is used for parsing input data, hosting status identification, and data correlation (such as geographical location and latency). The module for data storage is used for accessing SQLite database, data exporting, and data plotting on a map. For our application we used the Python 3 programming language with these main packages: `folium`, `dnspython`, `pyquery`, and `requests`.

The data used in this paper were collected through web crawling the public resources, including the lists of web addresses in public organization directories, available at `www.firmy.cz`, `www.toplist.cz`, and `www.sreality.cz`. With the first two, we crawled the public lists of organizations listed under specific categories. The data from the latter one were collected from a list of real estate agencies available at `www.sreality.cz/adresar` by Bulín (2017) and they form an additional category. The data related to IP address space primary come from regional Internet registry RIPE NCC accessed via the WHOIS database using the public server, available at `whois.ripe.net`. The data related to domain names come from the Czech domain name registry CZ.NIC accessed via `http://www.nic.cz/whois/`. For geographical data we used the free MaxMind (2017) database 'GeoLite2 City' with a local access.

In total, we have collected and processed data from more than 21 000 websites divided into nine categories. The numbers of crawled websites for each category are shown in Table 8. We selected the categories to cover both public and private sectors: i) private large sector – big companies (banks, insurance), ii) private small business sector – small companies (e-shops, real estate agencies, craftsmen), and iii) public sector (hospitals, government, schools, universities). This division is only indicative as e-shops and real estate agencies may fall into both big and small companies. Also hospitals, schools, and universities may fall in both public and private sector. We did not check the legal status and size of the entities. We rather evaluated each category independently and the numbers may further be combined based on specific needs.

Table 8 shows that most collected websites were from the private sector – small companies for these categories: e-shops, craftsmen, and real estate agencies respectively. Following was the public sector with categories of high schools and hospitals, respectively. These are the sectors for which the results

may be considered as 'strong' as a large number of websites was processed. The rest of the categories (banks, insurance companies, government institutions and universities) have smaller numbers given the size of the Czech Republic. The results are therefore only indicative and should be interpreted with the knowledge of the size of input data.

Table 8 Collected websites divided into categories

Sector*	Category	Websites
Private/large	Banks•	36
Private/small	E-shops	11 314
Public	Hospitals	469
Private/big	Insurance companies?	26
Private/small	Real estate agencies	1 660
Private/small	Craftsmen	6 563
Public	Government inst.?	109
Public	High schools	1 192
Public	Universities?	103
Total		**21 472**

Note: * – Indicative division; • – Indicative results.
Source: Own construction

4 RESULTS

4.1 Overall Data About Website Hosting

Using the tests described in Section 2 we have evaluated the hosting status of the collected websites. Table 9 shows the detail results of the test methods and the detected overall hosting status of the collected websites. The test results detect most of them as hosted. The highest numbers of websites evaluated as detected belong to e-shops and real estate agencies. The lowest number is for universities. The reason may be that larger organizations such as universities typically have their own IT departments, while those with the lowest number of detected hosted websites contain mainly smaller organizations.

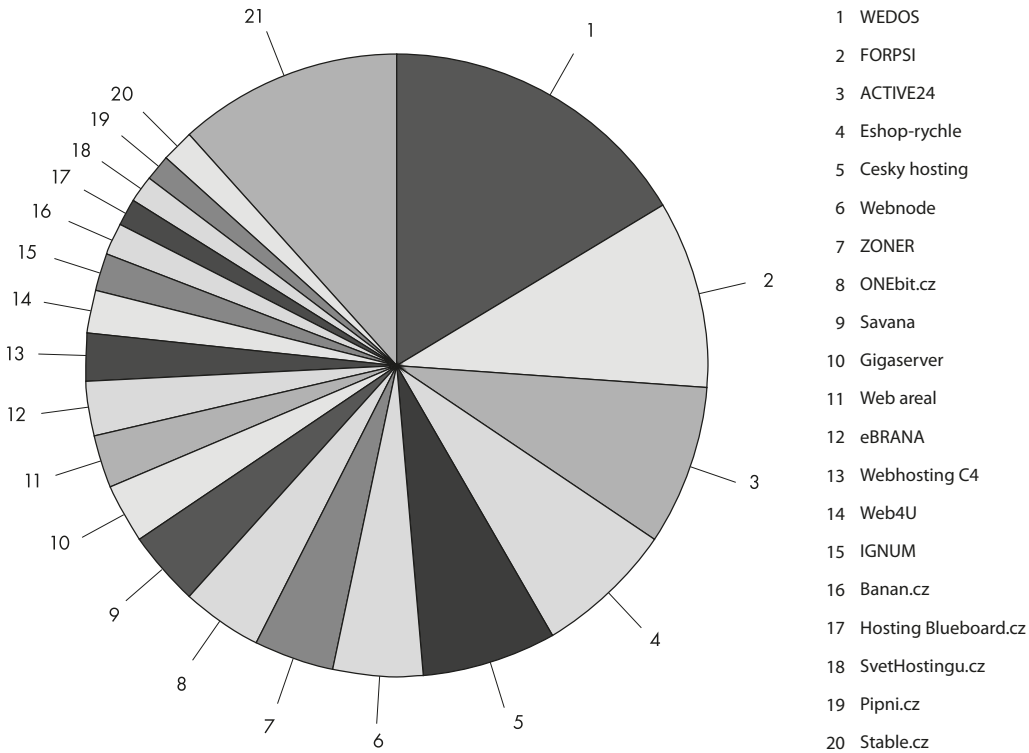
As for validation of the results, we randomly selected 400 websites from our dataset. There was the difference in the hosting status decision only in 14 of them. We may therefore state the classification accuracy of 95%.

Table 9 Percentage of websites detected as hosted in examined categories

Category	Evaluated	Percentage of detected hosted websites [%]				
		Test 1	Test 2	Test 3	Test 4	Overall result
Banks	36	52.78	5.56	86.11	77.78	75
E-shops	11 314	91.84	13.12	99.88	99.24	98.14
Hospitals	469	89.34	18.98	99.57	96.38	95.95
Insurance companies	26	76.92	3.85	96.15	84.62	84.62
Real estate agencies	1 660	91.27	16.27	99.64	98.07	98.13
Craftsmen	6 563	93.46	25.69	98.2	99.19	97.55
Government institutions	109	65.14	7.34	95.41	89.91	76.15
High schools	1 192	84.56	17.11	99.66	96.81	90.02
Universities	103	60.19	13.59	90.29	70.87	64.08
Total	**21 472**	**91.35**	**17.54**	**99.19**	**98.63**	**97.01**

Source: Own construction

Figure 2 Participation of webhosting providers



Source: Own construction

4.2 Web Hosting Providers

Based on the data from reverse DNS queries we have evaluated the numbers for web hosting providers, listed in Table 10. The numbers shown include the first 20 providers with the most detected hosted webs. The rest of providers (not shown) are summarized as 'Others'. To make it clear, the data with not-detected entries excluded are shown in Figure 3. The number of web servers where no hosting provider was detected is marked as 'Not detected'. This number also includes websites hosted by less known hosting providers that are not listed in Table 4. Our number of 'Not detected' websites is comparable with the 'Unknown' result provided by CZ.NIC in Table 1.

The websites counts for the biggest web hosting providers divided into the defined categories are shown in Table 11.

These data can be compared with the data by CZ.NIC, see Section 3. The contribution of the biggest providers is comparable. Some hosting providers are not mentioned by CZ.NIC since our list includes web hosting provider trademarks instead of the company full legal names. For example, CZ.NIC lists 'Gransy s.r.o.' but the trademark is 'Station webhosting'. The second example is that CZ.NIC uses 'THINline interactive, s.r.o.' and the trademark is 'Cesky hosting'.

Table 10 Detected websites counts for biggest hosting companies. Web hosting providers are labeled by their trademarks

Rank	Hosting provider	Webs	Share [%]
1	WEDOS	1 611	7.49
2	FORPSI	961	4.47
3	ACTIVE24	801	3.72
4	Eshop-rychle	726	3.38
5	Cesky hosting	667	3.1
6	Webnode	460	2.14
7	ZONER	424	1.97
8	ONEbit.cz	410	1.91
9	Savana	372	1.73
10	Gigaserver	301	1.4
11	Web areal	289	1.34
12	eBRANA	264	1.23
13	Webhosting C4	254	1.18
14	Web4U	200	0.93
15	IGNUM	192	0.89
16	Banan.cz	157	0.73
17	Hosting Blueboard.cz	147	0.68
18	SvetHostingu.cz	145	0.67
19	Pipni.cz	141	0.66
20	Stable.cz	137	0.64
–	Other providers	1 159	5.39
–	Not detected	11 693	54.36

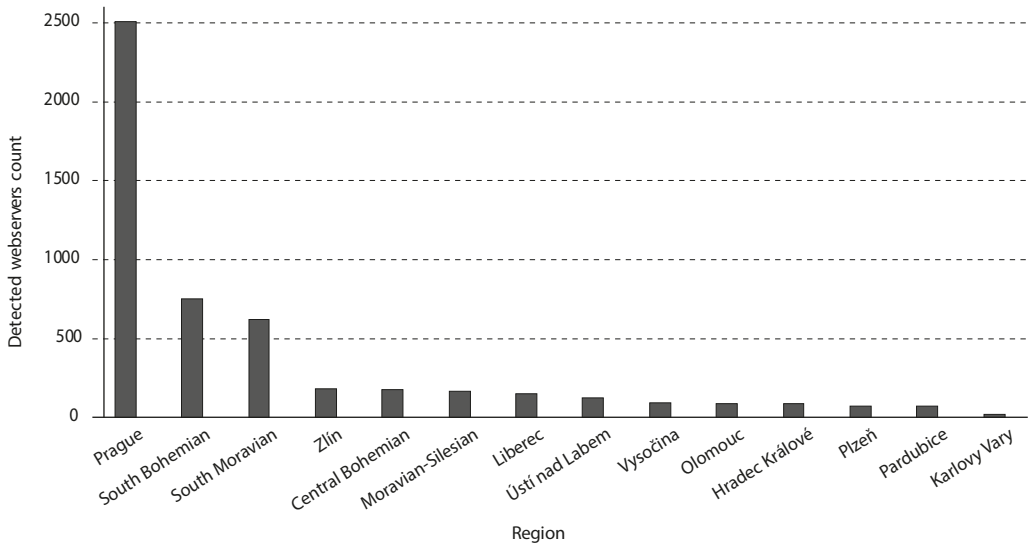
Source: Own construction

Table 11 Share of most common Czech web hosting providers in the examined categories

Category	Webhosting provider share [%]							
	WEDOS	FORPSI	ACTIVE24	Eshop rychle	Cesky hosting	Webnode	ZONER	ONEbit
Banks	0	0	2.78	0	0	0	0	2.78
E-shops	6.61	2.93	2.67	6.21	3.25	0.96	1.92	1.8
Hospitals	9.17	6.4	4.26	0	3.84	3.41	3.41	2.77
Insurance companies	0	0	3.85	0	3.85	0	0	0
Real estate agencies	6.99	4.94	3.86	0.12	2.59	1.33	1.45	2.11
Craftsmen	9.23	6.92	5.53	0.32	3.08	4.46	2.19	2.03
Government institutions	1.83	0.92	2.75	0	0.92	0.92	1.83	0.92
High schools	7.63	4.95	3.78	0	2.85	1.43	1.59	1.76
Universities	3.88	3.88	0.97	0	0	1.94	1.94	0.97

Source: Own construction

Figure 3 Web server location in the Czech regions

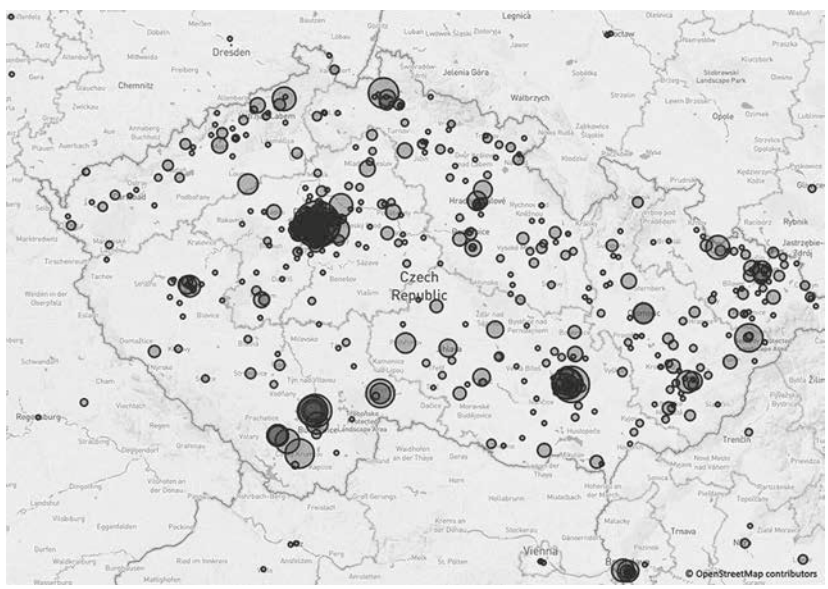


Source: Own construction

4.3 Geographical related data

The Czech Republic is a small country and the intra-country distance does not have any serious effect on web page loading delay. Almost half of the tested websites were hosted in the Prague region. The second one was the South Bohemian region, where servers of a large provider are situated. A map showing the location of servers in cities is shown in Figure 4.

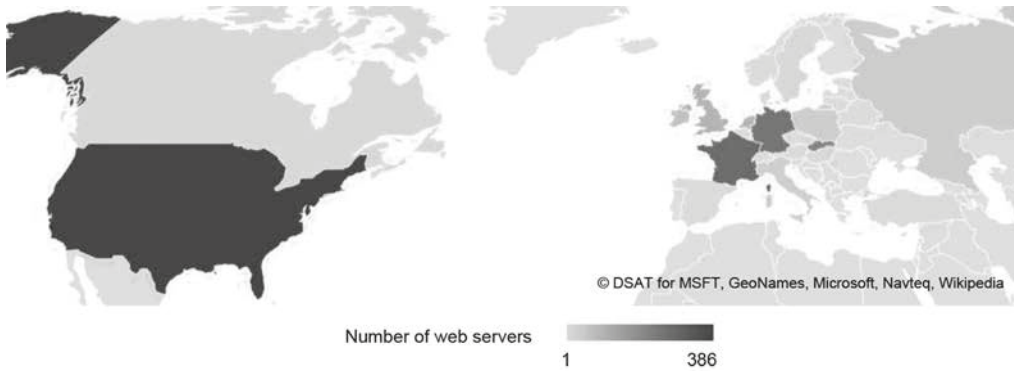
Figure 4 Web server location in Czech Republic



Source: Own construction

Although we focused on Czech domain zone, we detected 1 532 web serves hosted outside the Czech Republic. Most of these websites were hosted in the United States (386), France (274), Germany (247), and Slovakia (205). A map with major countries with the Czech websites hosted is shown in Figure 5.

Figure 5 World Czech website hosting server location



Source: Own construction

The median geographical distances from our university server in Brno to servers with a Czech domain in selected countries are listed in Table 12. For these distances, we estimated the minimal additional round-trip delay caused by data transmission in optical links over these distances. We used a simplified value of 5 us delay per 1 km (Coffey, 2017). As we calculate the minimum additional delay, we omitted the cable links inflation over distances and actual routing paths.

Table 12 Median distance and minimum additional RTT for servers running websites hosted outside the Czech Republic, rows are sorted by counts of hosted websites in each country

Country	Median distance		ExRTT*
	[km]	[miles]	[ms]
United States	7 489	4 653	75
France	1 039	645	10
Germany	579	360	6
Slovakia	122	76	1
United Kingdom	1 212	753	12
Netherlands	893	555	9
Ireland	1 648	1 024	16
Russia	1 748	1 086	17
Poland	361	224	4
Italy	689	428	7

Source: Own construction

4.4 Other Related Data

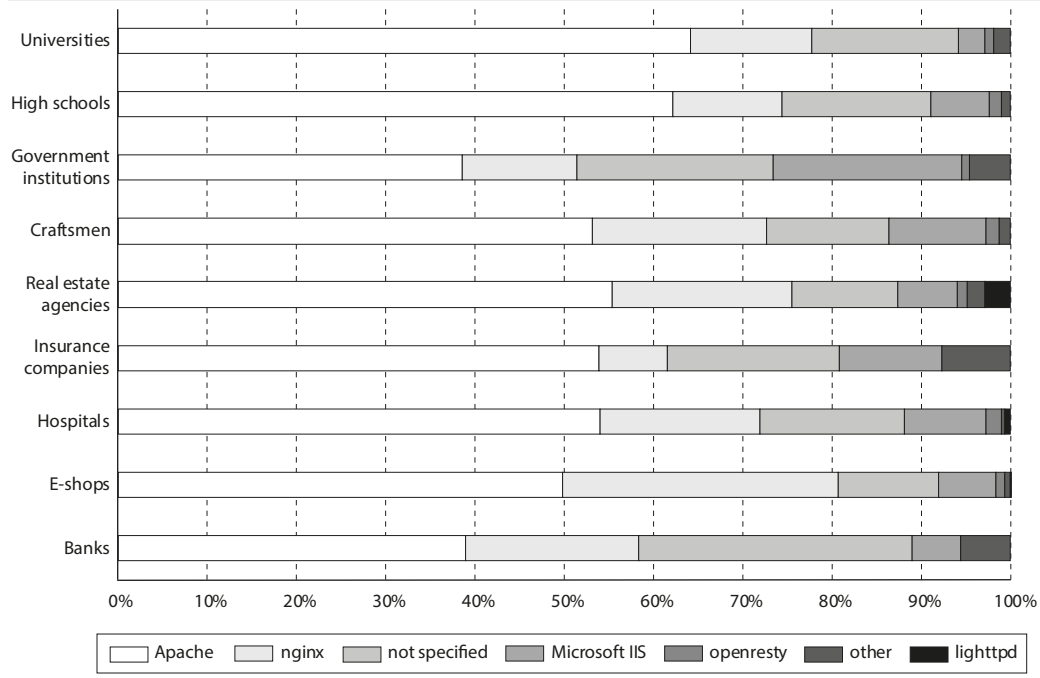
For other related data, we focused on the use of IPv6 and security implementations inthe Czech web servers. We detected the use of IPv6 by checking the existence of a DNS AAAA record and server HTTP-availability. For security implementations, namely HTTPS and DNSSEC, we considered a web server as HTTPS-compliant if there was a positive response for an HTTPS request and, also, a verified certificate was

present. For certificate verification, we used a list of certificates provided in the Debian `ca-certificates` package (2017). We obtained the DNSSEC data (DNSSEC keyset) from the WHOIS domain registry managed by CZ.NIC (2017, CZ domain registry). We considered a web server as DNSSEC-compliant if the keyset was present. We also found the web server software by inspecting the 'Server' field of the HTTP response. We again divided the results into nine organization categories.

The share of web server software is shown in Figure 6. We observed that the use of the free Apache is lower with government institutions and banks. Also, the data show that with the bank category, many web servers hide the information about used software for security reasons. We found that the use of the NGINX software is noticeably higher in the e-shop category.

We also detected version for some of the used software. The oldest detected version was for Apache '1.3.27', released in 2002. The use of such an old version can be dangerous due to known unfixed serious vulnerabilities (CVE Details, 2017).

Figure 6 Use of web server software



Source: Own construction

The use of IPv6 and security implementations is listed in Table 13. We included the numbers for IPv6 as its use is given by support of the hosting provider and, also, by DNS settings maintained by the administrator of the server domain. One could expect that universities would be the leading entity for the use of IPv6. However, the table shows that government institutions have the biggest number – 40%. This is probably given by implementation of the Czech government resolution (2009) about the use of IPv6. Banks and insurance companies have the highest percentage for the use of HTTPS as the entities offer secure services. The use of DNSSEC is the most significant with government institutions (72%, second is 50%). This is again probably given by implementation of the Czech resolution about the use of DNSSEC (2013).

Table 13 Use of IPv6 and secure implementations by organization categories

Category	Technology support [%]		
	IPv6	HTTPS	DNSSEC
Banks	5.56	66.67	30.56
E-shops	29.17	32.64	39.89
Hospitals	33.26	12.15	44.14
Insurance companies	3.85	69.23	42.31
Real estate agencies	24.88	14.82	49.34
Craftsmen	33.17	12.48	42.3
Government institutions	40.37	28.44	72.48
High schools	26.68	15.52	44.38
Universities	20.39	33.01	37.86

Source: Own construction

CONCLUSION

The paper presented a method for website hosting detection. The method consists of four partial tests with assigned weights. Large data were collected from the Czech Internet and processed for website statistical analysis. The data came from more than 21 000 websites that we have collected by web-crawling the public resources. We analyzed the websites in nine defined categories according to organizations they represent in their content. We also processed the data geographically to analyze the locations of the web servers. In addition, we focused on HTTPS, DNSSEC, and IPv6 protocols support. The used procedures applied in a country may be of use for marketing purposes, verification of fulfillment of legal directives, and for assessment of claimed web hosting plans.

The particular results detect 97% of the websites as hosted by another organization. The most used software was Apache followed by NGINX. Furthermore, 30 % of the crawled websites were available via IPv6, most of them in the category of government institutions. 24% of websites were available via HTTPS protocol, most of them in the categories of insurance companies and banks. The DNSSEC protocol was supported by 41.8% of the tested domains.

ACKNOWLEDGMENT

This work was supported by grant NSP LO1401 and the SIX Research Center.

References

BUILTWITH. *Web Hosting Usage Statistics* [online]. Sydney: BuiltWith, 2017. [cit. 5.1.2018]. <https://trends.builtwith.com/hosting>.

BULIN, M. *Analysis of Real Estate Market Using Information on Internet* [online]. Master thesis, Brno: Brno University of Technology, Faculty of Electrical Engineering and Communication, 2017. <http://hdl.handle.net/11012/65811>.

COFFEY, J. *Latency in optical fiber systems* [online]. White Paper, Hickory: Comm-Scope, 2017. [cit. 12.1.2018]. <https://www.commscope.com/Docs/Latency_in_optical_fiber_systems_WP-111432-EN.pdf>.

CVE DETAILS. *Apache 1.3.27 Security Vulnerabilities* [online]. CVE Details, 2017. [cit. 14.1.2018]. <https://www.cvedetails.com/vulnerability-list/vendor_id-45/product_id-66/version_id-8205/Apache-Http-Server-1.3.27.html>.

CZ.NIC. *CZ Domain Hosting Statistics* [online]. Prague: CZ.NIC, 2017. [cit. 2.1.2018]. <https://stats.nic.cz/stats/hosting>.

CZ.NIC. *CZ domain Registry (WHOIS)* [online]. Prague: CZ.NIC, 2017. [cit. 2.1.2018]. <https://www.nic.cz/whois/>.
DEBIAN. *Package: ca-certificates* [online]. Debian, 2017. [cit. 2.1.2018]. <https://packages.debian.org/en/stretch/ca-certificates>.
GOVERNMENT OF THE CZECH REPUBLIC. *Usnesení vlády České Republiky ze dne 8. června 2009 č. 729 ke Zprávě o přechodu na internetový protokol verze 6 (IPv6)* [online]. Prague: Government of the Czech Republic, 2009. [cit. 2.2.2018]. <https://kormoran.vlada.cz/usneseni/usneseni_webtest.nsf/0/6BFDE5B071A154C5C12575E5004024F1/$FILE/727%20uv090608.0727.pdf>.
GOVERNMENT OF THE CZECH REPUBLIC. *Usnesení vlády České Republiky ze dne 18. prosince 2013 č. 982 ke Zprávě o zavádění technologie DNSSEC a o plnění usnesení vlády ze dne 8. června 2009 č. 727 ke Zprávě o přechodu na internetový protocol verze 6 (IPv6)* [online]. Prague: Government of the Czech Republic, 2013. [cit. 2.2.2018]. <https://apps.odok.cz/attachment/-/down/VPRA9EVEBJYC>.
MAXMIND. *GeoIP2 Databases* [online]. Waltham: MaxMind, 2017. [cit. 2.1.2018]. <https://www.maxmind.com/en/geoip2-databases>.
SINGLA, A., CHANDRASEKARAN, B., GODFREY, B., MAGGS, B. The Internet at the Speed of Light. In: *Proceedings of the 13th ACM Workshop on Hot Topics in Networks,* New York: ACM Press, 2014, pp. 1–7.
TSOU, M. AND LUSHER, D. Mapping Web Pages by Internet Protocol (IP) Addresses: Analyzing Spatial and Temporal Characteristics of Web Search Engine Results. In: *Proceedings of the 17th International Symposium on Cartography in Internet and Ubiquitous Environments,* Tokyo: International Cartographic Association, 2015, pp. 1–20.
WANG, Y., BURGENER, D., FLORES, M., KUZMANOVIC, A., HUANG, C. Towards street-level client-independent IP geolocation. In: *Proceedings of the 2011 USENIX Annual Technical Conference,* Portland: USENIX, 2011, pp. 365–379.

ANNEX – List of web hosting providers considered by CZ.NIC

ACTIVE 24; AERO Trip PRO, s.r.o.; AIVision, s.r.o.; Amazon.com, Inc; Angel hosting; Axfone, s.r.o.; Banan, s.r.o.; Basefarm, AS; BEST-NET, s.r.o.; Blueboard.cz, s.r.o.; Bodis, LLC; business communication, s.r.o.; Bydzovsky, s.r.o.; Casablanca, Int.; Cesky hosting (THINline interactive, s.r.o.); Cesky server.cz, s.r.o.; CESKY WEBHOSTING, s.r.o.; CZOL media interactive, s.r.o.; Datahost, s.r.o.; DOMENY, s.r.o.; EBOLA Czech, s.r.o.; Explorer, a.s.; FlyNetwork, s.r.o.; FORPSI; Fortion Networks, s.r.o.; Gigaserver.cz; Google, Inc.; Gransy, s.r.o; Group NBT, plc; Happy Technik, s.r.o.; HEXAGEEK, s.r.o.; HOSTING90 systems, s.r.o.; HostingSolutions s.r.o; HUMLNET CREATIVE, s.r.o.; IglooNET, s.r.o.; Ignum s.r.o; ISOL Int., s.r.o.; IT Host.CZ, o.s.; KRAXNET, s.r.o.; LTweb s.r.o.; Luvenex plus, s.r.o.; Nethost, s.r.o.; NETIO Solutions, s.r.o.; Netlook, s.r.o.; Next Dimension, Inc.; Nodus Technologies, s.r.o.; OBSIDIAN, s.r.o.; ONEsolution, s.r.o.; OVH; PIPNI, s.r.o.; savana.cz s.r.o.; Savvy, s.r.o.; SecurityNet.cz, s.r.o.; Stable.cz; SuperNetwork, s.r.o.; SvetHostingu.cz; TELE3, s.r.o.; Telefonica O2 Czech Republic, a.s; Topweby; Trellian, Ltd.; UNIHOST, s.r.o.; united-domains, AG; Vas-hosting.cz; Web4ce, s.r.o.; Web4U, s.r.o; Websupport, s.r.o.; Web zdarma s.r.o.; WEDOS Internet, a.s.; W HOSTING, s.r.o.; WinSoft Company, s.r.o.; XHOSTING.CZ group, s.r.o.; ZONER software, a.s.

But Are those Numbers Correct? Some Suggestions for Appraising the Accuracy of Statistics

Simon Scott [1] | *Organisation for Economic Cooperation and Development, Paris, France*

Abstract

Knowing whether data are reliable is of fundamental importance to the establishment of knowledge, the formulation of explanatory hypotheses, and the development of effective policy. Yet there appear to be no standard, established tests to enable users to judge whether they should accept a given statistic as a fact. Numerous internationally agreed documents set out principles and practices to promote sound statistics, but they offer no direct guidance on whether to accept data as presented. Other documents discuss statistical quality, but focus largely on utilitarian considerations such as availability and timeliness; when they do discuss accuracy, they again consider processes (lists of good practices) rather than results (are the data correct?). This paper is a plea for, and a first attempt at, identifying some characteristics of data that may be accepted as true.[2]

Keywords

Statistical quality, accuracy, reliability, knowledge, truth

JEL code

C10

INTRODUCTION – THE EVOLUTION OF THINKING ON SOUND STATISTICS

The first recognisably modern steps towards ensuring statistical quality were taken in census legislation of the 18th and 19th centuries. These laws imposed obligations on citizens to answer questions truthfully, and also established some rights of privacy, e.g. to refuse to disclose religious belief. Census officials were required to maintain the secrecy of personal data and the confidentiality of commercially sensitive information. Such measures helped ensure an accurate count by compelling the respondent to provide correct information and assuring him that he would not suffer as a result.[3]

---

[1] Counsellor, Statistics and Data Directorate, Organisation for Economic Cooperation and Development (OECD), 2 rue André-Pascal, 75775 Paris Cedex 16, France. E-mail: simon.scott@oecd.org.

[2] This paper, presented in draft at the *2018 European Conference on Quality in Statistics* held at Krakow, Poland, during 27–29 June 2018, should not be reported as representing the official views of the OECD or of its member countries. The opinions expressed and arguments employed are those of the author only.

[3] Thus, for example, Presidential instructions for conducting the U.S. census of 1840 recognised in the following terms that accuracy required completeness, which in turn would be promoted by respect for privacy: "Objections, it has been

In the 19th and 20th centuries, statistical quality was further promoted through detailed instruction manuals regulating specialised statistical collections. Prominent examples include the standard guides on compiling the national accounts and the balance of payments, or government debt and deficit figures. To the extent that one could be sure that these instructions had been followed, and that no "tricks" had been played, one would be inclined to trust the resulting data – yet it would be difficult for any layman to make judgments on the fidelity with which the instructions had been followed, and in practice, separate judgments would be required on each country's implementation of each set of instructions.

It was in recognition of the generalised need for honesty and probity in statistical activities that, beginning in the 1990s, a new species of document emerged that abstracted from these laws and procedures principles thought to be of general application. The first such document was the *Fundamental Principles of Official Statistics in the Region of the Economic Commission for Europe* (ECE), agreed in 1992 (UNECE, 1992). This aimed to guide the newly free States of the former Soviet bloc to develop statistics that would command public trust because they respected "the fundamental values and principles which are the basis for any democratic society." The emphasis on producing information worthy of trust in a democratic society has persisted in this family of documents down to the present day. The UNECE principles were later adopted as a global standard in the *UN Fundamental Principles of Official Statistics* (UN 2014; original 1994) and the UN also developed *Principles governing international statistical activities* (UN, 2006). Key regional documents in the same group include the *European Statistics Code of Practice* (Eurostat, 2017), the *Code of Good Practice in Statistics for Latin America and the Caribbean* (ECLAC, 2011), and the especially comprehensive *Recommendation of the OECD Council on Good Statistical Practice* (OECD, 2015).

All these documents are "how to" guides designed to promote statistics suitable for a democratic society. In this respect they have many excellencies, but they do not attempt to establish the inherent characteristics of statistics that should be accepted as true.

For guidance on whether to accept data coming under our notice, we might turn in hope to another class of document, namely the presentations by various agencies of *dimensions of statistical quality*. Here again, however, we find no direct answer to the basic question of how to assess whether a number is true, or likely to be true. Indeed a UN study of international agencies' approaches to statistical quality (de Vries, 2002) even seems to imply that zeroing in on the question of accuracy is *dépassé*:

> *In statistics, quality used to be primarily associated with accuracy. It is now recognised that there are other important dimensions. Even if data are accurate, they do not have sufficient quality if they are produced too late to be useful, or cannot be easily accessed, or conflict with other credible data. Therefore, quality is increasingly approached as a multi-dimensional concept.*

As de Vries' analysis shows, such an insistence on multi-dimensionality has led to definitions of statistical quality that include timeliness, frequency, accessibility, relevance, coherence, interpretability, comprehensiveness, completeness, serviceability, integrity, credibility and clarity. Jostled by this throng of virtues, accuracy and reliability have retained only a minor place, and even so are defined in ways that somehow evade the question of whether the data are actually true. For example, de Vries reports

---

suggested, may possibly arise on the part of some persons to give the statistical information required by the act, upon the ground of disinclination to expose their private affairs. Such, however, is not the intent, nor can be the effect, of answering ingenuously the interrogatories. On the statistical tables no name is inserted – the figures stand opposite no man's name; and therefore the objection can not apply. It is, moreover, inculcated upon the assistant that he consider all communications made to him in the performance of this duty, relative to the business of the people, as strictly confidential" (Wright and Hunt, 1900, p. 145). It is especially easy to trace the evolution of census legislation in the United States of America, as the US Census Bureau has published on its website (US Census Bureau n.d.) detailed documentation, starting with the Constitution itself and the first census Act passed in 1790.

that the IMF defines *reliability* as merely "the closeness of the initial estimated value to the subsequent estimated value". The Fund's definition of *accuracy* – "the closeness between the estimated value and the true value that the statistics were designed to measure" – at first looks more promising, but it is then stated that "there is no single or overall measure of accuracy", and no attributes to gauge accuracy are offered. Moreover, insisting on the word "estimated" tends to suggest that absolute accuracy is impossible, no matter how simple the count might be.

Overall, it is clear that thinking about the soundness of statistics has focused on procedures by which official bodies can generate data useful to society.[4] This is, no doubt, a worthy objective, and governments now have a wealth of advice to follow about how to generate statistics that the public will see as possessing procedural integrity. Yet it is difficult to avoid the impression that something has been lost in the way the discussion has evolved. Nowhere do we find a checklist that citizens can use to judge whether any given statistic should be accepted. It is almost as if the need to secure public trust has discouraged the establishment of quality criteria which individual statistical series might fail. Paradoxically, however, the absence of such criteria may now be undermining that very trust – at least if we are to judge by how frequently we hear charges of *fake news, dodgy data*, *rubbery figures*, *alternative facts*, or *GIGO*.

This paper represents an initial attempt to identify some potential tests by which data users might judge whether to accept the numbers under their notice as knowledge. The discussion is divided into sections on *measurability* (how susceptible the target variable is to exact measurement), *measure* (the role of concepts and definitions in arriving at a correct representation of the target), and *measurement* (how the process of gathering and processing data may affect the accuracy of the resulting numbers). These are loose categories, and they overlap; they should not be seen as an attempt at typology but only as a means of giving structure to the argument.

1 MEASURABILITY

The simplest form of statistic is an *enumeration*. If I count the toes on my feet, or the apples in a barrel, then I shall arrive at an exact number, and if I do the job diligently, I may expect this number to be correct.

At the other end of the scale, some things cannot be enumerated, although attempts may be made to give them numerical expression. This especially applies to qualities rather than quantities. Business "confidence", employers' "willingness" to hire staff, the "liveability" of cities, as well as optimism, happiness, well-being, generosity, and other moods, intentions, or moral or ethical states, are not countable. Nevertheless, they are of interest, and must be expressed in numbers if they are to be compared over time and between parties. Hence they may be worked into figures by one technique or another, though the results must remain largely arbitrary.

The general rule is that *simplicity* and *tangibility* of objects improve their measurability. The most accurate statistics relate to countable objects. Objects in this sense may include animate objects, as long as their living nature does not impede their identification as objects of measurement. If one of my toes, or an apple in the barrel, is split or deformed, the question may arise whether it should be counted, or perhaps counted twice.

---

[4] For further references to relevant literature on statistical quality, focusing on international macroeconomic data, see the IMF's useful Data Quality Reference Site (IMF, n.d.). An anonymous reviewer has also rightly pointed to the existence of extensive broader literatures on data quality assessment (for a review, see Batini et al., 2009), as well as on metrology, on mathematical measurement theory, and on measure theory as a branch of mathematics. Each of these is a specialised and sophisticated discipline, geared towards the improvement of quantification through the application of professional skills and knowledge. While their discoveries and methods are, as a rule, beyond the grasp of the average citizen attempting to judge the reliability of a given statistic, they do repose on logical and empirical principles the study of which may suggest further practical tests to those sketched in this paper.

It is also important to appreciate the *temporality* of measurability. Measuring is an instantaneous act bringing together the measurer, the measure and the object of measurement. Only objects present at the moment of a measurement may be apprehended in that measurement. This means, first, that past states cannot be directly measured. Only the surviving evidence of that past state is available to be measured. Moreover, the quality of this evidence generally decreases with time, so that a count made from the present evidence of a past state becomes less reliable as that state recedes further into history. Nevertheless, the same target may eventually be estimated more accurately if new techniques improve the quality or measurability of evidence available about the past.

A further implication of the fact that measurability is a finite act occurring at a specific time is that, once they have been performed, measurements already relate to the past. By the same token, measurability does not exist now for future objects, since those objects do not yet exist. It follows that all *projections* should be treated from a scientific point of view as hypotheses rather than findings. Since direct measurement of the future is not possible, projections are often derived from models, which often include hypotheses about how variables relate to one another. Model outputs should be viewed as speculations, to be confirmed by actual measurements in future. In essence, they are not statistics at all.

In sum, from the point of view of the measurability of their targets, we may regard published figures as falling into one of two broad categories: knowledge, and hypotheses or speculation. Within these categories are certain gradations. Knowledge may be exact or vague, and hypotheses may be more or less grounded in existing observations.

The two categories are not quite watertight, and some forms of statistics straddle the divide. This particularly applies when the accessibility of information is taken into account. For many variables, the true figure is not known, and resort is made to *surveys*. Surveys are here taken to mean the collection of actual data, but from only a sample of the whole population concerned. The raw results of such surveys may be regarded as knowledge, but knowledge of a limited value since it does not relate to the totality of the category involved. Survey data are often presented as percentages, with the suggestion that the percentages can be taken to apply to the whole population, perhaps with an error margin based on the size of the sample and its share of the estimated total population. In fact, applying survey percentages to the whole population produces estimates, or speculations, the reliability of which depends on the size of the sample, the extent to which it represents the whole, the clarity of the measure and the diligence of the measurement.[5]

Beyond data generated by surveys conceived for statistical purposes, much use is now also made of data already available in existing records or through automatic logging of actions or transactions. Yet similar considerations of measurability apply to these sources, whether they be "administrative data", "big data", or data from scanners, webclicks, webscraping, or other sources. In all cases, the part of the variable accessible to measurement, and its relation to the whole, must be carefully assessed.

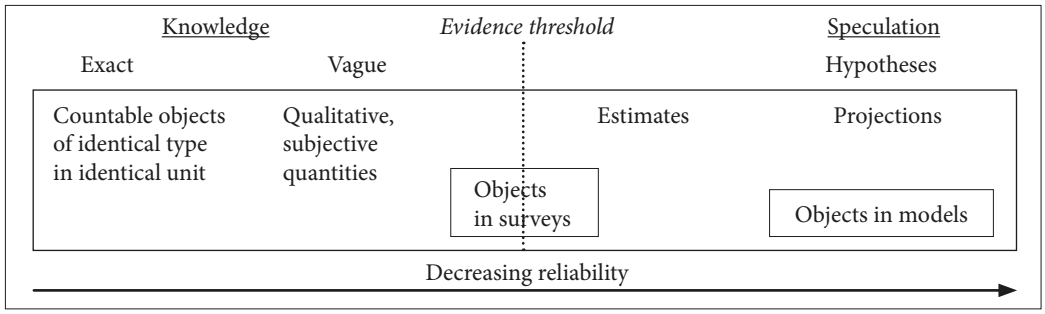
The essential features of the measurability scale implied by the above discussion are depicted in Figure 1. Items shown in the box are statistical quantities of various types. The line marked "evidence threshold" roughly marks the boundary between knowledge and speculation.

2 MEASURE

This section deals with how statistical concepts and definitions can promote or impede accurate measurement.

[5] US President Donald Trump evinced a realistic attitude to survey percentages in a speech at Wheeling, West Virginia on 29th September 2018. After citing numbers from an opinion poll showing strong support for a course of action he favoured, he observed "Hey, it's a poll. But we love those polls, don't we? I love polls. Only when they're good; when they're not good, I don't talk about them" (Trump, 2018).

Figure 1 Gradiations of the measurability of objects



Source: Own construction

Statistical measures are instruments which translate phenomena into numbers. So the first step in ensuring the reliability of a measure is to make its relation to the target phenomenon clear. The measure must define its object of measurement in a way that leaves no doubt what will be counted and what will not.

Sometimes a mere term will be sufficient. "Persons", "tonnes" or "dollars" are readily identifiable by all sane observers. Until recently the same might have applied to "men" and "women", despite some admitted marginal cases, but political discussions now cloud these categories. Wherever vagueness or ambiguity is present, mere terms will have to be supplemented by definitions that impose objective tests to consistently identify the objects of measurement.[6] Good and effective definitions possess both *exhaustiveness* and *exclusiveness*: they identify all and only those objects that are to be measured.

Tight definition is easier to achieve if the objects of measurement themselves form a logical and homogenous whole. Absolute homogeneity, or identity, is not required, but the objects counted under a measure must all possess some identifiable property which distinguishes them as a group and which separates them from other things which will not be counted. This identifiable quality must also be expressed in a single unit of measurement, such as tonnes, dollars, numbers of persons, etc. A single measure should never include within it different quantities, such as currency units of different countries, or real and nominal monetary units. A measure must always have one and only one *unit of measurement*; otherwise, the resulting number is meaningless, as it relates to no identifiable quantity.

Many different problems may arise in relation to units of measurement. A common error with money measures is to express them in "real" terms – i.e. at constant prices – without specifying the base year.[7] And data on technological subjects may be clouded by a misplaced urge to simplify units of measurement. Thus one sees the output of power plants expressed in terms of the number of "homes" they could serve, ignoring the fact that households' use of electricity varies by season and time of day and in any case accounts

[6] An example well-known to the author is that of "official development assistance" (ODA) – which has become the standard measure of government foreign aid. At first this was merely a descriptive term, qualified only by the observation that ODA was "intended to be concessional in character". Initial attempts to sharpen the definition focused on the source of the funds, but attention then shifted to the need to define "concessional". At one point, a qualitative definition requiring that the terms of ODA transactions be "significantly softer than the terms normally available for commercial transactions" was almost agreed. But in the end it was found necessary to introduce a strict mathematical test – loans would be reportable as ODA only if they embodied a grant element of at least 25 per cent, using a 10 per cent discount rate. The whole process took nearly four years, from early 1969 to late 1972 (see Scott, 2015).

[7] A Google search for "images" containing "constant prices" will disclose dozens of examples. See, for example, the Trading Economics page on Czech GDP (Trading Economics n.d.) which fails to supply the base year (2010) in the introductory statement, the chart, or the table – though it is finally mentioned in a box.

for only part of total demand. Press articles also often confuse megawatts, which measure instantaneous *power*, with megawatt-hours of *energy*, or temperature (the intensity of heat at a point) with enthalpy (the heat content of a system). The use of inappropriate units vitiates measurement.

Measures will typically also require specifications of time. *Stock* measures relate to a moment in time; *flow* measures to a period of time. Locations or *points of measurement* must also be defined so as to avoid multiple counting of the same item. For "stock" objects such as persons or commodities, this requires their unique localisation at the instant of measurement. For "flow" objects – and especially for money, which can pass through many hands before and after being exchanged for goods or services – careful thinking may be required to fix the point of measurement in a way that avoids unwarranted multiple recording.

The following may be considered as potential tests of the soundness of a statistical measure and hence of the reliability of associated data:

a. A good statistical measure starts with a sound and well-understood concept expressed in a definition which precisely identifies the target of measurement.
b. In general, the definition needs to be clear, unambiguous, exclusive and exhaustive. This may require sub-definitions of terms used, and explicit instructions about special cases.
c. If a definition requires multiple dimensions, then it must deal with all possible combinations of these dimensions in a way that clearly includes or excludes all potentially concerned phenomena.
d. A measure must never mix quantities: it should always possess a single, clear unit of measurement.
e. Units of measurement, points of measurement, the moment of a stock measurement, and the period of a flow measurement must all be specified.

3 MEASUREMENT

Certainty of identification remains an issue at the measurement stage. If identification is done by the enumerator, then some level of consistency may be expected, though the number and competence of the enumerators will also play a role. But if the targets of the enumeration identify themselves, then the prospects of a strictly accurate count are compromised. The degree of inaccuracy introduced may vary with the parameter involved. Statistics by age or sex may only be affected to the extent that respondents lie, are incapable of correctly identifying themselves, refuse to answer, or are of ambiguous sex. Statistics on religious faith or other beliefs will generally be more inaccurate, as the categories are more open to interpretation, and the self-image of respondents may diverge from the assessment of an enumerator. Even more inaccuracy is to be expected in responses on matters which may be the subject of pride, shame, reward or penalty.[8]

The method of measurement also has important implications for accuracy. As already mentioned, statistical measurements have traditionally been of two essential types: *censuses*, where the whole population is recorded, and *sample surveys*. In principle, censuses produce more reliable data, since estimation is limited to filling gaps created by non-responses. However, census data could only be perfectly accurate if all the target population were reached.

Traditional censuses have been the mainstay of official statistics throughout modern times but may now be dying out, as technology provides governments with all they need to know. Denmark has been a pioneer in this regard. Every individual, business and dwelling in the country is numbered, and data can be matched or extrapolated across the governmental system to produce information on population, employment, use of transport and government services etc. Other countries are heading in the same

[8] Huff (1954, pp. 132–133) gives the example of a survey of "8 000 representative British homes" which asked British men and women to say how often they took a bath. He rightly points out that "saying and doing may not be the same thing at all".

direction, but the digital transition is proving problematic. Australia still ran a census in 2016, but encouraged respondents to complete the forms online on the evening of 9th August. However, the system crashed at the vital time, leaving millions unable to file their returns. Access was not restored until nearly two days later, and the Prime Minister ordered an enquiry to determine "which heads roll, where and when"[9] (ABC News, 2016).

The contrasting experiences of these two countries show the advantages for statistical reliability of adopting a *single consistent approach to data collection*. This also applies in censuses of businesses, industries, or agricultural activities.

Sample surveys introduce issues of *representativity*: as already mentioned, any figures presented for the whole population from which the sample is drawn are merely estimates that depend for their accuracy not just on the extensiveness of the survey, but on the degree of conformity of the sample to the whole. Attaining representativity of a sample in all relevant dimensions is thus vital to ensure the reliability of a survey-based estimate.

Especially in surveys, it is important for accuracy that those collecting data do not have *personal or institutional incentives* to either exaggerate or minimise the phenomenon they are counting. In particular, data which violate the provision of the *OECD Recommendation on Good Statistical Practice* (OECD, 2015) that statisticians need to be "professionally independent from other policy, regulatory or administrative departments and bodies" should be treated with caution, especially if the measure in question has been made the subject of a *target*. Raising or spending predicted volumes of money, reducing waiting times for government services, improving clean-up rates for reported crime, or making the trains run on time may all become matters of announced targets, and figures showing whether the targets have been achieved will be more reliable to the extent that they are collected by officials with no incentive to "cook the books".

Sometimes no incentives are required for bias to be present. It is sufficient for enumerators to have a *firm opinion about the subject* of their count. If this is the case, one will almost always find that the figures published support the enumerators' prior opinion. This is the opposite of the "scientific principles and professional ethics" mentioned in the *UN Fundamental Principles* (UN, 2014), but it is common in academic debates and the bespoke data collections of think-tanks and lobby groups.

To sum up, accuracy of measurement can be assessed by examining:

a. The comprehensiveness of the count.
b. The number and competence of the enumerators.
c. The ease or difficulty in practice of making an unmistakeable identification.
d. Whether the identification is performed by the enumerator or the enumerated.
e. The presence or absence of institutional incentives or biases.
f. Personal biases towards obtaining one result or another.
g. The extent to which results are corroborated by other reliable measurements.

CONCLUSION

Current lists of statistical principles and good practices, instructions on how to collect specific statistics, and statements of the dimensions of statistical quality, do not provide – and do not attempt to provide – comprehensive guidance as to which statistics should be accepted as knowledge.

Yet in an era of "fake news", the public has never been in greater need of a set of objective criteria by which to judge the reliability of data as presented. This paper has therefore made a first attempt at suggesting potential aids to judgment. It has been organised according to three broad elements or

[9] The Prime Minister was speaking metaphorically as Australia had abolished the death penalty for federal offences in 1973.

stages of the statistical process, so as to offer guidance relating to the inherent measurability of the objects being quantified, the soundness of the statistical measure being applied, and the diligence and faithfulness of the act of measurement.

If further work is done in this area – whether by officials, academics, or civil society groups – then it may be possible over time to arrive at widely accepted checklists of statistical reliability, perhaps differentiated according to broad types of data or fields of enquiry.

One might hope for at least two benefits from such checklists. First, they could contribute to improving knowledge, especially by removing from consideration statistics that failed the criteria. Second, they could foster the elaboration of new and better data, by incorporating the desiderata on the checklists into the design of statistical collections. Both of these benefits could help improve the basis on which new hypotheses, research strategies, and policies are constructed.

Beyond these simple benefits, any patterns that emerge from the work of determining which statistics pass reliability tests may also contribute to eventual revisions of the existing general principles and codes of good practice.

References

ABC NEWS. *Census: Survey back online as angry Malcolm Turnbull warns "heads will roll"* [online]. Australian Broadcasting Corporation, 11th August 2016. <http://www.abc.net.au/news/2016-08-11/census-malcolm-turnbull-slams-abs-over-failure/7718584>.

BATINI, C., FRANCALANCI, C., CAPPIELLO, C., MAURINO, A. Methodologies for Data Quality Assessment and Improvement [online]. *ACM Computing Surveys*, 41(3), July 2009. <https://www.researchgate.net/publication/220565749_Methodologies_for_Data_Quality_Assessment_and_Improvement/download>.

DE VRIES, W. *Dimensions of Statistical Quality: A discussion note about the quality initiatives of some international organisations* [online]. New York: United Nations, 22nd August 2002. <https://unstats.un.org/unsd/accsub/2002docs/sa-02-6add1.pdf>.

ECLAC. *Code of Good Practice in Statistics for Latin America and the Caribbean adopted by the Statistical Conference of the Americas of the ECLAC* [online]. 6th meeting, Bávaro, Dominican Republic, November 2011. <https://repositorio.cepal.org/bitstream/handle/11362/16423/1/FILE_148024_en.pdf>.

EUROSTAT. *European Statistics Code of Practice* [online]. Luxembourg: Eurostat, 2017 [orig. 2005]. <https://ec.europa.eu/eurostat/documents/64157/4392716/Revised_CoP_Nov_2017.pdf>.

HUFF, D. *How to Lie with Statistics*. New York and London: W. W. Norton, 1954.

IMF. *Data Quality Reference Site* [online]. Washington: International Monetary Fund. <https://dsbb.imf.org/dqrs/approaches-to-data-quality>.

OECD. *Recommendation of the OECD Council on Good Statistical Practice* [online]. Paris: OECD, 2015. <http://www.oecd.org/statistics/good-practice-toolkit/Brochure-Good-Stat-Practices.pdf>.

SCOTT, S. *The Accidental Birth of Official Development Assistance* [online]. Paris: OECD, 2015. <https://www.oecd-ilibrary.org/development/the-accidental-birth-of-official-development-assistance_5jrs552w8736-en>.

TRADING ECONOMICS. *Trading Economics Website* [online]. <https://tradingeconomics.com/czech-republic/gdp-constant-prices>.

TRUMP, D. J. *President Trump EXPLOSIVE Speech at MASSIVE Rally* [online]. Wheeling, West Virginia, 29th September 2019. <https://youtu.be/x5sLPW0yYkI?t=553>.

UN. *Principles governing international statistical activities* [online]. New York: United Nations, 2006. <https://unstats.un.org/unsd/statcom/37th-session/documents/2006-13-principles-E.pdf>.

UN. *Fundamental Principles of Official Statistics* [online]. New York: United Nations, 2014. [cit. 29.1.2014]. <https://unstats.un.org/unsd/dnss/gp/fundprinciples.aspx>.

UNECE. *The Fundamental Principles of Official Statistics in the Region of the Economic Commission for Europe* [online]. Geneva: United Nations, 1992. <https://www.unece.org/fileadmin/DAM/stats/documents/e/1992/32.e.pdf>.

US CENSUS BUREAU. *Website of US Census Bureau* [online]. Washington. <https://www.census.gov/history/www/through_the_decades/overview/>.

WRIGHT, C. D. AND HUNT, W. C. *The History and Growth of the United States Census* [online]. Washington, Government Printing Office, 1900. <https://www.census.gov/history/pdf/wright-hunt.pdf>.

A Review on Measuring Digital Trade & E-Commerce as New Economic Statistics Products

Saeed Fayyaz [1] | *Statistical Center of Iran, Tehran, Iran*

Abstract

The Internet and digitalization are fundamentally changing the way people, businesses and governments interact. This has led to a new phase of globalization underpinned by the movement of data across national borders, changing the nature, patterns and actors in international trade in goods and services.

However, despite the growing importance of what is commonly referred to as digital trade, little empirical and internationally comparable information currently exists, inhibiting a full understanding of the scale and policy challenges of digital Trade, which has in turn raised concerns about the capacity of current statistics to fully capture and separate identify this phenomenon. In the other words, in digital trade economy there is a remarkable gaps and many of its statistics are not enough to fulfill the demands. It has been growing in importance, and with it, demands for detailed statistics from a number of policy areas including market access, trade facilitation, opportunities for SMEs (small and medium-sized enterprises), regulation, competition, digital data flows and privacy has remained yet.

Providing an overall review and making practical examples in case of Iran, this study tries to open other insight toward digital trade statistics as a new product of economic statistics. Beside, in this study some of the challenges like informal or semi-informal sectors toward digital trade will be addressed and an analysis on Organization for Economic Co-operation and Development (OECD) conceptual framework and an inventory of current measurement practices on digital trade will be presented. Finally, a case study on economic account new digital trade activity and startup based on the internet platform in Iran will be illustrated to describe the exponential growth rate of these kind of e-commerce in Iran. Producing reliable and accurate official statistics in this case is highly desirable. On the whole, obtained result show that e-commerce has been placed in top up-warding sector of economic growth in recent years in Iran.[2]

Keywords

Internet, digitalization, globalization, e-commerce, digital trade, informal or semi-informal economy

JEL code

L86, F60, L81, E20

[1] Statistical Center of Iran, F4, No. 1, Manufacturing, mining and infrastructural office, Dr. Fatemi, 1414663111, Tehran, Iran. Phone: (+98)2185102299, fax: (+98)2188970458, mobile: (+98)9126705085, e-mail: S_fayyaz@sci.org.ir.

[2] This article is based on contribution at the *IAOS (OECD) conference* in September 2018 in Paris, France.

INTRODUCTION

The Internet and digitalization are fundamentally changing the way in which people, businesses and Governments interact. This has led to a new phase of globalization underpinned by the movement of data across national borders, changing the nature and patterns of and the actors in the international trade in goods and services. While digitally related transactions in either goods or services have existed for many years, the current scale of transactions and the emergence of new and disruptive players (online platforms) are transforming production processes and industries, including many that were previously little affected by globalization (United Nations, 2018).

However, despite the growing importance of what is commonly referred to as digital trade, little empirical and internationally comparable information currently exists. This has inhibited a full understanding of the scale and policy challenges of digital trade, which in turn has raised concerns about the capacity of current statistics to fully capture and identify this phenomenon (United Nations, 2018). Moreover, the growing importance of enterprises with new business models as online platforms in Iran such as Digikala, Snapp, Tab30 etc. and other social networks like Telegram, Instagram, Facebook etc. gives rise to a number of additional complications, including in relation to the nature of their activities, for services trade and e-commerce policy. Besides, a number of efforts have been made by both governments and other stakeholders to develop definitions and classifications covering ICT-enabled services (UNCTAD, 2017).

In the side of official Statistics, an important impediment to the availability of data on digital trade – and certainly of statistics that are coherent with the current accounting frameworks *the 2008 System of National Accounts (2008 SNA)* and the *Balance of Payments and International Investment Position Manual, sixth edition (BPM6)* and are comparable across countries, is the lack of a common understanding of digital trade and of a comprehensive conceptual measurement framework. Therefore, as part of the collective efforts to address the broader measurement challenges, OECD (Organization for Economic Cooperation and Development) has developed a draft conceptual and measurement framework for digital trade, which provides a proposed typology of all digital trade flows that are considered "digital" (United Nations, 2018).

Following to introduction, section ii presents an overview of the conceptual measurement framework for digital trade based on OECD model and then an overview of the Iran's digital economy and existing data is provided in section iii. With considering conceptual framework, an analysis will be provided in the case of Iran in section iv. Finally, results can be presented.

1 CONCEPTUAL FRAMEWORK FOR MEASURING DIGITAL TRADE

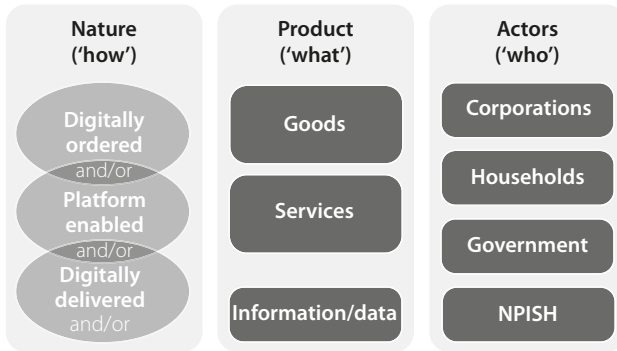
The focus on digital trade, brings to the fore new dimensions related to important characteristics of digitization, namely: the ordering and delivery processes (both of which can be digital), the nature of products (which products should be considered digital?) and the new actors involved, including not only digital intermediaries, but also households, given the increasing role played by consumers as unincorporated enterprises through the "sharing economy" (United Nations, 2018).

The conceptual framework (see the figure below) identifies those three key characteristics, or dimensions, as the nature of the transaction ("how"), the product ("what") and the partners involved ("who"). Central to the framework is the nature of the transaction, which builds on the common understanding that digital trade should encompass all digital trade transactions that are either digitally ordered, digitally facilitated (referred to as "platform-enabled") or digitally delivered (note that these are not necessarily mutually exclusive categories) (IMF, 2017).

As a magnificent challenge, there is no classifications of goods and services that aim to identify 'digital' products (e.g. ICT goods and services, ICT enabled services, trade in ideas etc.) that hopefully will be proposed in near future by OECD. It is widely acknowledged that measuring trade in services is more difficult than measuring trade in goods. Services are intangible, hard to define and unlike goods they

leave little or no administrative trail when crossing the border. As a result, data on trade in services lack the product and geographical detail available for trade in goods, a problem that creates significant and well-known knowledge gaps (UNCTAD, 2015).[3]

Figure 1 Dimensions of Digital Trade



Note: NPISH – Non-Profit Institutions Serving Households.
Source: OECD Handbook

Table 1 Different dimension of digital nature in conceptual model

	Digitally ordered transactions (E –commerce)	Digitally facilitated transactions (Digital trade flows)	Digitally delivered transactions
Definition	Sale or purchase of a good or service, conducted over computer networks by methods specifically designed for the purpose of receiving or placing orders	Cross-border trade flows facilitated by online platforms such as Facebook, Master Card, Visa etc.	Services and data flows that are delivered digitally as downloadable products.
Challenge in Iran	due to foreign currency exchange rate, online prices fluctuate and its getting hard to provide official statistics down by product and their average prices.	– Uncertainties about whether underlying transactions are recorded as cross-border trade or as income flows. – Cross-border transactions should be recorded as "gross" (including the value of underlying services provided between residents) or as "net" (i.e., including only the value of the intermediation fee as cross-border).	– Nonobserve economy is significantly bigger than observant economy. – Transactions' data do not have a monetary flow.

Source: Own construction

The second dimension identified in the framework ties into the first by identifying whether the products being traded relate to goods or to services, and also introduces a separate category referred to as information, or data (United Nations, 2018; UNCTAD, 2015). The last dimension concerns the actors involved. It's worth mentioning that the nature of transaction in digital process may be digital, the product either be digital (e.g. music files, article, etc.) or it may not be (e.g. clothes ordered online). A comprehensive overview on the conceptual framework in the case of Iran will be provided in next parts.

[3] See examples in UNCTAD (2017).

2 DESCRIPTION OF DIGITAL TRADE AND E-COMMERCE IN IRAN

ICT and digital trade provide potential opportunities for economic growth not only in developed counties as an origins of ICT developers but in developing countries as comparable as Iran by facilitating online purchasing/sale. Iran with approximately 80 million populations, is one of the biggest developing countries in the middle east which has been mounting both fix and mobile the broadband internet connection. Significant infrastructure investment has developed widespread nationwide coverage of mobile networks and a national fiber-optic backbone. There is a high level of basic mobile access and mobile broadband has been growing rapidly since its recent deployment. Due to the fast and rapid improvement in ICT technology, the fix and mobile broad band internet connections has changed the way of communication and business in Iran. The internet penetration rate has risen dramatically during recent years and reach from 22.73% on 2012 to 61.03 on 2017 and it is estimated that it drives to 69% at the end of 2018. Rapid increase in the internet penetration rate, which obtain from proportion of internet users on population with six and more, shows the significant growth in the number of internet users and as a result considerable growth in the number of monetary transactions. As a supplementary index, the ICT development index (IDI), also was calculated on three criteria include access, skill and use of the internet, similarly confirm the growth of internet users in Iran the IDI index went up from 4.97 on 2015 to 5.85 on 2017.

Regarding to the ICT users, there are different types of enterprises which indirectly affect the digital market and e-commerce activities. These enterprises produce and serve ICT goods and services and composite the ICT share in Iran's GDP and play enabler roles for the other sectors as well. Based on 2017 survey, the distribution of ICT enterprises in Iran is described in Table 3. In these classification, some sectors wholly are digital like Electronic disseminations and others have indirect correlation and provide infrastructures for digital trade. In Electronic disseminations part, both state own and private organizations were placed for example IRANDOC[4] is a governmental organization which sales thesis, papers, articles, magazines online. It registers all identifications data of their customers and use special gateway for online transitions. All goods and services are available just online.

Table 2 Different types of ICT enterprises in Iran based on 2017 surveys

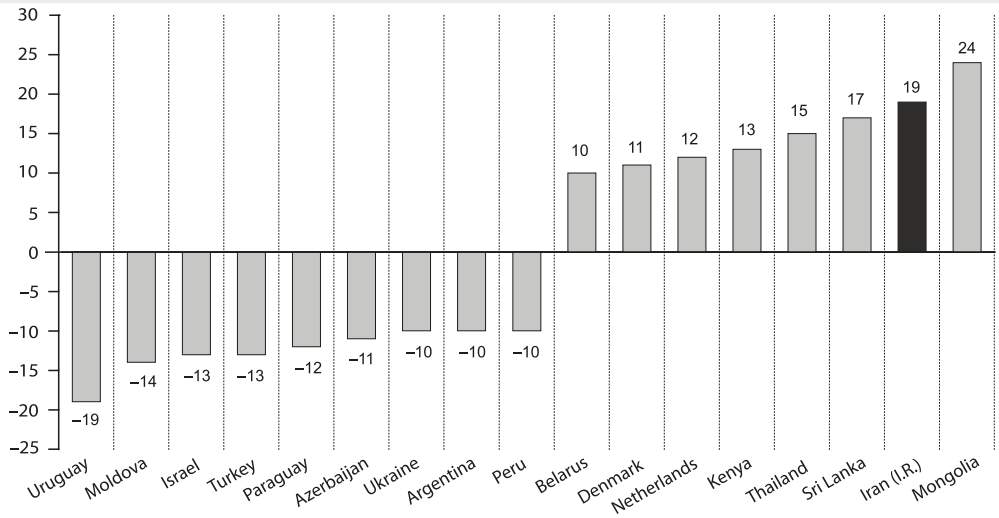
Type	Number	Value added (Rial)	Activity / service type
CT services providers	5 834	~34×10^{12}	Cafe net, ISP, ISDP, FCP, FWA, Serveco, SAP, MVNO, mobile rooming.
IT services providers	1 949	~7×10^{12}	Software programming, computer consultancy, data ware house, web hosting , computer services.
Post and communications	14 393	~8×10^{12}	Mobile, Fixed telephone, Post, Delivery.
Computer and telecommunications equipment installation and repair	4 561	~11×10^{12}	Mainframe computer, telecommunications equipment, personal computers.
produce ICT equipment	257	~16×10^{12}	Electronic equipment and devices, computers, telecommunication equipment, magnetic media, fiber optic cable, telecommunication cable, USP.
Wholesale ICT goods	1 145	~1×10^{12}	Computers, telecommunication equipment.
Electronic disseminations	410	~2×10^{12}	Online files, books, magazine etc.
Sum	28 549	~79×10^{12}	ICT's goods and services.

Note: ISP – Internet Service Provider, ISDP – Internet Service Distribution Provider, FCP – Fixed Communication Provider, FWA – Fixed Wireless Access, Serveco – Service communication, SAP – Satellite Access Provider, MVNO – Mobile Virtual Network Operator.
Source: SCI 2018 ICT report

[4] <www.irandoc.ac.ir>.

In recent years, the dramatic growth of internets users and as a results online businesses and digital trade has drawn the national attentions to this prominent sector in Iran as far as Iran placed in top 10 developing economies in UNCTAD (United Nation Conference on Trade and Development) E-commerce Index, 2017. Any indicator that is a proxy for online payment affects the index for economies where there is a high incidence of cash used to pay for e-commerce purchases (cash on delivery accounted for 7 per cent of global payments in 2015). In the Impact of new payment indicator, Iran with score +19 was placed in rank two after Mongolia with +24 endorsing the progress of new payment methods in Iran as shown below (UNCTAD, 2017) while the proportion of ICT sector from Iran's GDP (Gross Domestic Product) on 2016 was approximately close to 10% (there a gigantic non- observe economy in this part).

Figure 2 Impact of new payment indicator by UNCTAD



Source: UNCTAD B2C 2017 report

Table 3 UNCTAD B2C E-commerce Index for Iran's data, 2016

2016 Rank	Share of individuals using Internet (2014 or latest)	Share of individuals with credit card (15+, 2014 or latest)	Secure Internet servers per 1 million people (normalized, 2014)	UPU postal reliability score (2013–14)	UNCTAD B2C e-commerce Index value 2015	2014 Rank
77	39	11	38	82	42.6	69

Source: UNCTAD B2C 2017 report

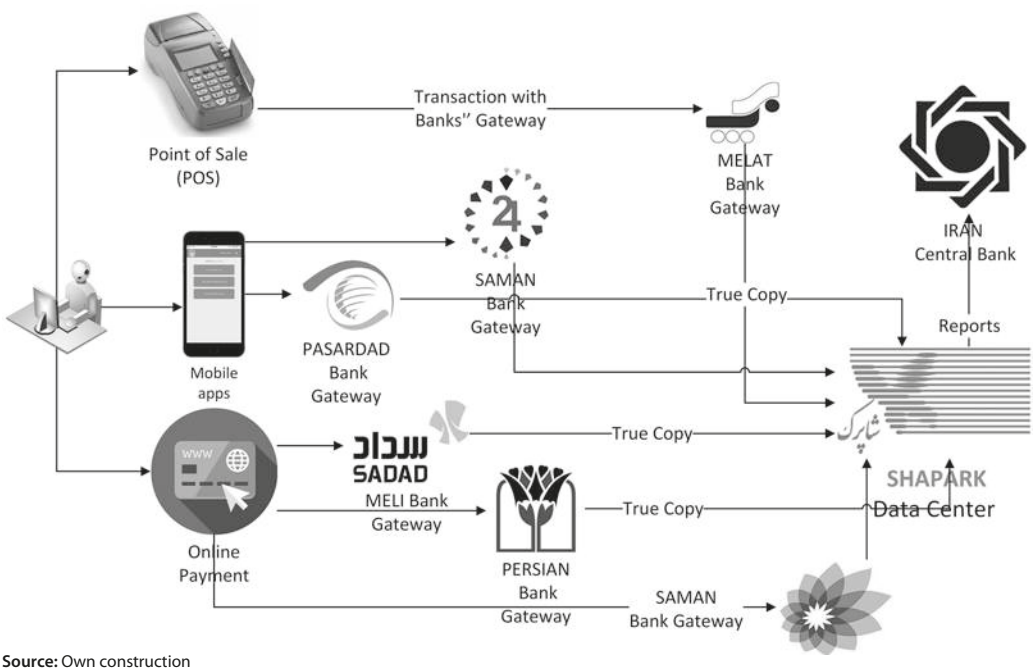
Table 4 Iran Placed in top 10 developing economies in the UNCTAD B2C E-commerce Index, 2017

2017 Rank	Share of individuals using Internet (2016)	Share of Individuals with an account (15+, 2014 or latest)	Secure Internet servers per 1 million people (normalized, 2016)	UPU postal reliability score (2016)	Index Value (2016 data)	Index Value (2015 data)	Index Rank (2015 data)
48	53	92	45	86	69	65	52

Source: UNCTAD B2C 2017 report

For more details, all legitimate websites whether are created or managed by individuals or enterprises have been tagged with E-NAMAD sign in their websites which shows its legal permission. Besides, all of the both governmental and private banks in Iran receive their activity permission and authorizations from Iran Central Bank (ICB). These Banks have their identifiable gateways for online payments which is recognized as unique getaway. All banking transactions' true copy will send to central database (SHAPARAK Co. on the behalf of ICB directly. Up to now, there are 12 legalized Banking payment service provider (PSP) for online payments under the support of different banking system. Clearly, SHAPRAK is responsible for saving, managing and maintaining a true copy of the all monetary transactions in the both banking system and all online transactions under the supervision of ICB whether they have done in website, mobile applications and points of sale (POS) devices.

Figure 3 Process description of different payment methods in Iran

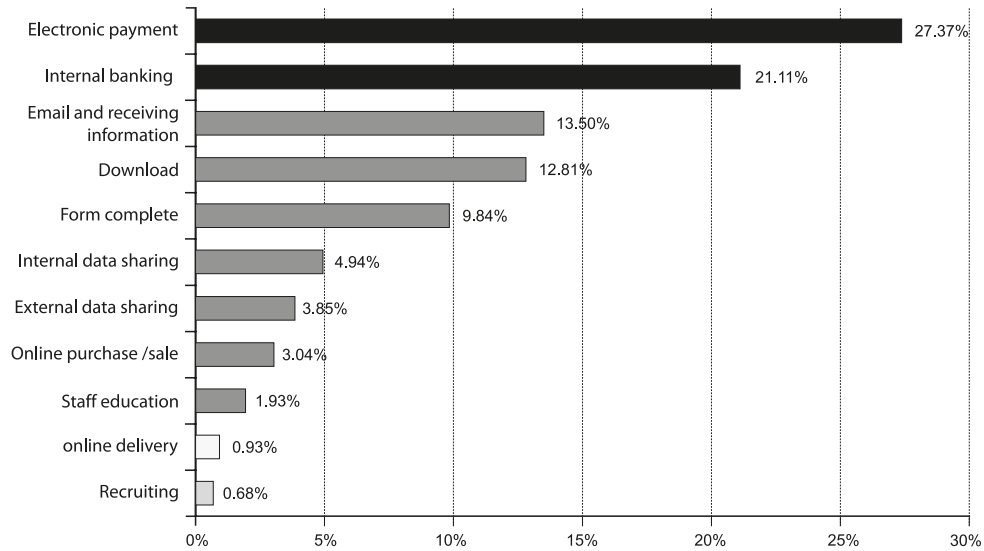


Source: Own construction

In 2017 B2C E-commerce target there are 4 indicators: Internet users (ITU – International Telecommunication Union), Secure servers (World Bank), Account penetration (World Bank Global FINDEX Database) and Postal reliability score (UPU – Universal Postal Union). In order to measure the digital trade and e-commerce in Iran, a survey for enterprises conducted on 2012 which include sample size 3834 from 5854 enterprises and it covered four ISIC code 7910, 5811, 6612 and 5510 (ISIC Rev.3).[5] This survey focused on measuring digital trade in mentioned codes though it was not fully covered the whole enterprises. Regarding to obtained results, 60% of enterprises had internet based purchase and sale (equal to 3.048) (SCI, 2012).

[5] International Standard for Industrial Classification (ISIC).

Figure 4 Activity type by using the internet in Iran's enterprises on 2017



Source: SCI 2018 ICT report

Besides, digital-trade can be gone through both formal and informal divisions in Tables 5 and 6.

Table 5 Different dimensions of digital trade and e-commerce in formal sector of Iran

	Method	Description	Data source
Formal sector	Business-to-Business (B to B)	In case of online ordering or purchasing of other businesses' good and services either raw material or final products, the transaction can be tracked in both enterprises' accounting systems or in the SHAPARK database if they always use identifiable unique gateways.	1. ICT and e-commerce enterprise surveys 2. Financial Statement (Registration data)
	Business-to-Consumer (B to C)	The data of those enterprises receive ordering or sale good or services online can be gathered from different surveys as well as their registered accounting systems.	1. ICT and e-commerce enterprise surveys 2. Household surveys on internet use 3. Financial Statement (Registration data)
	Government-to-Business (B to G) Government-to-Consumer (C to G)	E-Government: now many of government's services are accessible online. Businesses or individuals who are looking for online services, also can pay online. All the online payments have done in one of the banking gateways with a true copy sending to the Central database in SHAPARK Co. Thus, all these transactions can achievable by registered data.	Administrative registers: e.g. trade statistics in custom clearance register, financial statement by Supreme Audit Court

Source: Own construction

Contrary to digital trade in formal and legitimate situation, there are an unlimited number of digital trade and e-commerce activities in informal and casual conditions.

Table 6 Non-observe digital economy of informal sector of Iran

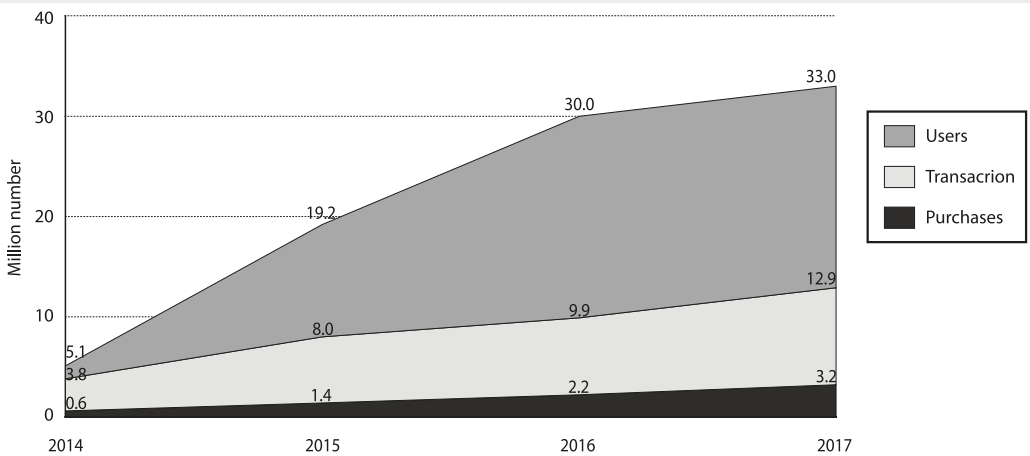
	Method	Description	Data source
Informal sector (non-observe digital economy)	Business-to-Consumer (B to C)	Many of businesses sale their product both good and services informally in the internet. They offer their products in illegal websites or social network like Instagram, telegram, Facebook etc. All above transactions have not been navigated and tracked by other administers like tax office, statistical agencies and more although all the monetary transaction happen in banking system. In the other words, it is implausible to distinct those banking transactions in center database (SHAPARAK Co.).	Not available
	Consumer-to-Consumer (C to C)[6]	According to the astonishing growth of the internet penetration rate in Iran, majority of people use their own privacy for doing business. They use weblogs, social network like telegram groups and channels, Instagram, illegal website etc. to sale their goods and services online. Informally in most of cases, people give their own bank card number in order to pay/receive the money. Although these kinds of monetary transactions register in banking system and main database, it would be impossible to recognized them.	Generally, NA but partly: 1. Household surveys on internet use (partly) 2. Enterprise and household expenditure surveys (partly)

Source: Own construction

3 CASE STUDY: BAZAR CAFÉ STARTUP AS ONLINE PLATFORM

Bazar Café Startup has nominated as one of the pioneer's mobile application which provide an online platform for internet user in Iran with focusing on android software. Bazar Café application offer a wide range of very exciting mobile applications comprise games, sport, entertainments, health, ethical, science, economic, social media etc. and it could have had 33 million users (approximately 45% of the total internet users and 1.4 person in each Iranian family) in Iran as shows in Figure 5. In 2017, Bazar Café published 21 500 developers and 111 600 applications which shows potential capacity to make substantial investment in this part. During 2017, Bazar Café app was used 4.8 billion times and 1.4 billion times installed or updated. Besides, by the end of the year, 1 000 billion Rial was the developers' income.

Figure 5 The number of Bazar Café Users and transactions between 2014 and 2017



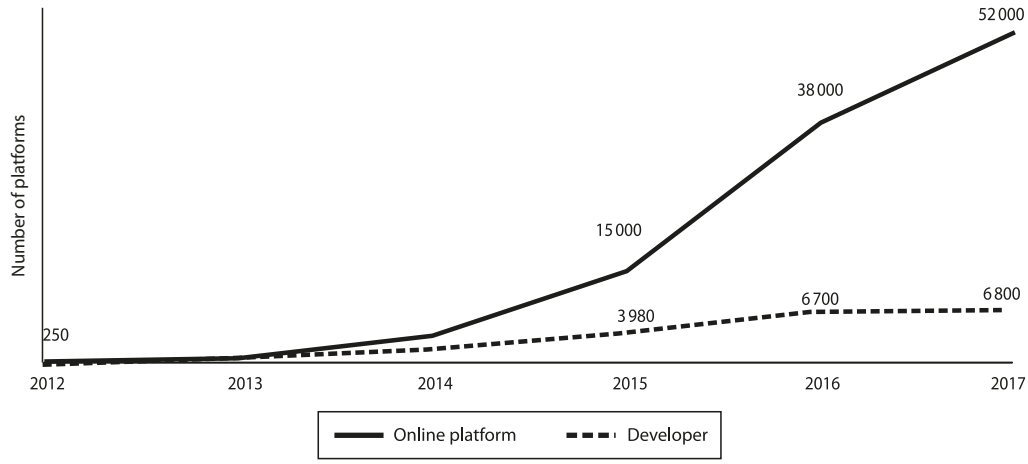
Source: Bazar Café 2017 report

[6] Due to the economic and political imposing sanctions, individuals haven't been permitted to have foreign exchange account or credits card in Iran.

As a biggest app stores for android in Iran, it offers both free and payable apps (Bazar Café Startup, 2017a). In order to measure the digital trade, it is not important haw many online platforms (Apps) had been produced but the processes that statistical system encounter with them, how they have been registered, tracked, updated and what type of data flow they have will be crucial.

Meanwhile, this startup could provide 20 700 and 10 350 direct and indirect employees (23% women and 77% are men) respectively by 1900 developer team and 1400 small-scale firms (Bazar Café Startup, 2017b).

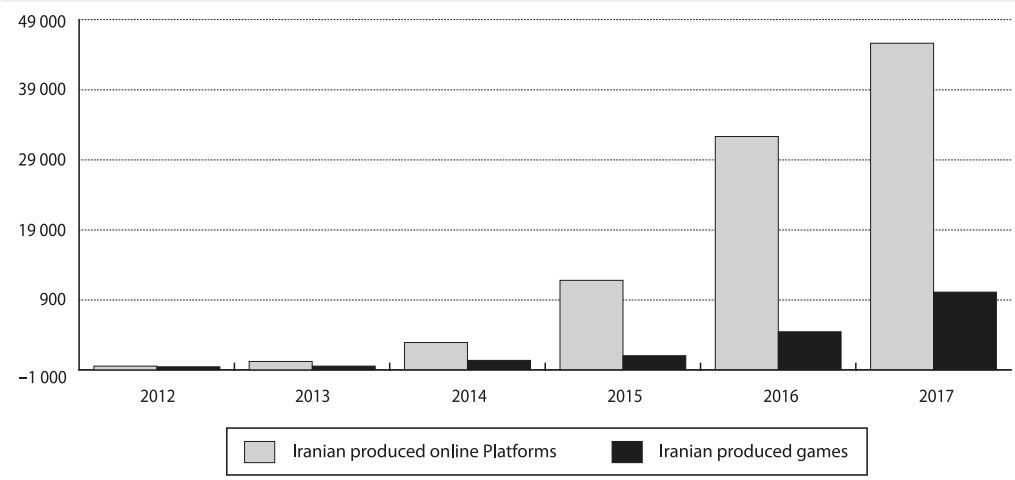
Figure 6 The number of Bazar Café online platforms versus Apps developer



Source: Bazar Café 2017 report

According the statistical report (Bazar Café Startup, 2017c), Bazar Café startup has been used by foreign users in other countries with 2% of all of the users. These foreign users comprise of 10% Turkey, 11% USA, 21% Afghanistan, 24% Iraq and 34% other countries.

Figure 7 The number of Iranian applications and games in Bazar Café app between 2012 and 2017



Source: Bazar Café 2017 report

4 CASE STUDY: DIGITAL TRANSACTIONS REPORT BY SHAPARAK CO.

SAPARAK Co. is the last resort in all banking transactions and play as a main datacenter which maintain all transactions records. Meanwhile, based on latest report on 2018 (SAPARAK Rerport, 2018), 93.61% of recorded transactions were purchasing and 6.39% related to bills and mobile charge. Totally, proportion of SHAPARAK's transactions on liquidity was equal to 10.81 while amazingly proportion of these recorded transactions on GDP was 119.02% (it is impossible in digital transaction to distinct the intermediate and final consumption). This astonishing rate was from 1 647 380 290 transactions with 1 760 068 425 million Rial.

Regarding to variety of digital gateways, all transactions were registered by Point of Sale (POS), Mobile Payment Gateway and the Internet Payment Gateway. The portion of each device from total transactions were 3.82%, 6.85% and 89.34% for the internet platform, Mobiles' applications and POSs respectively while the transactions' per capita population with are equal and greater than 18 were 27.62, 24.67, 1.89 and 1.05 for total devices, POSs (Points of Sale), Mobiles' applications and the internet platform respectively.

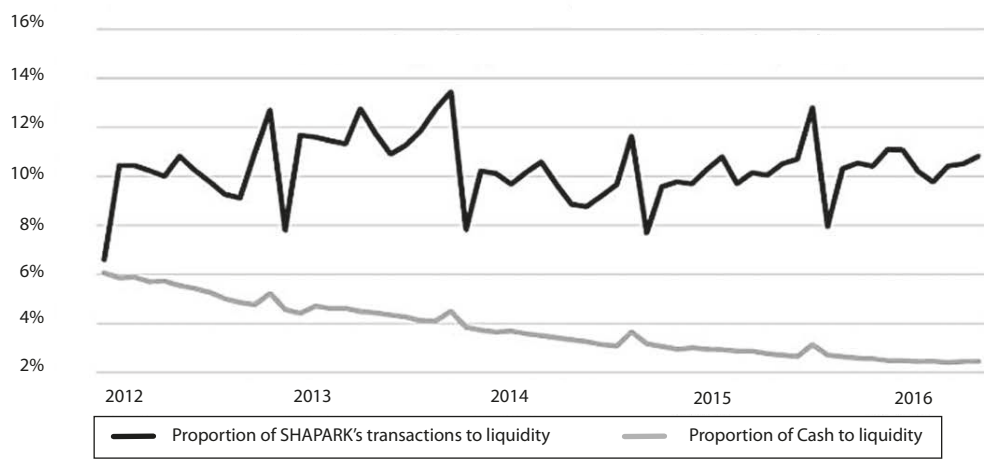
Table 7 Description of digital transaction gateways in May 2018 of Iran

Device type	Share %	Transactions	Transaction per device	Transaction value (Million Rial)	Transaction value Per device (MR)
Internet platforms	3.82%	62 868 993	85	185 336 910	250.60
Mobile Payment	6.84%	112 803 859	117	7 799 551	8.07
POSs	89.34%	1 471 707 438	212	1 592 931 964	229.42
Sum	100.00%	1 647 380 290	190	1 786 068 425	206.51

Source: Shaparak 2017 report

According to the transactions number and physical cash spending for purchasing good and services during the different years, buyer has been continuously replacing digital payment instead of cash which approved in Table 7. Interestingly the amount of transactions falls at the beginning of each year which might be due to the new year's vacation in Iran.

Figure 8 Comparison between digital transactions and cash



Source: Shaparak 2017 report

Digital payment and transactions has been snowballing trend as shows in Figure 8 and it has ripple effect on other economic sectors. In the other word ICT and digital trade as role in Iran economic. For more illustration, in May 2018, 1.647 million transactions with value of 1.768 billion Rial has done in comparison with previous month with 1.449 million transitions and value of 1.178 billion Rial has 13.67% and 51.59% growth in the number of transactions and its value correspondingly. Similarly, May 2018 has significant growth in comparison with similar month on 2017 by increasing 28.29% and 34.44% in transactions and its value.

Table 8 Problems toward digital trade in Iran

1	Surveys do not yet provide a detailed breakdown of the value of digital transactions.
2	In practice it may be difficult to separate the intermediation fees from the value of the services provided. In the other words, most users of digital services cannot separate the value of digital services from physical goods that they buy.
3	Measuring the non-observe digital economy (e.g. trade in social networks like telegram, Instagram etc.) has turned to the big challenge in Iran. ICT group has been looking for conducting a survey to build an estimation framework of this part.
4	How to separate digital and non-digital goods or services? What kinds of classifications or threshold should be defined?

Source: Own construction

Table 9 Suggestions toward digital trade in Iran

1	Using the potential of big data produced by the evolving digital economy, in particular by technologies such as the Internet of Things, cloud computing, and artificial intelligence. An intuitively straightforward option would be to add questions to surveys regarding the breakdown of online purchases and sales into domestic and international transactions
2	Online surveys with anonym netizens as the respondents should emphasized to better understand the dynamics of digital transactions as well as profiles of digital transaction actors.
3	Customs clearance or internationals post offices (DHL, TNT etc.) database can be beneficial to separate those goods and services which had been digitally ordered.
4	Iran's NSO conducted "Household access and the internet individual users" (reference year 2017) with supplementary one-page questionnaire to gather online purchasing buy households 'member so as to provide HH19, HH20 and HH21 ITU's Indexes. Next year, the extra questions will be added to separate directly the value of goods and services, intermediate consumption, related costs, payment details, origin, kind of platform, type of products etc. so as to better calculate the added value in this sector.
5	Conducting a survey for online platforms (those active in digitally ordered, enabled and delivered) and legalized and registered websites (those using legal banking gateways) with input-output approach

Source: Own construction

CONCLUSION

Digital trade and e-commerce has turned to an inseparable part of the economy and due to the developing trend, National statistical system and especially NOSs should try to provide official statistics for tracing and preparing adequate internationally comparable statistics. Regarding to the impediment current capacities to provide related statistics, this is study attempts to provide an outlook of digital trade capacities and infrastructure in Iran with focusing more on gaps and challenges. As a developing country, Iran has faced with remarkable difficulties like non observe digital economy to fully measure the digital trade and ecommerce. As far as recently published manuals and continuously under progressing issues, dealing with challenges and problems to some extent describing in this paper, can help statisticians remarkably. Hopefully, the results would be worthwhile for other developing countries.

References

UNITED NATIONS, ECONOMIC AND SOCIAL COUNCIL. *Report of the Inter-Agency Task Force on International Trade Statistics*. Statistical Commission, 49th session, 6–9 March 2018.

IMF. *13th Meeting of the IMF Committee on Balance of Payments Statistics.* Paris, France, October 24–26, 2017.

UNCTAD. *International trade in ICT services and ICT-enabled services: proposed indicators from the Partnership on Measuring ICT for Development.* 2015.

STATISTICAL CENTER OF IRAN. *Results of the Survey on "ICT Access and Use by Households and Individuals of Iran".* Statistical Center of Iran (SCI) report, 2017.

UNCTAD B2C E-COMMERCE INDEX. *UNCTAD Technical Notes on ICT for Development.* TN/UNCTAD/ICT4D/09, 2017.

STATISTICAL CENTER OF IRAN. *Results of the Survey on "Digital trade and e-commerce in Iran's Enterprise".* Statistical Center of Iran (SCI) report, 2012.

BAZAR CAFÉ STARTUP. *Bazar Co. report on "mobile apps in Iran on 2017 – applications' report".* Bazar Café Startup Statistical annual report, 2017a.

BAZAR CAFÉ STARTUP. *Bazar Co. report on "mobile apps in Iran on 2017 – employment report".* Bazar Café Startup Statistical annual report, 2017b.

BAZAR CAFÉ STARTUP. *Bazar Co. report on " mobile apps in Iran on 2017 – the internet and networks report".* Bazar Café Startup Statistical annual report, 2017c.

Report on "economics of SHAPARAK Co.". Vol. 35, May 2018.

UNCTAD, DIVISION ON TECHNOLOGY AND LOGISTICS SCIENCE, TECHNOLOGY AND ICT BRANCH, ICT ANALYSIS SECTION. *International Trade in ICT Services and ICT-enabled Services.* Proposed Indicators from the Partnership on Measuring ICT for Development, 2017.

UNCTAD. Technical Notes on ICT for Development. In: *Search of Cross-Border E-Commerce Trade Data,* 2017.

ICT SURVEY RESULTS. *ICT survey on enterprises; manufacturing, service providing, ICT indicators, online dissemination.* Conducted by Iran NSO with corporation of Ministry of ICT, 2017.

The EuroGroups Register

Agne Bikauskaite [1] | *Eurostat, Luxembourg, Luxembourg*
August Götzfried [2] | *Eurostat, Luxembourg, Luxembourg*
Zsolt Völfinger [3] | *Eurostat, Luxembourg, Luxembourg*

Abstract

Globalisation presents significant statistical challenges, particularly for small and open economies in terms of measuring statistical indicators and communicating the results to users. The European Statistical System allocated high priority to the better measuring of globalisation in the statistical processes and output, in business or macro-economic statistics.

Some concrete actions were already undertaken such as setting up of the EuroGroups Register of multinational enterprise groups and the putting in place of a so-called Early-warning System for monitoring restructurings of the groups.

This paper focuses on the EuroGroups Register (EGR), the central statistical business register of Eurostat and the EU and EFTA countries' statistical authorities. The EGR is part of the EU statistical infrastructure and has been built up to better capture globalisation effects as well as for improving the consistency of national data on enterprise groups.[4]

Keywords

Globalisation, EGR, statistical business register, multinational enterprise groups

JEL code

F6

INTRODUCTION

Globalisation presents significant statistical challenges, particularly for small and open economies in terms of measuring statistical indicators and communicating the results to users. The European Statistical System allocated high priority to the better measuring of globalisation in the statistical processes and output, in business or macro-economic statistics. Some concrete actions were already undertaken such as setting up of the EuroGroups Register of multinational enterprise groups and the putting in place of a so-called Early-warning System for monitoring restructurings of the groups.

This paper focuses on the EuroGroups Register (EGR), the central statistical business register of Eurostat and the EU and EFTA countries' statistical authorities. The EGR is part of the EU statistical infrastructure and has been built up to better capture globalisation effects as well as for improving the consistency of national data on enterprise groups.

The EGR covers multinational enterprise groups operating in Europe. It provides the statistical authorities of the EU and EFTA countries with yearly population frames of multinational groups. The Register's main function is to provide the statistical authorities with a harmonised picture of multinational groups for their national statistics. The EGR has been growing in terms of quality over

[1] Eurostat, 5 Rue Alphonse Weicker, 2721 Luxembourg, Luxembourg. E-mail: Agne.Bikauskaite@ext.ec.europa.eu.
[2] Eurostat, 5 Rue Alphonse Weicker, 2721 Luxembourg, Luxembourg. E-mail: August.Goetzfried@ec.europa.eu.
[3] Eurostat, 5 Rue Alphonse Weicker, 2721 Luxembourg, Luxembourg. E-mail: Zsolt.Volfinger@ec.europa.eu.
[4] This article is based on contribution at the *IAOS (OECD) conference 'Better statistics for better lives'* in September 2018 in Paris, France.

years and now covers more than 110 000 multinational enterprise groups in the EU. When at least one legal unit of a multination enterprise group is registered in the EU or EFTA country, the group is in the scope of the EGR.

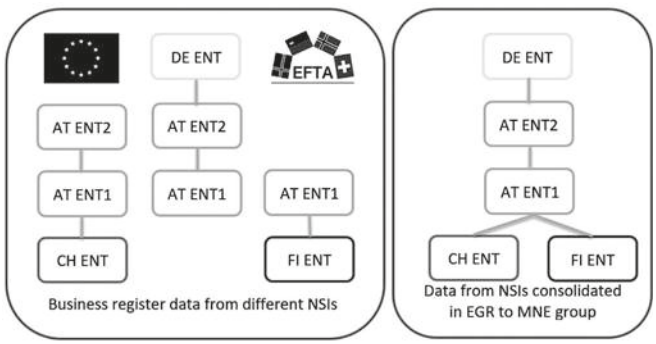
This paper will present the scope, the processes, the data sharing and use of the EGR, in particular how information is produced on the overall structure of multinational enterprise groups in the EU using the EGR.

1 SCOPE OF THE EGR

The fragmented picture that statistical authorities of EU and EFTA countries have of multinational enterprise groups operating on the EU market causes growing harmonisation problems for several types of statistics affected by globalisation.

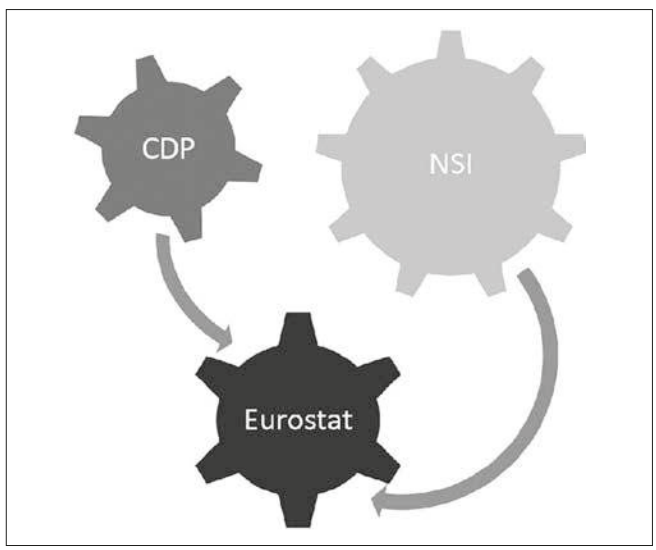
The EGR is designed to provide a unique survey frame for these and other statistics and therefore to serve as the basic tool for improving them. National statistical institutes (NSIs) and national central banks can use EGR as consolidated and harmonized frame when producing national statistics related to globalization.

Figure 1 Compilation of the groups in EGR



Source: Authors' own construction

Figure 2 Contributors to the EGR



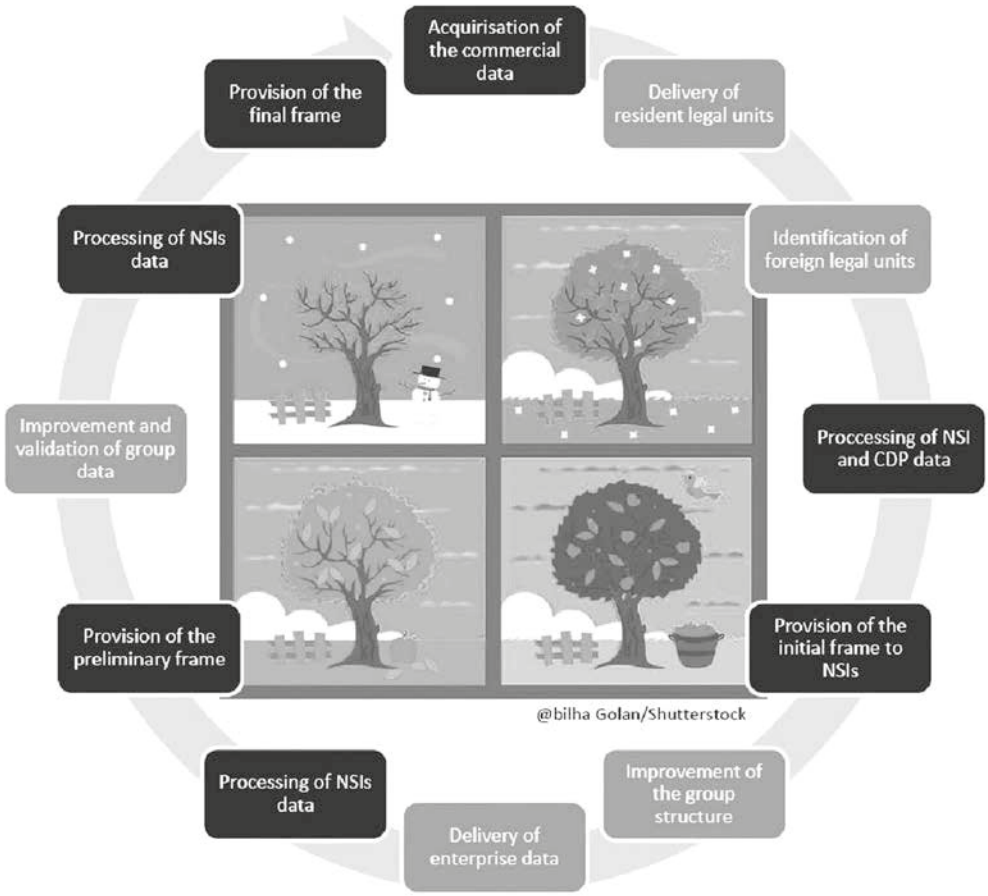
Source: Authors' own construction

2 PROCESSES OF THE EGR

The EGR produces annual data, reflecting the status of the registered gro ups as of 31st December of the reference year. The collection and validation of the EGR data are complex and long processes. The EGR final data for the reference years are currently available 15 months after the end of each reference year.

To create the EGR frames, Eurostat collects input information on enterprise groups from the national statistical business registers of EU and EFTA countries, and from one commercial data source. National statistical registers deliver micro data on the constituent units of the groups and on their relationships. The EGR also acquires data from one commercial data provider (CDP); this acquisition ensures coverage for units outside the EU and EFTA, and full consistency of the largest enterprise groups.

Figure 3 The annual data exchange cycle



Source: Authors' own construction

After validation of input data from the different sources Eurostat via the EGR calculates the consolidated output on the legal units (companies), on their relationships and on their enterprises (the statistical representation of the legal units). Complementary and conflicting information from the sources are treated in the EGR process with predefined priority settings and business rules. In the EGR production, the highest priority data are the data from the country of the subsidiaries (bottom-up view), followed by the data from the country of the parent companies (top-down view), followed and complemented by data from commercial data provider.

Having final information on legal units and relationships for a reference year the EGR compiles the enterprise group structures. The enterprise groups are built on those legal units, which are linked together by control relationships, where the voting rights are above 50 per cent.

The following enterprise group characteristics are calculated in the EGR process for each group: EGR identification number, group main activity code, group employment, group turnover and country of global decision centre (country of the group).

Before publishing the final data on the groups, the calculated group structures and group characteristics are revised and validated by the participating countries and Eurostat.

The EGR 1.0 system was in production since 2009, data of multinational groups were processed with this system for reference years 2008 to 2011. In EGR 1.0 data of large multinational enterprise groups were acquired from two commercial sources and these data were validated by NSIs. As the initial data had only partial coverage, the EGR 1.0 output also had limitations. EGR 1.0 produced 4 yearly frames for reference years 2008 to 2011; the process was improved year by year.

The EGR process was redesigned to the EGR 2.0 from 2012, the new process was fully implemented in 2015. The switch to the EGR 2.0 improved both the EGR process and the IT system. Data for reference years 2014, 2015 and 2016 were produced completely with the EGR 2.0 system. EGR 2.0 aims to cover all relevant multinational enterprise groups present in the EU.

The current EGR 2.0 system has four modules:

- the EGR IS (Identification Service) that supports users in identifying legal units,
- the EGR CORE (Core application) that consolidated input data and generates the EGR frames,
- the EGR IM (Interactive Module) that provides users with interactive web interface for data repair and validation,
- the EGR FATS (output interface) that provides users with web interface to consult the EGR output.

The EGR data enable information to be produced on the overall structure of multinational enterprise groups in the EU, while preserving the confidentiality of micro-data.

3 THE EGR COVERAGE

The EGR provides harmonised information on:

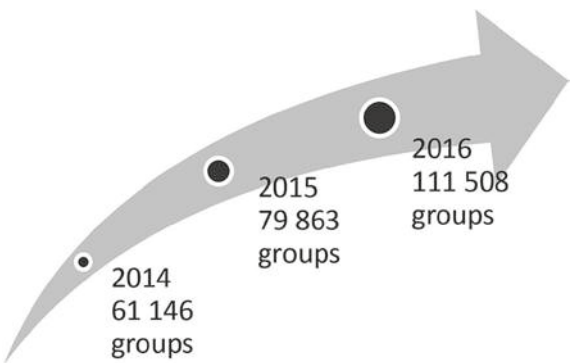
- multinational enterprise groups, enterprises and their legal units (identification, demographic and economic characteristics),
- relationships and control of enterprises within the groups,
- shareholdings of at least 10 per cent.

The EGR coverage has increased over the years, due to the launching of EGR 2.0 system in 2015 and the on-going quality improvements. The coverage of the last available EGR 2016 frame enlarged compared to the 2015 reference year, especially for Europe. The coverage of outside Europe entities slightly decreased, as less complete groups were acquired from commercial sources.

The data acquisition policy of EGR targets to collect year by year more data from NSIs and publicly available sources and less data from commercial sources.

For reference year 2014 EGR produced data of 61 000 multinational enterprise groups, for 2015 data of 80 000 groups. For the last available reference year, 2016 EGR produced data on 111 000 groups covering 944 000 legal units (companies) and 779 000 enterprises (statistical representation of the companies).

Figure 4 The EGR coverage over the reference years

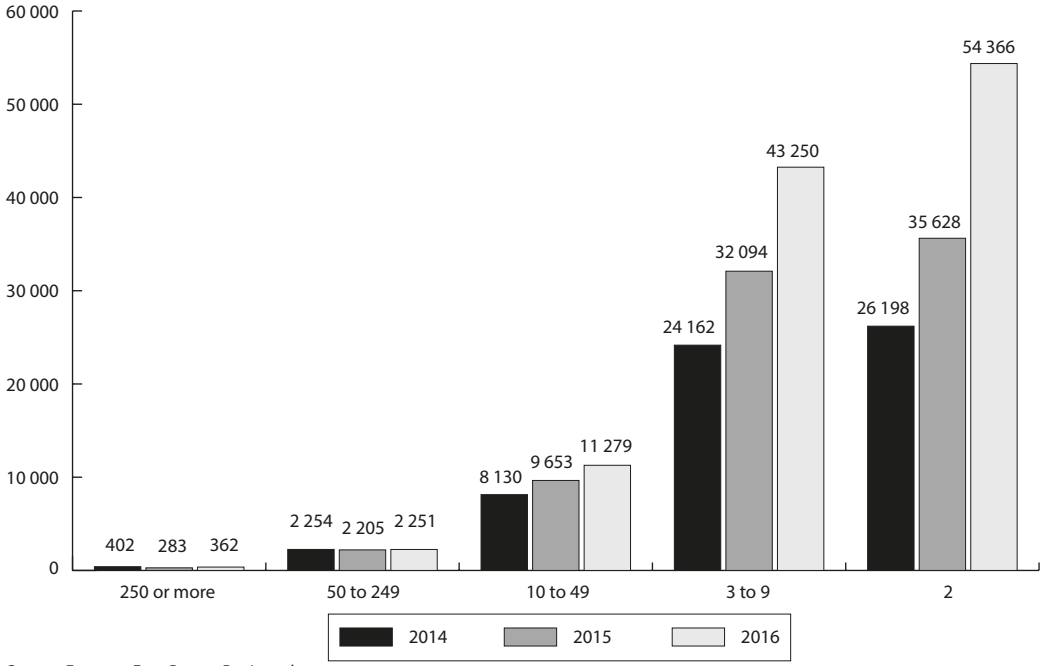


Source: Eurostat, EuroGroups Register data

The EGR has very broad coverage of large and medium-sized groups. However small groups and their constituent enterprises are covered to a lower extent. The analysis is based on micro-level comparisons of the EGR population with national data of EU countries on foreign controlled enterprises and enterprises with foreign subsidiaries. These comparisons show that EGR covers very well the large and medium sized multinational groups (groups having 250 or more employed persons). In contrary there

are small groups, mostly present in two countries only, which are part of the national populations of foreign controlled enterprises and enterprises with foreign subsidiaries, but those are not yet present in the EGR. To reach the targeted full coverage, EGR should cover more small multinational groups.

Figure 5 Multinational enterprise groups in EGR by number of legal units



Source: Eurostat, EuroGroups Register data

4 DATA SHARING AND USE OF THE EGR

The EGR is a statistical register with restricted use. The EGR data with the harmonised the multinational groups are only accessible for statisticians producing national statistics in the EU and EFTA NSIs and national central banks.

The EGR provides harmonised information for the users on:

- enterprise groups, enterprises, legal units (identification, demographic and economic characteristics),
- relationships and control,
- shareholdings of at least 10 per cent.

Statistical data producers can use the EGR as:

- coordinated frame population,
- database to derive consistent statistical output,
- auxiliary source to improve quality of statistical output,
- input for measuring global activities,
- database of European enterprises who are part of multinational groups.

The EGR data serve only for statistical purposes, i.e. for the production of the official statistics related to globalisation in the EU Member States and EFTA countries, for example:

- Statistics on foreign affiliates (FATS);
- Foreign direct investment statistics (FDI);
- Statistics on small and medium sized enterprises (SBS);

- Statistics on trade by enterprise characteristics (TEC);
- National accounts – statistics on gross national income (GNI);
- Short term statistics (STS).

5 NEW DATA SOURCES FOR THE EGR

Additional sources of information such as crowdsourcing platforms, web crawling and different open data projects are seen as further opportunities to increase the quality of the EGR, its completeness and accuracy namely with the units outside of the EU and EFTA as well as the aggregate indicators on the whole group level.

Under the umbrella of Eurostat BIG DATA project, Eurostat EGR Team is investigating these additional data sources. Eurostat is collaborating with Leipzig University to explore the possibility of using DBpedia as new additional source of data of MNE groups. DBpedia is a project which extracts structured information from Wikipedia to make it publically available in a format that allows to ask sophisticated queries against Wikipedia and to link different data sets to Wikipedia data. The objective of this project is to set-up a mostly automatized at a large extend the collection of aggregated group figures using as input the names of the enterprise groups.

Eurostat also follows the works of the Global Legal Entity Identifier Foundation (GLEIF) on the globally unique legal entity identifiers as well as on the GLEIF level-2 data on relationships of legal units. Eurostat also collaborates with OECD that develops a project on ADIMA database. This cooperation can be mutually beneficial in the future, when ADIMA will reach certain level of maturity. This additional data source could benefit for the EGR when providing information on units for non EU and non EFTA countries that are related by control relationships to the European enterprises.

6 PUBLICLY AVAILABLE DATA DISSEMINATION FROM THE EGR

In an effort to better respond to users' needs Eurostat publishes experimental statistics on the related Eurostat website based on the EGR data.

The EGR experimental statistics article focuses on the population of the active multinational enterprise groups in the EU and EFTA countries.

The EGR defines an active multinational enterprise group a group that has employees in two or more countries, at least one of them in the EU or EFTA. According to the EGR, 47 621 active multinational groups were registered in the EU in 2016.

Of these groups, 40 081 were from EU Member States (their global decision centres were in the EU Member States) and the rest from other countries, mostly from Switzerland (2 843) the United States (2 200), Norway (441), followed by Japan (350) and Canada (200).

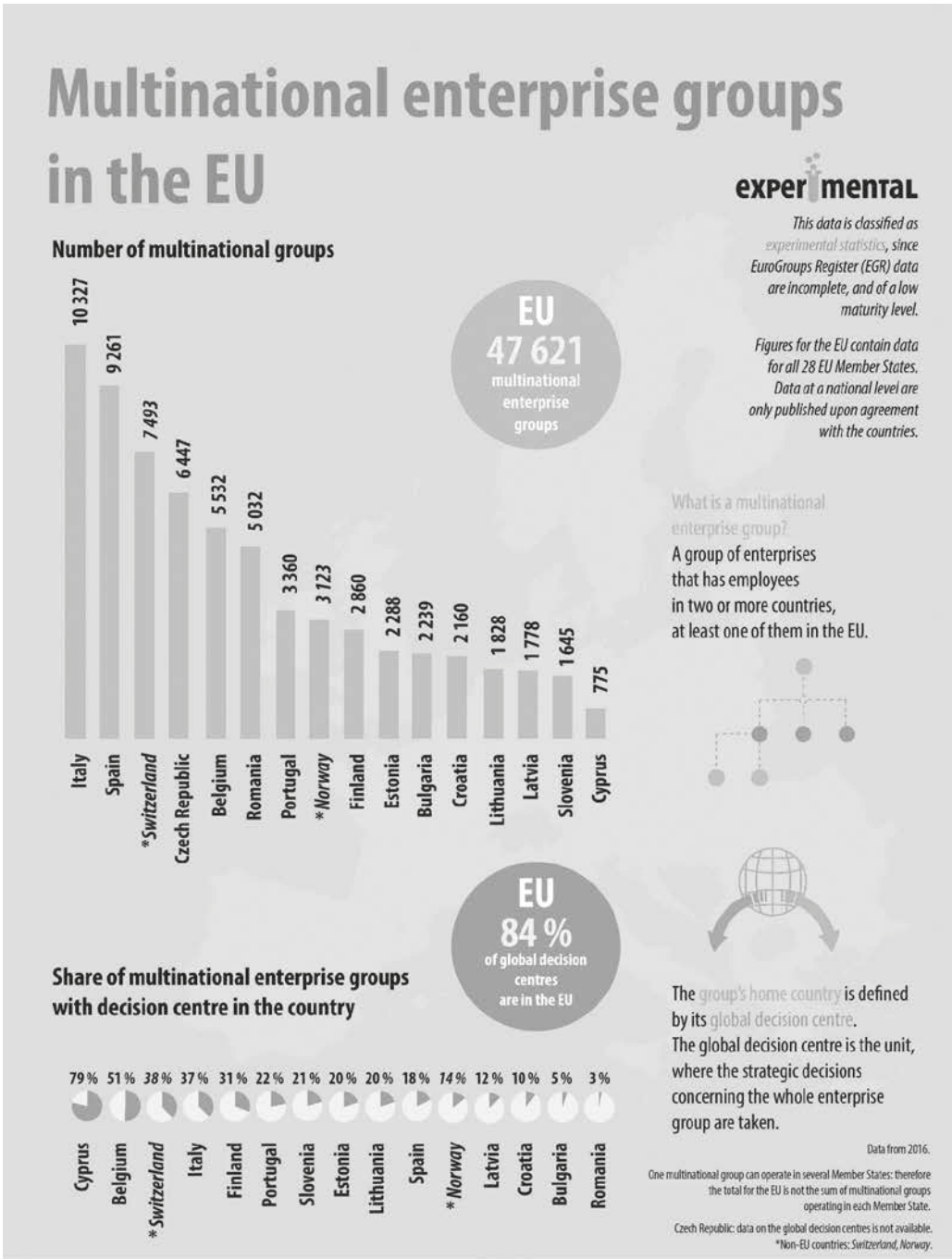
The EGR is a very rich data source, Eurostat and the NSIs are continuously working to improve its quality. Having sufficient quality level in the future, the EGR can be the source of very relevant and valuable European official statistics.

CONCLUSION – OUTLOOK FOR THE EGR PRODUCTION AND USE

Eurostat and NSIs are continuously working to improve the quality of EGR. Further improvements are planned on the EGR processes, improvement of the continuity algorithm, on the EGR coverage, accuracy and timeliness. Additional data sources will be analysed and used in the future.

The main goal of Eurostat is to enhance the use of EGR in official statistics. Statistical compilers of the EU and EFTA countries can use EGR as coordinated frame population of multinational enterprise groups and their units to derive consistent statistical output when measuring activities related to globalisation.

Figure 6 Infographics on multinational enterprise groups in the EU



Source: Eurostat, EuroGroups Register 2016 data

References

EUROSTAT. *EuroGroups Register* [online]. Luxembourg, 2018. <https://ec.europa.eu/eurostat/web/structural-business-statistics/structural-business-statistics/eurogroups-register>.

EUROSTAT. S*tatistics Explained, EuroGroups register* [online]. Luxembourg, 2018. <https://ec.europa.eu/eurostat/statistics-explained/index.php/EuroGroups_register>.

EUROSTAT. *Statistics Explained, EuroGroups register FATS online interface* [online]. Luxembourg, 2018. <https://ec.europa.eu/eurostat/statistics-explained/index.php?title=EuroGroups_register_FATS_online_interface>.

EUROSTAT. *Statistics Explained, EuroGroups register identification service* [online]. Luxembourg, 2018. <https://ec.europa.eu/eurostat/statistics-explained/index.php?title=EuroGroups_register_identification_service>.

Eurostat. *Statistics Explained, Structure of multinational enterprise groups in the EU* [online]. Luxembourg, 2018. <https://ec.europa.eu/eurostat/statistics-explained/index.php/Structure_of_multinational_enterprise_groups_in_the_EU>.

A Centenary of the State Statistical Office

Prokop Závodský [1] | *University of Economics, Prague, Czech Republic*
Ondřej Šimpach [2] | *University of Economics, Prague, Czech Republic*

Abstract

In the Austrian part of the Habsburg monarchy, official statistics were controlled by the Vienna central authorities (the Central Statistical Commission, Ministry of Commerce and other ministries) and statistical offices of local authorities in the individual countries (Bohemia, Moravia, Silesia and others), whose competences were very limited. Insufficient cooperation and coordination of activities of these statistical offices had been subject to criticism as well as the lack of general reporting obligations.

The Czechoslovakian Act regulating the state statistics approved by the Revolutionary National Assembly, three months after the establishment of a new state centralized statistical service to the Statistical State Council as a Deputy Authority and the State Statistical Office as an executive body, and also enacted general reporting obligations and guarantees of consistent protection of individual data. The current statute of the state statistical service in the Czech Republic is actually based on the same principles as the Act approved a century ago. The paper describes the emergence of Act No. 49/1919 and a long and complicated path leading to its fulfilment. The protracted problem was also the post of the chairman of both statistical authorities and the provision of a suitable object for activity of the State Statistical Office.

Keywords

Statistical State Council, State Statistical Office, Dobroslav Krejčí, Karel Engliš, František Weyr

JEL code

B16, B23, B31

INTRODUCTION

During the celebration of 100th anniversary of the birth of an independent Czechoslovak state and its most important bodies held on 28th January 2019 also statisticians remembered 100 years since the adoption of the Act on Czechoslovak State Statistics, which established the State Statistical Office, a predecessor of current Czech Statistical Office.

The founding of the Czechoslovak state statistical service over the past century has been dealt with by a number of authors in specialized statistical literature, often at the occasion of various anniversaries. For example, Podzimek (1974 and 1989), *70 let* (1989), Závodský (1999) etc., some of which were partly uncritical to the information and opinions received from the "father of Czechoslovakian statistics" D. Krejčí. In our contribution, based on original sources, we try to bring a somewhat new look at the events a hundred years ago.

[1] University of Economics, Prague, Faculty of Informatics and Statistics, Department of Statistics and Probability, W. Churchill Sq. 4, 130 67 Prague 3, Czech Republic. E-mail: prokop.zavodsky@vse.cz.
[2] University of Economics Prague, Faculty of Informatics and Statistics, Department of Statistics and Probability, W. Churchill Sq. 4, 130 67 Prague 3, Czech Republic. E-mail: ondrej.simpach@vse.cz.

1 IN THE REVOLUTIONARY NATIONAL ASSEMBLY

On 14th November 1918, after the birth of the Czechoslovak Republic, the Revolutionary National Assembly (RNA), the non-elected state's supreme representative body, started its activity. It operated until the spring 1920, when the constitution was adopted, and the first parliamentary elections were held (when choosing its name, the representative body was inspired by the legislative assembly that came to light during the Great French Revolution – l´Assemblée nationale.) The mandates were distributed among the individual Czech political parties based on the results of the last (relatively democratic) elections to the House of Deputies of the Vienna Imperial Council in the year 1911. Politicians from Slovakia, who had not yet recognized democratic elections, delegated their representatives by themselves. Minority nationalities were not represented in the RNA.

Figure 1 Conference Hall of the Revolutionary National Assembly at Thun Palace in 1919

Source: *Národní shromáždění* (1919)

The Revolutionary National Assembly met in the building of the former Landtag in Bohemia in Sněmovní Street, in today's the seat of the Chamber of Deputies of the Parliament of the Czech Republic. Already the 22nd meeting of RNA took place on 28th January 1919. Roughly at 17 o'clock the chairman of RNA F. Tomášek interrupted the discussion about the issue of municipal elections and invited Karel Engliš, Member of the Parliament, to present a report of the social-political committee on its draft law on the organization of the statistical service in the new state. Professor of national economy at the Czech Technical University in Brno and Member of Parliament for the Czech Constitutionalist Party, K. Engliš, in his speech explained briefly and concisely the principles of the proposed law and the need for its rapid adoption. He also recommended that the Statistical State Council (SSC) should be established as an advisory and quaint body, State Statistical Office (SSO) as an executive body. The above mentioned authorities should centralize all the statistical service in the Republic and directly report to the Prime Minister. A more detailed organization cannot be enacted at present but will be defined by government regulation later. More details were available to the Members of Parliament in the report of the social-political Committee of 10th January (print 323).

Present participants (more than two hundred from 256 Members of Parliament) listened to Engliš's speech. No comments have been made, so it immediately passed to the vote. The Members of Parliament that agreed, stood up at the call of the chairman of RNA and he then proclaimed that it is the majority (it was voted four times, always on several successive paragraphs). Immediately thereafter, the chairmen asked "gentlemen" (actually 8 women were members of RNA), to rise, if they agree with the text of the law also in the second (last) reading. He again declared the majority and thus the adoption of the law.[3]

Figure 2 Karel Engliš as the first rector of Masaryk University (1919–1920)

Note: The law on the establishment of Masaryk University in Brno was approved by RNA also on 28th January 1919.
Source: *Čs. statistický věstník,* 1930, Vol. XI, p. 494a (F. Hlavica portrait)

The law was signed by the president T. G. Masaryk (it is not the first law signed by him, (as sometimes indicated) and also by the Prime Minister Kramář on behalf of the interior minister Švehla. The law was published within the stated eight-day period on 5th February 1919 in the Collection of Laws and Regulations (figure XI/1919) as Law No. 49 of 28th January 1919 Coll., about the Organization of Statistical Service, which became effective.

2 STATE AND LAND STATISTICS IN AUSTRIA-HUNGARY

The organizational structure of official statistics in the Habsburg monarchy was rather complicated. Both Austria and Hungary had their own, entirely independent statistical service, organized differently in many ways. In the Austrian part of the monarchy, the representative of the state statistics was the I. r. Central Statistical Commission (K. k. statistische Zentralkommission), that was incorporated under the Ministry of Religious Affairs and Education. State statistics was, however, decentralized to a large extent. Majority of Austrian ministries founded their own statistical offices or bureaus. The most important was the Department of Commerce, with the competences covering statistics of foreign trade and trade with Hungarian part of monarchy, social statistics and partly also the statistics of industry and transport.

The individual countries of Cisleithania had some limited autonomy. First, various civil servants dealt with statistics of various provincial affairs (provincial elections, health and charity institutes, public buildings, etc.), later, specialized Land Statistical Offices were established (Bohemia 1897, Moravia 1899,[4] Silesia 1907). They overtook also the agricultural statistics under their competences in their countries.[5]

[3] *Společná česko-slovenská digitální parlamentní knihovna – Národní shromáždění československé* – stenoprotokoly.
[4] Sometimes stated year 1893 is a mistake.
[5] Closer see: Závodský and Šimpach (2018).

The Land Statistical Offices operated independently of the Vienna Central Statistical Commission and were jealously guarding their small competencies. Their activity was partly coordinated by the Conference on Land Statistics (Konferenz für Landesstatistik). Its sessions took place alternately in various countries of Cisleithania and published statistical yearbooks of provincial governments.

In the Hungarian part of the monarchy the statistical service was centralized into the Royal Central Statistical Office of Hungary (Magyar Királyi Központi Statisztikai Hivatal) based in Budapest. It was subordinate to the Hungarian Ministry of Commerce. The only territory in Hungary with a certain autonomy – the Croatian-Slavonic Kingdom[6] – had its own statistical office in Zagreb. Save for some exceptions (e.g. foreign trade statistics), this autonomous authority conducted all statistical surveys on the territory of Croatia-Slavonia exclusively by itself. In case of events related to Hungarian territory it was directed by the instructions from statistical office in Budapest.

Figure 3 Dobroslav Krejčí

Source: Wikipedia (open encyclopedia), Dobroslav Krejčí [cit. 15.1.2019]

Work conditions of the Hungarian Statistical Office were codified in 1897 by a special law.[7] The detailed plan of statistical surveys was submitted to the Hungarian Parliament for approval by the Minister of Trade. Persons and legal entities were also required to provide proper information for such investigations. The penalty for non-compliance could be the search for necessary data at the expense of the culprit or fine. Interesting was the enforced obligation of village teachers to work as census commissioners in censuses and surveys on education. The law also provided the protection of individual data.

Let's return to the territory of today's Czech Republic. In the framework of its limited competencies and resources, the Land Statistical Office of the Bohemian Kingdom had developed extensive activities. It consisted of the Land Statistical Committee as the expert body of the advisory and quieter and the Land Statistical Bureau as an executive body. Since 1905 Dobroslav Krejčí[8] had been the head of the office. He had been employed there since its foundation and managed the office for whole 14 years only as Deputy Chief. The post of the Chief of Staff was not occupied to avoid a dispute with representatives of the German minority in Bohemia who demanded German office management. Even so, no consensus was found, many German districts and municipalities in Bohemia refused to provide required data.

In no area of activity of the Land Statistical Offices, a reporting duty was enacted, which meant permanent problems for these authorities. Reporting duty was imposed in the monarchy in only four cases by special legal norms. These included the census, the census of agricultural and trade businesses, foreign

[6] It is part of the territory of today's Croatia. Podzimek (1974, p. 109) replaced Slavonia for Slovenia, which was not part of Hungary.

[7] Here we take the information from publication of former director of the Hungarian Statistical Office – Buday (1915).

[8] Dobroslav Krejčí (1869–1936) graduated at the Faculty of Law of the Czech University in Prague (doctorate in 1891), then worked as a trainee lawyer.

trade statistics and "inter-trade" (trade between the Austrian and Hungarian parts of the monarchy). The Central Statistical Commission (not having any branches outside the capital) and other Viennese offices could, however, carry out their investigations in individual countries with the help of governorship and district governors, whose extensive powers ensured that respondents would not refuse to provide the required data. The authority of the Landtag in Bohemia and its executive body – the Land Committee, which the land statisticians could turn to, was, on the other hand, small. The Landtag was, in fact, blocked by the obstructions of German deputies from 1908, and, finally, in the sum-mer of 1913, it was together with the Land Committee dissolved by the imperial patents. It did not restore its activities until the demise of Austria-Hungary.

3 THE PATH TO THE CZECHOSLOVAK STATISTICAL LAW

The Land Statistical Office performed less activity in Moravia. Besides the finances and deeper interest of the provincial politicians, it lacked a qualified and statistically enlightened personality, as was in Prague D. Krejčí with his co-workers.

The hope for change for the better appeared in June 1913 when Karel Engliš, extraordinary professor of national economy and statistics at the Czech Technical University in Brno, was elected to the Moravian Landtag.[9] On 3rd February 1914, Engliš submitted a proposal of statistical law for Moravia, that would impose on municipalities and district road committees[10] the obligation to provide the Land Statistical Office with the information for investigation approved by the Land Committee. As a penalty for non-provision of statistical data, it was determined that he necessary data were obtained at the cost of the offending institution (§ 3). This proposal signed another 41 Members of the Parliament.

Engliš's proposal was approved with minor change by two committees of Landtag. On 28th February, it easily passed also its final reading in Moravian Landtag and after having been signed by the emperor and competent Austrian ministers it was promulgated on 18th December 1914 as Law No. 92/1914 in the Moravian Land Code of Law. For the first time, a general (abstract) reporting duty was enacted in our territory.

By the irony of fate, the law came into force only at the time when the Land Statistical Office in Brno actually shut down its activities – after many of its staff was called to the army in relation with the outbreak of War. Engliš's detailed proposal of the new Statute of the Land Statistical Office, which should have ensured its extension and improvement of its work, had failed to be discussed by the Landtag, because it was closed due to the declaration of war and its activity were not restored.

The Land Statistical Office in Bohemian continued to a limited extent its work also during the First World War. It focused mainly on statistics of agriculture and food production, in which it was supported by the Vienna Ministry of Agriculture and the Prague governorship. Publishing activity was limited due to the lack of finance and confidentiality of survey results. Problems of war economy and especially rapid worsening supplies of the army and the population had increased the interest in economic statistics throughout the monarchy. The situation had evoked a discussion in professional circles about the need to reorganize the statistical service in the Austrian part of the monarchy. Vienna Professor of National Economics and senior representative of the Central Statistical Commission, Karl Přibram,[11] published in monthly Statistische Monatschrift an article

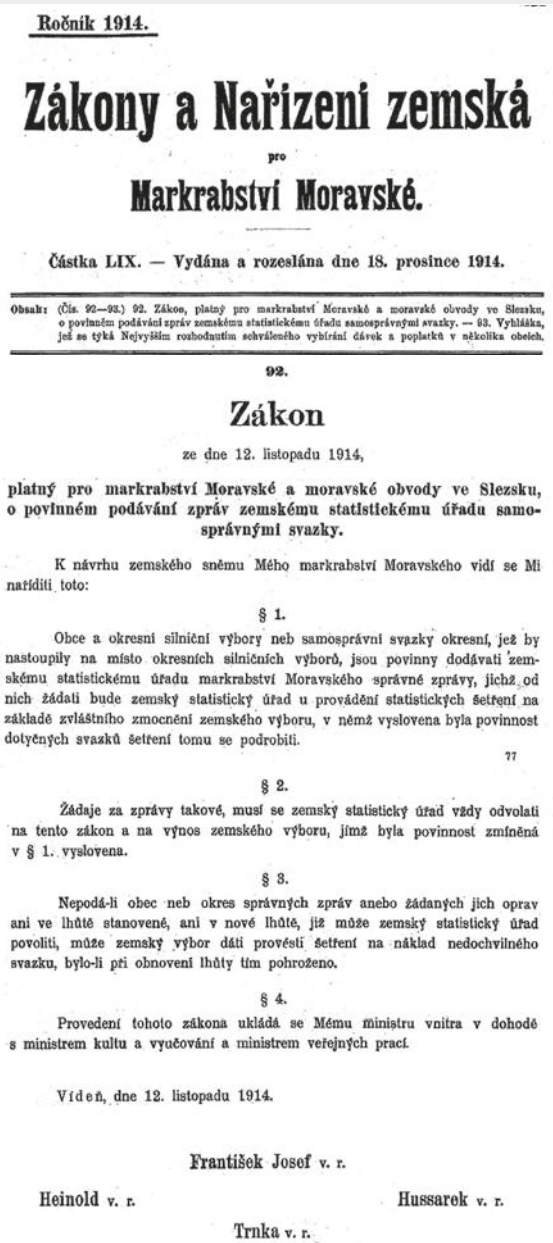
[9] Karel Engliš (1880–1961) was concerned mainly by statistics since his graduation at law faculty (1904) – at the Land Statistical Office in Prague (1904–1908) and at ministry of trade in Vienna (1908–1911).

[10] Unlike Bohemia, there were no self-government authorities at the district level in Moravia; there were only district road committees, which were responsible only for the construction and management of district roads.

[11] Karl Přibram (1877–1973) was born in an educated Jewish family in Prague and graduated here at the Law Faculty of the German University (1900).

Die amtliche Statistik Österreichs *am Scheidewege (Austrian official statistics at the crossroads)*, in which he criticized unsatisfactory state of the statistical service in Cisleithania (Přibram, 1915). He proposed, according to the organization of the Hungarian official statistics, also in the Austrian part of the monarchy to merge the statistical services of individual ministries with the Central Statistical Commission into one office superior to all land and municipal statistical offices.

Figure 4 Moravian statistical law



Ročník 1914.

Zákony a Nařízení zemská

pro

Markrabství Moravské.

Částka LIX. — Vydána a rozeslána dne 18. prosince 1914.

Obsah: (Čís. 92—93.) 92. Zákon, platný pro markrabství Moravské a moravské obvody ve Slezsku, o povinném podávání zpráv zemskému statistickému úřadu samosprávnými svazky. — 93. Vyhláška, jež se týká Nejvyšším rozhodnutím schváleného vybírání dávek a poplatků v několika obcích.

92.

Zákon

ze dne 12. listopadu 1914,

platný pro markrabství Moravské a moravské obvody ve Slezsku, o povinném podávání zpráv zemskému statistickému úřadu samosprávnými svazky.

K návrhu zemského sněmu Mého markrabství Moravského vidí se Mi nařídíti toto:

§ 1.

Obce a okresní silniční výbory neb samosprávní svazky okresní, jež by nastoupily na místo okresních silničních výborů, jsou povinny dodávati zemskému statistickému úřadu markrabství Moravského správné zprávy, jichž od nich žádati bude zemský statistický úřad u provádění statistických šetření na základě zvláštního zmocnění zemského výboru, v němž vyslovena byla povinnost dotyčných svazků šetření tomu se podrobiti.

77

§ 2.

Žádaje za zprávy takové, musí se zemský statistický úřad vždy odvolati na tento zákon a na výnos zemského výboru, jímž byla povinnost zmíněná v § 1. vyslovena.

§ 3.

Nepodá-li obec neb okres správných zpráv anebo žádaných jich oprav ani ve lhůtě stanovené, ani v nové lhůtě, již může zemský statistický úřad povoliti, může zemský výbor dáti provésti šetření na náklad nedochvilného svazku, bylo-li při obnovení lhůty tím pohrožено.

§ 4.

Provedení tohoto zákona ukládá se Mému ministru vnitra v dohodě s ministrem kultu a vyučování a ministrem veřejných prací.

Vídeň, dne 12. listopadu 1914.

František Josef v. r.

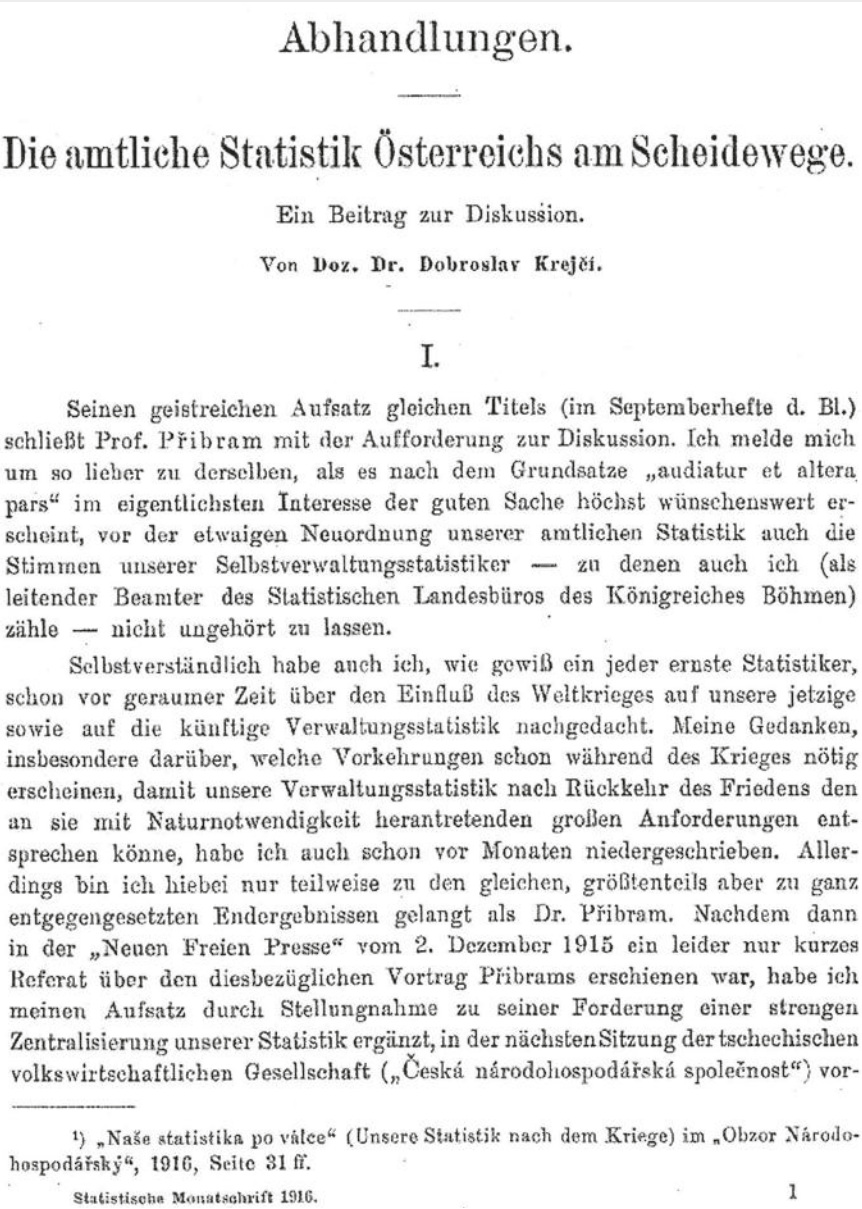
Heinold v. r. Hussarek v. r.

Trnka v. r.

Source: Zákony a Nařízení zemská pro Markrabství Moravské (Laws and Land Regulations for Margrave Moravia), Vol. 1914, Iss. LIX, pp. 421

At the invitation of the discussion in the conclusion of Přibram's article, D. Krejčí responded by a Czech article (Krejčí, 1916b) and later by its extended version published in *Statistische Monatschrift* under the same name as Přibram's contribution (Krejčí, 1916a). Krejčí agreed with the opinion of Přibram about the need to enact a general reporting obligation. He also published his own concrete draft solution – experts should prepare as soon as possible a concise and clear "statistical law" that could be issued urgently during the time of Parliament's closure in form of imperial regulation.

Figure 5 Title sheet of Krejčí's article in *Statistische Monatschrift*



Abhandlungen.

Die amtliche Statistik Österreichs am Scheidewege.

Ein Beitrag zur Diskussion.

Von Doz. Dr. Dobroslav Krejčí.

I.

Seinen geistreichen Aufsatz gleichen Titels (im Septemberhefte d. Bl.) schließt Prof. Přibram mit der Aufforderung zur Diskussion. Ich melde mich um so lieber zu derselben, als es nach dem Grundsatze „audiatur et altera pars" im eigentlichsten Interesse der guten Sache höchst wünschenswert erscheint, vor der etwaigen Neuordnung unserer amtlichen Statistik auch die Stimmen unserer Selbstverwaltungsstatistiker — zu denen auch ich (als leitender Beamter des Statistischen Landesbüros des Königreiches Böhmen) zähle — nicht ungehört zu lassen.

Selbstverständlich habe auch ich, wie gewiß ein jeder ernste Statistiker, schon vor geraumer Zeit über den Einfluß des Weltkrieges auf unsere jetzige sowie auf die künftige Verwaltungsstatistik nachgedacht. Meine Gedanken, insbesondere darüber, welche Vorkehrungen schon während des Krieges nötig erscheinen, damit unsere Verwaltungsstatistik nach Rückkehr des Friedens den an sie mit Naturnotwendigkeit herantretenden großen Anforderungen entsprechen könne, habe ich auch schon vor Monaten niedergeschrieben. Allerdings bin ich hiebei nur teilweise zu den gleichen, größtenteils aber zu ganz entgegengesetzten Endergebnissen gelangt als Dr. Přibram. Nachdem dann in der „Neuen Freien Presse" vom 2. Dezember 1915 ein leider nur kurzes Referat über den diesbezüglichen Vortrag Přibrams erschienen war, habe ich meinen Aufsatz durch Stellungnahme zu seiner Forderung einer strengen Zentralisierung unserer Statistik ergänzt, in der nächsten Sitzung der tschechischen volkswirtschaftlichen Gesellschaft („Česká národohospodářská společnost") vor-

[1]) „Naše statistika po válce" (Unsere Statistik nach dem Kriege) im „Obzor Národohospodářský", 1916, Seite 31 ff.

Statistische Monatschrift 1916. 1

Source: *Statistische Monatschrift*. Vol. XXI (1916, p. 1)

However, Krejčí argued with Přibram's proposal on centralization of all statistical services in Austrian part of the monarchy. He suggested to solve the current unsatisfactory state by the coordination of the work of state and self-governing statistical authorities according to a single plan of all statistical actions. That would be gain for the land (or newly established) statistical authorities, which better know the specifics of their countries, are closer to the data sources and, therefore, have better prospects for successful data collection and control.

Polemics of Přibram and Krejčí continued in the spring 1916 in Vienna during their common visit of the President of Central Statistical Committee Viktor Mataja who took a neutral position in the dispute and called for further discussion with wider participation of experts (Krejčí, 1919, p. 2). Polemics, however, did not continue anymore. It was clear that the future reorganization of statistical service in the monarchy will necessarily depend on the reform of the entire state administration.

Already before 28th October 1918, D. Krejčí prepared the first version of the outline of the law on the organization of state statistics in the new state, which he presented after the putsch (2nd November) at economic-statistical committee of the National Committee. However, this commission soon ceased its activities. The National Committee was extended to the Revolutionary National Assembly. A constitutional constitution was adopted and a leading government Karel Kramář was appointed as PM of the government. Krejčí submitted to him his draft law together with an explanatory report on 22nd November (Krejčí, 1919, pp. 4–5).

In spite of dramatic international and internal political situation of the new state, Krejčí pushed through the relentlessly rapid adoption of the law on state statistics. Four days later, he found a member of the RNA Engliš and passed his suggestions to him. Engliš (after the agreement with the Prime Minister Kramář) submitted a draft law as his own deputy proposal. However, Engliš made Krejčí's proposal briefer and simpler to increase the chances for rapid adoption of the law. At that time, the law could contain only the main principles of the organization of state statistics, as there has not yet been a decision on the system of central offices of the republic nor on its division into territorial units. Engliš's law proposal signed also by other 22 members of parliament[12] was with a brief explanation printed as a parliamentary print number 147 dated 1st December 1918. Presidency of RNA passed the draft law for discussion to social-political committee (10th December).[13]

During December and the beginning of January in a number of places Engliš slightly amended the text of the concise statistical law, partly also according to the comments of D. Krejčí and his colleagues (J. Auerhan, J. Mráz). Social-political committee of RNA discussed the draft law on 9th January. Engliš, as the committee's reviewer, explained in detail and justified the principles of the new law. Czechoslovak state statistics must be centralized into the Statistical State Council (SSC) as an advisory and decision-making body and the State Statistical Office (SSO) as an executive body. Chairman of SSC will be jointly a chairman of SSO and both bodies will be directly subordinated to the Prime Minister. Organization and method of work of SSC and SSO will be stated later by government regulations, as well as the structure of possible regional statistical authorities.

However, the law already requires a clear reporting obligation for all the investigations decided by SSC. Now, it was possible to be inspired by statistical laws in Moravia, Hungary and some foreign states. The law must also include possible sanctions for non-compliance (fine, investigating the charges of guilty, possibly imprisonment), but also sufficient guarantees of protection of the provided individual data, which cannot be disclosed or given to another office (namely the tax office). Land Statistical Bureau in Prague with its staff and inventory will become the basis for building a much larger SSO. The law does not mention the future fate of the Land Statistical Offices in Brno and Opava(see Figure 6).

[12] Minimum established by the rules of procedure of RNA was 20.

[13] *Společná česko-slovenská digitální parlamentní knihovna – Národní shromáždění československé* – stenoprotokoly.

Figure 6 Act on the Organization of Statistical Service in the Collection of Laws and Regulations

Ročník 1919. 59

Sbírka zákonů a nařízení státu československého.

Částka XI. Vydána dne 5. února 1919.

Obsah: (Čís. 49.—56.) 49. Zákon o organisaci statistické služby. — 50. Zákon, kterým se zřizuje druhá česká universita. — 51. Zákon o upotřebení části správních přebytků hromadných sirotčích pokladen. — 52. Zákon, jímž se poskytují drahotní přídavky aktivním a pensionovaným učitelům veřejných, obecných a měšťanských škol, jakož i vdovám a sirotkům po těchto učitelích na dobu od 1. ledna 1919 do konce června 1919. — 53. Nařízení, jímž se zrušuje císařské nařízení ze dne 9. srpna 1915, z. ř. č. 234, o zcizování hospodářských a lesních pozemků. — 54. Nařízení o prodloužení lhůty ku dodatečnému započítání služební doby, kterou kancelářští oficianti před svým ustanovením strávili jako kancelářští pomocníci nebo ve vojenské službě presenční. — 55. Nařízení o přísaze státních úředníků, úředníků podniků a zřízenců již ustanovených. — 56. Nařízení o obsazování míst radů a projednávání osobních a disciplinárních záležitostí u vrchního zemského soudu v Praze a obsazování míst u soudů a státních zastupitelstev.

Čís. 49.

Zákon ze dne 28. ledna 1919

o organisaci statistické služby.

Na základě usnesení Národního shromáždění se nařizuje:

§ 1.

Službu státní statistiky organisovati, říditi i prováděti bude pro obvod celé Československé republiky

a) Statistická Rada Státní, jakožto orgán poradní a usnášející,

b) Státní Úřad Statistický, jakožto orgán výkonný.

§ 2.

Statistická Rada Státní a Státní Úřad Statistický jsou bezprostředně podřízeny předsedovi ministerstva. V čele Statistické Rady Státní a zároveň Státního Statistického Úřadu stojí odborový přednosta jako předseda.

§ 3.

Složení, kompetence a způsob jednání Statistické Rady Státní upraví se nařízením celého ministerstva. Nařízením ministerského předsedy stanoví se též složení a působnost Státního Úřadu Statistického, jakož i nižších orgánů statistických.

§ 4.

Všechny statistické a jiné úřady ve státě, státní, samosprávné, vojenské, církevní atd. jsou povinny podporovati Státní Úřad Statistický při plnění jeho úkolů a dbáti jeho pokynů v mezích usnesení Statistické Rady Státní.

§ 5.

Každý obyvatel státu, vyjímaje osoby požívající exteritoriality, jest povinen podávati správně, pravdivě a včasně všechny údaje a zprávy, jež bude po něm požadovati Státní Úřad Statistický přímo nebo nepřímo jinými orgány neb úřady podle usnesení Statistické Rady Státní.

Porušení této povinnosti se bude trestati politickými úřady první stolice pokutou do 1000 K a při opětování nebo za přitěžujících okolností vězením do šesti měsíců a pokutou bez omezení.

Státní Statistický Úřad může mimo to provésti šetření přímo nebo nepřímo na útraty provinilého.

§ 6.

Ustanovení § 5. platí též pro všechny veřejné místní i zájmové svazky ve státě, jakož i pro spolky, družstva, a výdělečné společnosti všeho druhu. Za splnění jejich statistické povinnosti zodpovědní jsou osobně ti, kdož po zákonu jsou oprávněni zastupovati příslušný svazek, spolek atd. na venek.

§ 7.

Uveřejňováním statistických dat nesmí býti odhalovány individuální soukromé poměry. Uveřejnění individuálních statistických dat Státním Statistickým Úřadem předpokládá schválení Statistické Rady Státní.

Neuveřejněné individuální statistické výkazy nesmí býti ani k nahlédnutí přenechány žádnému jinému úřadu, zvláště ne úřadům berním. Individuálních statistických výkazů nesmí býti použito za základ pro ukládání daní.

Soukromé zpracování úředního statistického materiálu za podmínek § 8. smí připustiti předseda Státní Rady Statistické.

11

60

§ 8.

Všichni, jichž prostřednictvím se statistická data sbírají, zpracovávají neb uveřejňují, jsou povinni zachovávati o všech věcech při statistickém šetření zjištěných, pokud se dotýkají poměrů soukromých, přísnou mlčenlivost vůči každému, také vůči jiným nežli statistickým úřadům, zvláště berním.

Porušení této povinnosti, jakož i každé zneužití zpráv a údajů, získaných pro účely statistické, se trestá bez újmy případné soukromé náhrady škody řádnými trestními soudy pokutou do 10.000 K nebo vězením do dvou let, po prvé jako přestupek, při opětování nebo za přitěžujících okolností jako přečin.

§ 9.

„Zemská statistická kancelář král. Českého v Praze“ béře se se vším personálem i zařízením a se všemi svými právy i závazky do služeb státních a pověřuje se úkoly Státního Úřadu Statistického. „Zemská Statistická komise král. Českého“ se zrušuje.

§ 10.

Zákon tento nabývá účinnosti dnem vyhlášky. Provedení jeho se ukládá předsedovi ministerstva.

T. G. Masaryk v. r.

Švehla v. r., v zast. min. předsedy.

Source: *Sbírka zákonů a nařízení státu československého.* XI/1919 (pp. 59–60)

The social-political committee of the Revolutionary National Assembly unanimously approved the law submitted (in other words print 147 with minor changes done and justified by Engliš) and recommended the plenary of RNA to approve it. The final draft of the law, with its extensive Engliš's justification, was printed with the date 10th January 1919. It was handed out to the Members of Parliament as a print 323. The discussion of the law was listed as forth point of the program of 21st meeting of RNA held on 24th January. However, it was submitted at the next meeting on 28th January, when the law was accepted unmodified without objections (see above).

4 THE FIRST STEPS OF THE STATE STATISTICAL OFFICE

The approved Act on the state statistical service envisaged the issue of statutes of the Statistical State Council and the State Statistical Office in form of a government regulation and the appointment of the joint chairman of the two institutions (by the President of the Republic at the proposal of the Prime Minister). In this sense, the law had not been carried out for 10 months.

D. Krejčí with K. Engliš prepared a draft statute of the SSC and the SSO at the beginning of February, but the Prime Minister's office decided to postpone the establishment of the two institutions in the approval of the forthcoming amendment to Act No. 49. The draft amendments to the law, prepared by the Ministry of the Interior, envisaged the limitation of SSO's independence, its subordination to the Ministry f the Interior, and some other changes. The draft amendment was finally withdrawn, and Act No. 49/1919 Coll. was then applied throughout the First Republic. There has been no progress in organizing state statistics.

Figure 7 The oldest seat of SSO in Šeříková street No. 4/618

Source: Photo V. Puci (2019)

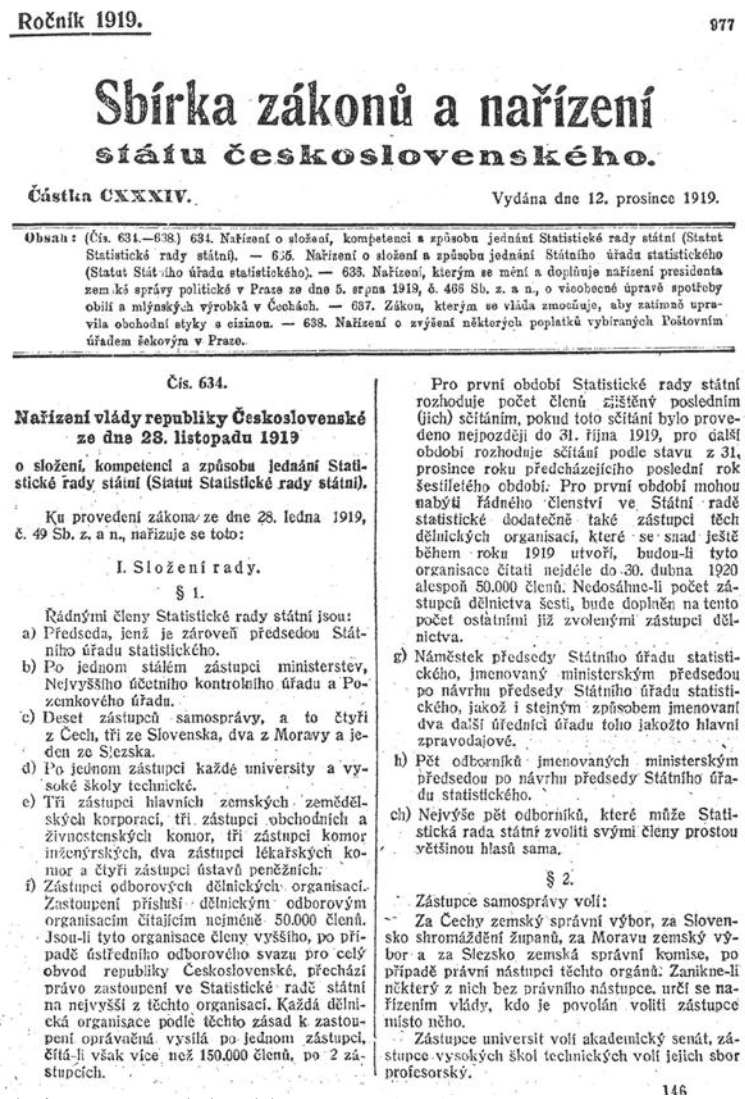
Inpatient Krejčí constantly presented and urged his proposals. The long-term presence of the Prime Minister at peace talks in Paris, riots at the border, conflicts with Hungary, and the economic and social problems of the new state had led to the permanent postponement of the organization's statistical service.

Paragraph § 9 of the Act No. 49 was the only legal basis for action of SSO: *"Land Statistical Bureau of the Kingdom of Bohemia in Prague … is delegated by the tasks of the State Statistical Office" …* Despite that, the "father of Czechoslovakian statistics" Krejčí and his collaborators managed to develop to certain extent the activity of SSO. Gradually, new workers were recruited (during year 1919 increased the number of employees from 34 on 220). Among the most important we can mention Robert Kollar, Jan Náhlovský and Antonín Boháč. The biggest problem was the lack of space equipment. SSO worked only in several modified rooms of the apartment building in Šeříková street at Malá Strana (Figure 7) that were "inherited" by the Land Statistical Bureau.

The solution to the government's failure to apply the statistical law has not improved even after the appointment of the new govern-ment of Vlastimil Tusar. Disappointed Krejčí announced him the resignation to the managing post of SSO on 14th July 1919 (he has not been officially nominated yet to the leading position). Resignation was not accepted. Krejčí received a decree appointing him

as chairman of the SSO on 13th August. Since he was not nominated as the head of the SSC and as chief of a subdivision (sub-department) (both assumed the law No. 49), Krejčí perceived it as *difficult, unjustified humiliation of his person.*[14] He refused the nomination and insisted on his resignation. The resignation was not decided during many months. The SSO leadership was actually taken over by Jan Auerhan.[15]

Figure 8 Statute of the Statistical State Council in the Collection of Laws and Regulations (1st page)



Ročník 1919. 977

Sbírka zákonů a nařízení státu československého.

Částka CXXXIV. Vydána dne 12. prosince 1919.

Obsah: (Čís. 634.—638.) 634. Nařízení o složení, kompetenci a způsobu jednání Statistické rady státní (Statut Statistické rady státní). — 635. Nařízení o složení a způsobu jednání Státního úřadu statistického (Statut Státního úřadu statistického). — 636. Nařízení, kterým se mění a doplňuje nařízení presidenta zemské správy politické v Praze ze dne 5. srpna 1919, č. 466 Sb. z. a n., o všeobecné úpravě spotřeby obilí a mlýnských výrobků v Čechách. — 637. Zákon, kterým se vláda zmocňuje, aby zatímně upravila obchodní styky s cizinou. — 638. Nařízení o zvýšení některých poplatků vybíraných Poštovním úřadem šekovým v Praze.

Čís. 634.

Nařízení vlády republiky Československé ze dne 28. listopadu 1919

o složení, kompetenci a způsobu jednání Statistické rady státní (Statut Statistické rady státní).

Ku provedení zákona ze dne 28. ledna 1919, č. 49 Sb. z. a n., nařizuje se toto:

I. Složení rady.

§ 1.

Řádnými členy Statistické rady státní jsou:

a) Předseda, jenž je zároveň předsedou Státního úřadu statistického.

b) Po jednom stálém zástupci ministerstev, Nejvyššího účetního kontrolního úřadu a Pozemkového úřadu.

c) Deset zástupců samosprávy, a to čtyři z Čech, tři ze Slovenska, dva z Moravy a jeden ze Slezska.

d) Po jednom zástupci každé university a vysoké školy technické.

e) Tři zástupci hlavních zemských zemědělských korporací, tři zástupci obchodních a živnostenských komor, tři zástupci komor inženýrských, dva zástupci lékařských komor a čtyři zástupci ústavů peněžních.

f) Zástupci odborových dělnických organisací. Zastoupení přísluší dělnickým odborovým organisacím čítajícím nejméně 50.000 členů. Jsou-li tyto organisace členy vyššího, po případě ústředního odborového svazu pro celý obvod republiky Československé, přechází právo zastoupení ve Statistické radě státní na nejvyšší z těchto organisací. Každá dělnická organisace podle těchto zásad k zastoupení oprávněná vysílá po jednom zástupci, čítá-li však více než 150.000 členů, po 2 zástupcích.

Pro první období Statistické rady státní rozhoduje počet členů zjištěný posledním (jich) sčítáním, pokud toto sčítání bylo provedeno nejpozději do 31. října 1919, pro další období rozhoduje sčítání podle stavu z 31. prosince roku předcházejícího poslední rok šestiletého období. Pro první období mohou nabýti řádného členství ve Státní radě statistické dodatečně také zástupci těch dělnických organisací, které se snad ještě během roku 1919 utvoří, budou-li tyto organisace čítati nejdéle do 30. dubna 1920 alespoň 50.000 členů. Nedosáhne-li počet zástupců dělnictva šesti, bude doplněn na tento počet ostatními již zvolenými zástupci dělnictva.

g) Náměstek předsedy Státního úřadu statistického, jmenovaný ministerským předsedou po návrhu předsedy Státního úřadu statistického, jakož i stejným způsobem jmenovaní dva další úředníci úřadu toho jakožto hlavní zpravodajové.

h) Pět odborníků jmenovaných ministerským předsedou po návrhu předsedy Státního úřadu statistického.

ch) Nejvýše pět odborníků, které může Statistická rada státní zvoliti svými členy prostou většinou hlasů sama.

§ 2.

Zástupce samosprávy volí:

Za Čechy zemský správní výbor, za Slovensko shromáždění županů, za Moravu zemský výbor a za Slezsko zemská správní komise, po případě právní nástupci těchto orgánů. Zanikne-li některý z nich bez právního nástupce, určí se nařízením vlády, kdo je povolán voliti zástupce místo něho.

Zástupce universit volí akademický senát, zástupce vysokých škol technických volí jejich sbor profesorský.

146

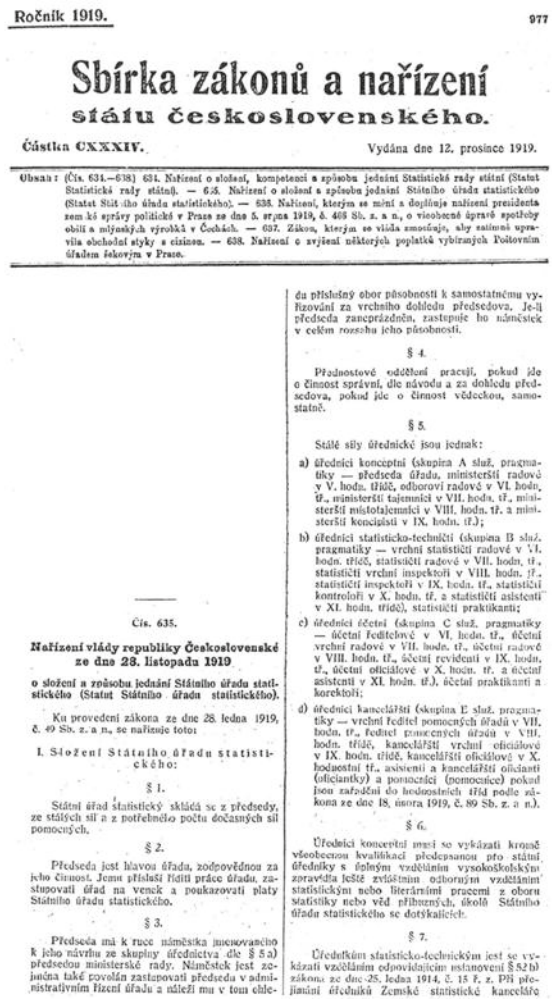
Source: *Sbírka zákonů a nařízení státu československého.* CXXXIV/1919 (p. 977)

[14] Krejčí (1919, p. 16). Krejčí actually managed the Land Statistical Bureau already since 1905 without being appointed its head.

[15] Jan Auerhan (1880–1942) studied law at the Czech University in Prague (graduation 1904) and worked in the Land Statistical Bureau in Prague since year 1906. He served in army in years 1914–1918.

According to Act No. 49 the SSC was competent to determine the program and methods of statistical surveys. Until it was established, a special law had to be adopted in urgent cases. When the RNA discussed the government draft law on foreign trade statistics on 18th September 1919, it approved on the recommendation of its Financial Committee (the rapporteur was K. Engliš) besides the government draft law also two resolutions. They challenged the government to "*with the fastest acceleration finally put into the life both authorities of state statistical service* (SSC and SSO) *established by the Act from* 28*th January 1919, No. 49 Coll.*", appoint their chairman, issue their statutes and to "*immediately provide adequate room and sufficient staff*" for SSO. At the same time, RNA demanded the necessary autonomy for SSB in its activities.[16] Government inaction in the matter of enforcement of Act No. 49 was also a subject of interpellation of a Member of RNA Rudolf Beran (and 21 others) on 14th November.[17]

Figure 9 Statute of the State Statistical Office in the Collection of Laws and Regulations (1st page)



Ročník 1919. 977

Sbírka zákonů a nařízení
státu československého.

Částka CXXXIV. Vydána dne 12. prosince 1919.

Obsah: (Čís. 634.—638.) 634. Nařízení o složení, kompetenci a způsobu jednání Statistické rady státní (Statut Statistické rady státní). — 635. Nařízení o složení a způsobu jednání Státního úřadu statistického (Statut Státního úřadu statistického). — 636. Nařízení, kterým se mění a doplňuje nařízení presidenta zemské správy politické v Praze ze dne 5. srpna 1919, č. 405 Sb. z. a n., o všeobecné úpravě spotřeby obilí a mlýnských výrobků v Čechách. — 637. Zákon, kterým se vláda zmocňuje, aby zatímně upravila obchodní styky s cizinou. — 638. Nařízení o zvýšení některých poplatků vybíraných Poštovním úřadem šekovým v Praze.

Čís. 635.

Nařízení vlády republiky Československé ze dne 28. listopadu 1919

o složení a způsobu jednání Státního úřadu statistického (Statut Státního úřadu statistického).

Ku provedení zákona ze dne 28. ledna 1919, č. 49 Sb. z. a n., se nařizuje toto:

I. Složení Státního úřadu statistického:

§ 1.

Státní úřad statistický skládá se z předsedy, ze stálých sil a z potřebného počtu dočasných sil pomocných.

§ 2.

Předseda jest hlavou úřadu, zodpovědnou za jeho činnost. Jemu přísluší řídit práce úřadu, zastupovati úřad na venek a poukazovati platy Státního úřadu statistického.

§ 3.

Předseda má k ruce náměstka jmenovaného k jeho návrhu ze skupiny úřednictva dle § 5 a) předsedou ministerské rady. Náměstek jest zejména také povolán zastupovati předsedu v administrativním řízení úřadu a náleží mu v tom ohledu příslušný obor působnosti k samostatnému vyřizování za vrchního dohledu předsedova. Je-li předseda zaneprázdněn, zastupuje ho náměstek v celém rozsahu jeho působnosti.

§ 4.

Přednostové oddělení pracují, pokud jde o činnost správní, dle návodu a za dohledu předsedova, pokud jde o činnost vědeckou, samostatně.

§ 5.

Stálé síly úřednické jsou jednak:

a) úředníci konceptní (skupina A služ. pragmatiky — předseda úřadu, ministerští radové v V. hodn. třídě, odboroví radové v VI. hodn. tř., ministerští tajemníci v VII. hodn. tř., ministerští místotajemníci v VIII. hodn. tř. a ministerští koncipisti v IX. hodn. tř.);

b) úředníci statisticko-technický (skupina B služ. pragmatiky — vrchní statističtí radové v VI. hodn. třídě, statističtí radové v VII. hodn. tř., statističtí vrchní inspektoři v VIII. hodn. tř., statističtí inspektoři v IX. hodn. tř., statističtí kontroloři v X. hodn. tř. a statističtí asistenti v XI. hodn. třídě), statističtí praktikanti;

c) úřednící účetní (skupina C služ. pragmatiky — účetní ředitelové v VI. hodn. tř., účetní vrchní radové v VII. hodn. tř., účetní radové v VIII. hodn. tř., účetní revidenti v IX. hodn. tř., účetní oficiálové v X. hodn. tř. a účetní asistenti v XI. hodn. tř.), účetní praktikanti a korektoři;

d) úřednící kancelářští (skupina E služ. pragmatiky — vrchní ředitel pomocných úřadů v VII. hodn. tř., ředitel pomocných úřadů v VIII. hodn. třídě, kancelářští vrchní oficiálové v IX. hodn. třídě, kancelářští oficiálové v X. hodnostní tř., asistenti a kancelářští oficianti (oficiantky) a pomocníci (pomocnice) pokud jsou zařaděni do hodnostních tříd podle zákona ze dne 18. února 1919, č. 89 Sb. z. a n.).

§ 6.

Úřednící konceptní musí se vykázati kromě všeobecnou kvalifikací předepsanou pro státní úředníky s úplným vzděláním vysokoškolským zpravidla ještě zvláštním odborným vzděláním statistickým nebo literárními pracemi z oboru statistiky nebo věd příbuzných, úkolů Státního úřadu statistického se dotýkajících.

§ 7.

Úředníkům statisticko-technickým jest se vykázati vzděláním odpovídajícím ustanovení § 52 b) zákona ze dne 25. ledna 1914, č. 15 ř. z. Při přejímání úředníků Zemské statistické kanceláře

Source: *Sbírka zákonů a nařízení státu československého.* CXXXIV/1919 (p. 980)

[16] *Společná česko-slovenská digitální parlamentní knihovna – Národní shromáždění československé* – stenoprotokoly.
[17] Ibidem, print 1884; answer V. Tusar, print 2043.

On 28th November 1919 the statute of SSC and SSO was finally approved, both government regulations came into force by publishing in the Collection of Laws and Regulations on 12th December under No. 634 and 635. The deputies of ministries, other central authorities, representatives of universities and technicians, local authorities, trade unions and statisticians including senior SSO were appointed as members of SSC. Tenure was six years. Members worked in committees for various statistical branches. The SSC Plenary Meeting approved the SSO's annual work plan. It was binding also for natural and legal persons that were the subjects of reporting obligations. Resolution of SSC could be cancelled only by the government.

Statute of the SSO laid down the work conditions of the office according to the resolutions of the SSC, the relation of the SSO to the Office of the Prime Minister, to ministries etc. At the beginning, the SSO consisted of the presidium and 6 departments, which gradually started their activities, most often in the autumn of 1919.

At the end of 1919 a critical lack of space for an action of SSO was addressed. To the benefit of the SSO a large object of the Deaf Institute in Smíchov (former Karlova, present Holečkova street No. 4/104) was confiscated on 26th October. However, it was vacated and partially adapted for the purposes of statistical office gradually over the next two years. However, the eviction of a charitable institution from its own building did not contribute to the popularity of the statistical office in the public. In addition, an object with large halls, but lack of small workrooms, was not very suitable for the SSO.

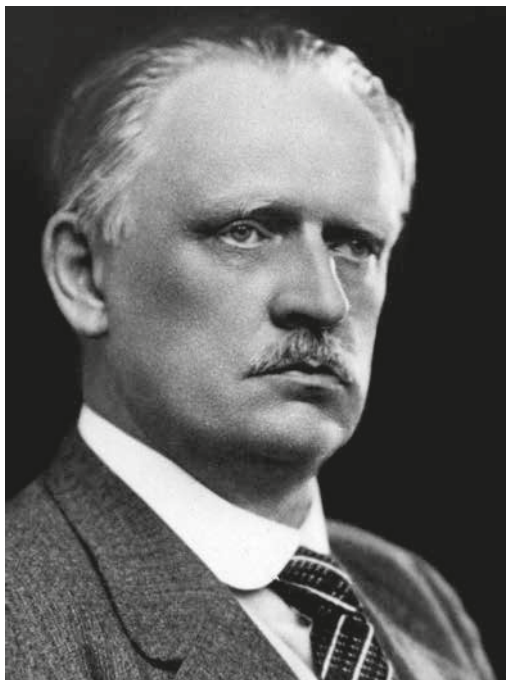
Figure 10 Seat of State Statistical Office of in years 1920–1929 (contemporary postcard)

Source: The collection of postcards of P. Závodský

In February 1920, the issue of management of the two central statistical authorities was finally resolved. On 3rd February resignation of D. Krejčí was accepted and on 19th February his successor, František Weyr, was installed to the office. He was at that time a professor of constitutional law at the Faculty

of Law in Brno and, at the same time, its first dean.[18] Since he could not simultaneously perform two posts in the state administration, he was appointed chairman of SSC and was only commissioned by the post of a chairman of SSO (he also received only a minor part of the salary on this position). Jan Auerhan was appointed vice-chairperson of SSO. Thanks to F. Weyr and K. Engliš a position of a professor of statistics at the Faculty of Law in Brno for D. Krejčí[19] was created.

Figure 11 František Weyr



Source: Photo from Archive of Masaryk's University in Brno

Figure 12 Jan Auerhan

Source: Sekanina (1927, p. 154)

New head of SSO F. Weyr was entirely a different personality than F. Krejčí who was of tireless and of pedantic nature. As he mentioned in his memoirs (Weyr 2001, pp. 128–131) he led the SSO under the motto of the administrative of an antient Roma "*minima non curat praetor*".[20] In practice, it meant that he commuted to the statistical office from Brno usually every other week for less than three days. He then chaired the meetings of some committees of the SSC, had appointments with the senior officials of the office, made decisions on personal and other important issues and negotiated at the office of the Prime Minister, eventually with some ministers. He delegated management of current office affairs to the heads of SSO departments and other conceptual staff.

[18] Also, František Weyr (1879–1951) studied law at the Czech University in Prague (doctorate 1904). He also worked at the Central Statistical Commission in Vienna (1905–1908) and in Land Statistical Bureau in Prague (1909–1912). Since year 1912 he was professor of law sciences at Czech Technical University in Brno, and since the year 1919 a professor at Law Faculty of Masaryk's University. In years 1918–1920 sat in RNA (for the Czech Constitutionalist Party – like K. Engliš). He was as a member of the constitutional committee one of the authors of the new constitution of Czechoslovakia.

[19] Weyr (2001, pp. 55–57). Krejčí had lectured before (since 1909) as honoured associate professor an agricultural statistics at Czech Technical University in Prague.

[20] Praetor (a high official in Rome) does not care about insignificant issues.

During the first months of the year 1920 the Czechoslovak state statistics began to fully operate. On 24 March the first (extraordinary) meeting of the SSC of 49 members delegated by central authorities and other institutions under the current statute of SSC was held, 5 other members from ranks of renowned statisticians chose the plenum of the SSC. According to the proposal of deputy chairman of SSO, J. Auerhan, the members of SSC approved a work plan of SSO for the year 1920 and appointed 14 committees of SSC for various sectoral statistics. In the coming weeks, SSC committees have already systematically started their work. Number of SSC members gradually increased. Further enlargement (up to 80 employees) happened after an amendment of the SSC statute by a government decree of 20 August 1920 (see References). Member of SSC were at that time almost all important representatives of statistics and related branches in Czechoslovakia.

Second (1st regular) plenary meeting of the SSC, held on 25th June 1920, discussed and approved the work plan of the SSO for the year 1921, when the first Czechoslovak census was planned (on 15th February).

According to the Statistics Act No. 49/1919 Coll., the Land Statistical Bureau became part of the SSO since 5th February 1919, but formally the employees were transferred already by the government decree on 21st May 1920. Number of employees of SSO was further increased. By the end of the year 1920 the office employed as many as 505 persons (from which were 295 of women, none of them in any important position).

Even under difficult conditions, the SSO managed to launch publishing activities in 1920. A statistical magazine Československý *statistický věstník* (Czechoslovakian Statistical Bulletin) was published in the first year. It was the predecessor of today's quarterly journal *Statistika*. In June 1920 *Statistická příručka republiky* Československé (Statistical handbook of Czechoslovak Republic) was published with 106 pages 235 tables providing pre and post-war data about the area of a new state. At the same time, individual sheets *Zprávy Státního* úřadu *statistického* (Reports of State Statistical Office) began to appear. They offered various, often preliminary, results of elaborated statistical surveys. By the end of the year 1920 a number of 22 issues of *Zprávy* (Reports) were published.[21]

CONCLUSION

In this article we focused on the establishment and initial development of the Statistical State Council and the State Statistical Office, especially in terms of legislation and organization. In the follow-up paper (in any of the next volumes of the *Statistika* journal) we will focus in detail on professional and scientific activities of both bodies of state statistics in Czechoslovakia. We will pay attention to the most important persons and publications of SSO to the roughly first half of 20s of the previous century.

ACKNOWLEDGEMENT

This paper was processed with the long-term institutional support of research activities by the Faculty of Informatics and Statistics, University of Economics, Prague.

References

AUERHAN, J., eds. Statistika. In: *Deset let Československé republiky*, Prague: Státní tiskárna v Praze, 1928, Iss. 3, pp. 579–638.
BUDAY, L. V. Die Entwicklung der ungarischen amtlichen Statistik. *Allgemeines statistisches Archiv,* 1915, Vol. 9, pp. 197–212.
Československý statistický věstník. 1920–1922, Vol. I–III.

[21] *Statistická příručka* (Statistical handbook) and *Zprávy SSO* (Reports of SSO) were issued also in French and German versions. In *Čs. statistický věstník* were bilingually – in Czech and French – published the full text of statistical laws and government regulations, the content of the magazines and some statistical tables.

DUBSKÝ, S. et al., eds. *70 let československé státní statistiky: sborník.* Prague: FSÚ: Výzkumný ústav sociálně ekonomických informací a automatizace v řízení, 1989.

KOLLAR, R. Organisace československé služby statistické u srovnání s organisací statistické služby v bývalém Rakousko-Uhersku. *Čs. statistický věstník*, 1922, Vol. IV, Iss. 1–3, pp. 52–64.

KREJČÍ, D. Die amtliche Statistik Österreichs am Scheidewege. Ein Beitrag zur Diskussion. *Statistische Monatschrift*, 1916a, Vol. XXI, pp. 1–17.

KREJČÍ, D. Naše statistika po válce. *Obzor národohospodářský*, 1916b, Vol. XIX, Iss. 1, pp. 31–40.

KREJČÍ, D. *K historii vzniku Státního* úřadu *statistického.* Manuscript 1919.

Národní shromáždění československé v prvním roce republiky (L'assemblée nationale tchécoslovaque dans la première année de la république) 1918–1919. Prague: Předsednictvo Národního shromáždění, 1919.

Nařízení vlády republiky Československé č. 634 Sb. z. a n., ze dne 28. listopadu 1919 o složení, kompetenci a způsobu jednání Statistické rady státní (Statut Statistické rady státní).

Nařízení vlády republiky Československé č. 635 Sb. z. a n., ze dne 28. listopadu 1919 o složení a způsobu jednání Státního úřadu statistického (Statut Státního úřadu statistického).

Nařízení vlády republiky Československé č. 493 Sb. z. a n., ze dne 20. srpna 1920, jímž částečně se změňuje, popř. doplňuje nařízení vládní ze dne 28. listopadu 1919, č. 634 Sb. z. a n. o složení, kompetenci a způsobu jednání Statistické rady státní (Statut Statistické rady státní).

PODZIMEK, J. *Vývoj čs. statistiky do vzniku Státního úřadu statistického*. Prague: SNTL, 1974.

PODZIMEK, J. *Vývoj československé statistiky 1919–1945 v událostech a datech.* Prague: Výzkumný ústav sociálně ekonomických informací, 1979.

PODZIMEK, J. Z historie prvního čs. zákona o statistice. *Statistika*, 1989, Vol. 26, Iss. 1, pp. 15–21.

PŘIBRAM, K. Die amtliche Statistik Österreichs am Scheidewege. *Statistische Monatschrift*, 1915, Vol. XX, pp. 483–494.

SEKANINA, F., eds. *Album representantů všech oborů veřejného života československého.* Prague: Umělecké nakladatelství Josef Zeibrdlich, 1927.

Společná česko-slovenská digitální parlamentní knihovna – Národní shromáždění československé – stenoprotokoly [online]. [cit. 15.1.2019]. <https://www.psp.cz/eknih/1918ns/ps/stenprot/index.htm>.

WEYR, F. *Paměti 2. Za republiky (1918–1938).* Brno: Atlantis, 2001.

Zákon č. 49 Sb. z. a n., ze dne 28. ledna 1919 o organisaci statistické služby.

Zákon č. 518 Sb. z. a n., ze dne 18. září 1919 o statistice zahraničního obchodu.

ZÁVODSKÝ, P. 85 let od vzniku státní statistické služby na území České republiky. *Statistika,* 2005, Vol. 42, Iss. 3, pp. 254–262.

ZÁVODSKÝ, P. AND ŠIMPACH, O. 120th Anniversary of Founding the Land Statistical Office of the Bohemian Kingdom [online]. *Statistika, Statistics and Economy Journal,* 2018, Vol. 98, Iss. 2, pp. 185–201.

Comparison of Statistical Yearbooks of Czechoslovakia, Czech Republic and Slovak Republic 1920, 1925 and 2017

Jiří Pelej [1] | *Czech Statistical Office, Prague, Czech Republic*

Abstract

The article focuses on comparison of contents of statistical handbooks published by the Czechoslovak State Statistical Office in the years 1920 and 1925 with that of statistical yearbooks of Czechoslovakia's successor states – Statistical Yearbook of the Czech Republic 2017 and Statistical Yearbook of the Slovak Republic 2017.

Keywords

Statistical handbook, statistical yearbook, Czech Statistical Office, Statistical Office of the Slovak Republic

JEL code

Y10, Z00, C10, C49

1 STATISTICAL HANDBOOK OF CZECHOSLOVAKIA 1920

Not long after the birth of the Czechoslovak Republic on 28 January 1919, the State Statistical Office was established by the Act No. 49 of the Collection of Laws and Ordinances. Roughly half and a year later, the State Statistical Office issued the first yearbook in a new independent state: Statistical Handbook of the Czechoslovak Republic.

As opposed to the expectations the issue was delayed due to late establishment of the State Statistical Council and also due to final determination of the new republic's borders to which some issues of the yearbook were linked. The published data presented mainly the results of the latest census conducted in 1910, i.e. before the WW I. In this respect it is useful to mention the statement of the then president of the Statistical Office, Jan Auerhan: "*Data, presented in the Handbook are mostly outdated, coming from the pre-war period. It is understandable given that the State Statistical Office due to a disastrous lack of space could not be properly built and only some departments could be put into operation. Many offices that were charged under the previous administration with the duty to collect statistical data, had failed to fulfil the task because of the lack of time and staff. This situation had caused that from many branches the latest data were missing. After all, the pre-war data, i.e. data from the years 1910–1914, cannot be considered outdated, on the contrary, they offer even more realistic and correct picture of our republic than recent data proving that we still suffer from war consequences.*"

[1] Czech Statistical Office, Na padesátém 81, 100 82 Prague 10, Czech Republic. E-mail: jiri.pelej@czso.cz.

This handbook has 106 numbered pages structured into 14 sections + appendices with tables:

I. Area and description of the population	VIII. Mining and metallurgy
II. Occupation of the population	IX. Small businesses, industry and trade
III. Movements of the population	X. Transportation
IV. Health care	XI. Credit granting
V. Education	XII. Insurance
VI. Administration	XIII. Justice
VII. Agriculture	XIV. Finance

The biggest space and attention is devoted to the data from section I Area and description of the population (15 p), followed by section VII Agriculture (12 p), section IX Small businesses, industry and trade and XI Credit granting (both 11 p). Movements of the population in 1914–1918 in Bohemia, Moravia and in Opava regions, review of the results of elections to the National Assembly in 1920, data on coal mining in 1919, distilleries and breweries in the Czech lands in 1918–1919, etc., are shown in appendices. Majority of data, however, as stated above, refer to the period before 1918, most often to the year 1910. The above suggests that with respect to the contents and scope (only 106 p) the first Czechoslovak statistical handbook is not suitable for comparison with present yearbooks.

2 STATISTICAL HANDBOOK OF CZECHOSLOVAKIA 1925

If we wish to compare the contents of yearbooks at the beginning of the existence of Czechoslovakia with the present yearbook, it is more proper to compare the yearbooks of the Czech Republic and Slovak Republic (2017) with the second *Statistical handbook of the Czechoslovak Republic* (1925). Numerical section of this publication is divided into two parts: introductory page of the first part is called "Data from 1918–1923", followed by the second section: "Retrospective data".

The publication by its scope (hundreds pages) comes closer to a usual scope of statistical yearbooks (e.g. Statistical Yearbook of the Czech Republic 2017 has 822 pages).

In the handbook of the year 2015 the following sections are included:

Section I: Data of the years 1918–1923	
I. Educations and enlightment	XI. Administration. Elections.
II. Agriculture	XII. Military forces and security service
III. Mining and metallurgy	XIII. Area and description of the population
IV. Industry and trade licences	XIV. Occupations
V. External trade	XV. Movements of the population
VI. Prices, indices, standard of living	XVI. Healthcare and physical training
VII. Banking	XVII. Social statistics
VIII. Insurance	XVIII. Finance
IX. Transportation	XIX. International surveys
X. Justice	
Section II: Retrospective data	
I. Area and description of the population	II. Occupations

3 STATISTICAL YEARBOOK OF THE CZECH REPUBLIC 2017

The Czech statistical yearbook 2017 has the total of 819 pages, of which 753 pages include tables, charts and cartograms. The remaining part represents the introduction, subject index and a list of persons contributing to the yearbook containing the following numbered chapters:

1. Key national economy indicators	17. Construction
2. Area and climate	18. Trade, hotels and restaurants
3. Environment	19. Tourism
4. Population	20. Transportation, information and communication
5. National accounts	21. Market services
6. Finance	22. Information society
7. Currency and balance of payments	23. Science, research and innovation
8. Prices	24. Education
9. Household income and expenditure	25. Health
10. Labour market	26. Social security
11. External trade	27. Culture
12. Organizational structure of the national economy	28. Sport
13. Agriculture	29. Justice, crime, accidents
14. Forestry	30. International comparisons
15. Industry	31. Selected indicators of cohesion regions and regions
16. Energy	32. Elections

4 STATISTICAL YEARBOOK OF THE SLOVAK REPUBLIC 2017

The second successor state of Czechoslovakia is the Slovak Republic. The structure of the yearbook is similar to that of the yearbook of the Czech Republic: it has 661 numbered pages followed by graphic annex (13 charts, 5 cartograms). Its contents is structured into the below chapters:

1. Primary Economic Indicators	17. Industry
2. Territory and Climate	18. Energy
3. Population	19. Construction
4. Labour Market	20. Trade, Catering and Accommodation
5. Income, Expenditure and Consumption of Households	21. Transport, Storage and Postal Services
6. Education	22. Information and Communication, Information Services
7. Health Care	23. Selected Market Services
8. Social Protection	24. Tourism
9. Culture and Religion	25. Management of Non-Financial and Financial Corporations
10. National Accounts	26. Environment
11. External Trade	27. Science and Technology
12. Prices	28. Criminality and Justice
13. Fixed Assets	29. Accidents and Disasters
14. Organizational Statistics	30. Gender Equality
15. Agriculture	31. International Surveys
16. Forestry	

5 TRANSCARPATHIA

Transcarpathia was a part of Czechoslovakia in the years 1918–1938. Its territory was occupied by the end of the 1930's by Hungary and after the WW II, in 1945 it was enacted to the Union of Soviet Socialist Republics (Ukraine). Some data about former Transcarpathia can be found under the region called Zakarpattya in publications *Statistical Yearbook of Ukraine* and *Ukraine in figures* (accessible e.g. in the library of the Czech Statistical Office).

CONCLUSION

If we compare statistical handbook of 1925 with chapters of the Czech yearbook 2017, in the handbook the following parts are missing: Selected indicators of national economy, Environment, National accounts, Currency and balance of payments, Household income and expenditure, Energy, Tourism and Selected indicators of cohesion regions and regions.

In the Czech yearbook we do not find a counterpart to Slovak chapter no. 13 Fixed assets, 25 Economic activity of non-financial and financial corporations and 30 Gender equality.

In the Slovak yearbook we do not find an independent counterpart to the following Czech chapters: 6 Finance (about budget activity, from data provided by the Ministry of Finance of the Czech Republic), and 7 Currency and Balance of payment (from data of the Czech National Bank), chapters 28 Sport and also 31 Selected indicators of cohesion regions and regions.

A significant influence on formation of the structure of the statistical offices' publications had and still has the expanding application of international classification of national economy braches, introducing unified systems of national accounts of which our between-war statistics did not have a single idea, further, the statistics of environment and tourism, general development of science, research and development, and, naturally, the information society.

References

CZECH STATISTICAL OFFICE. *Statistická ročenka České Republiky 2017* (Statistical Yearbook of the Czech Republic 2017). Prague: CZSO, 2017.

STATISTICAL OFFICE OF THE SLOVAK REPUBLIC. *Štatistická ročenka Slovenskej republiky 2017* (Statistical Yearbook of the Slovak Republic 2017). Bratislava: VEDA, 2017.

STATE STATISTICAL OFFICE. *Statistická příručka Republiky Československé I.* (Statistical Handbook of Czechoslovakia I). Prague, 1920.

STATE STATISTICAL OFFICE. *Statistická příručka Republiky Československé II.* (Statistical Handbook of Czechoslovakia II). Prague, 1925.

How to Turn Quality into a Habit in the Statistical Production?

Paula Silva [1] | *Banco de Portugal, Lisboa, Portugal*
Margarida Pinto [2] | *Banco de Portugal, Lisboa, Portugal*
António Agostinho [3] | *Banco de Portugal, Lisboa, Portugal*

Abstract

One of the main purposes of the Statistics Department of Banco de Portugal is to ensure a statistical production with high quality standards aiming at fully meeting users' needs, aligned with the best practices and procedures recommended by the international organizations. Following its commitment to quality, one of the Bank's priorities is to develop a wide set of quality control procedures that ensure high levels of regular and thorough review of the key statistical outputs.

Statistical quality control is based on different procedures and working arrangements that make sure that processes are effective and efficient and the risks are mitigated. In order to achieve higher quality statistics, there are several quality indicators performed by the primary statistics' compilers.

This paper will present the main quality indicators used and the ongoing process to improve the model of regular and systematic quality controls.[4]

Keywords

Statistical quality control, quality assessment, quality indicators

JEL code

G20, L15, M42

INTRODUCTION

The statistical data published by Banco de Portugal complies with the quality management guidelines and best practices laid down in national and international documents like ESCB Public Commitment, European Statistics Code of Practice, ECB Statistics Quality Framework, and IMF Data Quality Assessment

[1] Banco de Portugal, Statistics Department, Statistics Audit Unit, R. Francisco Ribeiro 2, 1150-165 Lisboa, Portugal. E-mail: paasilva@bportugal.pt.
[2] Banco de Portugal, Statistics Department, Statistics Audit Unit, R. Francisco Ribeiro 2, 1150-165 Lisboa, Portugal. E-mail: mmpinto@bportugal.pt.
[3] Banco de Portugal, Statistics Department, Statistics Audit Unit, R. Francisco Ribeiro 2, 1150-165 Lisboa, Portugal. E-mail: afagostinho@bportugal.pt.
[4] This paper is based on contribution at the *European Conference on Quality in Official Statistics (Q2018)* in June 2018 in Krakow, Poland. The opinions expressed in the article are those of the author(s) and do not necessarily coincide with those of Banco de Portugal or the Eurosystem. Any errors and omissions are the sole responsibility of the author(s).

Framework (DQAF). DQAF includes a set of prerequisites, and five dimensions of data quality: (i) assurances of integrity, (ii) methodological soundness, (iii) accuracy and reliability, (iv) serviceability and (v) accessibility.

With these dimensions in mind, Banco de Portugal statisticians have implemented effective and efficient statistical procedures throughout the statistical production chain, in line with principle 8 of ESCB Public Commitment – "Appropriate statistical procedures". Moreover, their routines are driven by "High output quality" principles (towards relevant, accurate, reliable, timeliness, consistent and accessible statistics).

Nevertheless, it is difficult to assess to what extent they are daily engaged to principle 4 of the same document – "Commitment to quality" (it systematically and regularly identifies strengths and weaknesses to continuously improve process and product quality). For that purpose, Banco de Portugal is currently defining a model of quality indicators to systematically measure the quality of its statistical systems and outputs.

Turning the performance of quality indicators into a habit will allow to: (i) document quality control procedures, (ii) make comparisons amongst different statistical domains and along time series, (iii) reveal the weaknesses of the systems, and (iv) set priorities when planning the statistical activities for the coming years.

In "The Power of Habit" by Charles Duhigg, the award-winning business reporter for The New York Times explains why habits exist and how they can be changed. "The process within our brains is a three-step loop. First, there is a *cue*, a trigger that tells your brain to go into automatic mode and which habit to use. Then there is the *routine* (…). Finally, there is a *reward* (…)." When figuring out how to spark a craving, the author argues that it is easier to convince someone to adopt a new behaviour, if the same cue and reward is kept.

To foster new habits, the new routine that comes with the proposed model of quality indicators must be easy to implement, and it must be accepted by all stakeholders as an added value process.

1 CUE

Whenever a new production cycle begins, the cue for statisticians in the Statistics Department of Banco de Portugal is to produce high quality statistics, following the motto of Banco de Portugal's Strategic Plan for 2017–2020: "Always do better".

2 ROUTINE

To pursue this objective (high quality statistics), the Statistics Department of Banco de Portugal created the Statistics Audit Unit (SA Unit) in 2004, with a specialized team responsible for regularly assessing data quality and legal provisions' full compliance. One of the ways to address statistical quality control is carrying out statistics audit operations, with the purpose of evaluating the efficiency of procedures in place, and promoting the sharing of good practices.

Once a year this Unit also produces statistics quality reports (compiler oriented) for the main statistics to assess the quality of the current statistical compilation and, ultimately, to identify opportunities for improvement and future developments towards statistical efficiency.

After all that is being done, what is missing? Quality reports shed light on systems' constraints and include suggestions/recommendations to change and improve current practices. But since they are annually produced, during most part of the year, statistical producers stay focused on presenting results for a specific domain, and a specific time period, following the same specific quality controls. One might thing "if a thing ain't broken, don't fix it" but this should not prevent statisticians to dedicate time to re-think routines and pursue new best practices.

That is why consideration is being given to changing the actual routine and turn these annual quality reports into a regular model of indicators, ready to be quickly and systematically updated by statistics compilers.

3 EXAMPLES OF REGULAR QUALITY INDICATORS

First and foremost, it must be stressed that this model is a preliminary and non-exhaustive proposal of the SA Unit, still to be thoroughly discussed with the production units prior to its implementation.

In this exploratory study, the ongoing model of regular quality indicators consists of:

- Indicators already produced in annual quality reports (or during the production cycle) and new indicators queued for further implementation;
- Indicators computed with a quarterly frequency, but respecting the periodicity of the underlying statistic/phenomenon (i.e. monthly statistics should have monthly indicators, produced with a quarterly frequency);
- Comparable indicators across domains as well as domain specific indicators.

The model should be divided into seven categories:

– Pre-requisites of quality (PR);
– Accuracy and completeness (AC);
– Plausibility and outlier analysis (PO);
– Reliability and revision studies (RR);
– Consistency (C);
– Timeliness and punctuality (TP); and,
– Accessibility (A).

These categories were inspired by DQAF but adjusted to better fit the statistical domains' idiosyncrasies, and produce more intuitive and measurable results. For each category, a brief explanation and a preliminary, non-exhaustive, sample of indicators is therefore presented (symbols classify the actual status of indicator's performance: ● – already exists; o – to be implemented; ●/o – exists but not in a systematic way).

– Pre-requisites of quality: (i) indicators are in place to evaluate the degree to which legal and institutional environment is supportive of statistics and resources are commensurate with statistical programs and used efficiently; (ii) existing statistics are regularly checked to ascertain whether they can be produced in a more cost-effective way or the burden on reporting agents can be reduced.

Table 1 Pre-requisites of quality indicators

PR1. Number of aborted or failed job runs in IT systems (per month)	o
PR2. Percentage of confidential statistical information series flagged (as at the last reviewing date)	o
PR3. Number of days assigned to data exploration stage (per month)	o
PR4. Number of accesses to databases	●

Regarding PR4. a comprehensive list of granted accesses to statistical databases is validated and updated on a yearly basis, as a control activity within the rules on data confidentiality and to ensure the integrity of information.

– Accuracy and completeness: (i) source data and statistical techniques are sound and statistical outputs sufficiently portray reality; (ii) the largest and most material subset of the required information is available.

Table 2 Accuracy and completeness indicators

AC1. Percentage of estimated non-response	●
AC2. Percentage of adjustments/imputation to stocks at the end of period (breakdown by reporting entity, by country,[5] by institutional sector,[6] by unit records)	o
AC3. Percentage of failed 1st level data checks until the version used in the production stage (breakdown by reporting entity)	o

Data checks referred in indicator AC3. are only applicable to statistics with direct report (i.e. monthly MFI[7] and BOP[8] data) and include, for instance, tests on basic logical identities.

– Plausibility and outlier analysis: (i) the absence of unjustified outliers in data; (ii) values that markedly deviate from the usual pattern of the series are detected, isolated and further analysed.

Table 3 Plausibility and outlier analysis indicators

PO1. Monthly rate of change in stocks/transactions/OCVP[9] greater than X% (breakdown by instrument type, by institutional sector debtor and/or creditor, by reporting entity)	●/o
PO2. Year-on-year rate of change in stocks/transactions/OCVP greater than X% (breakdown by instrument type, by institutional sector debtor and/or creditor, by reporting entity)	●/o

Monthly and year-on-year rates are generally computed during the production cycle, when statisticians validate their first estimates. But the systematic documentation, respecting comparable standards, as well as the calibration of thresholds (X%) by phenomenon, and statistical domain, is still to be defined.

– Reliability and revision studies: (i) revised values of statistic are close to the initial value released; (ii) revisions are tracked and mined for the information they may provide.

Table 4 Reliability and revision studies indicators

RR1. MAPE – Mean Absolute Percentage Error	●
RR2. MARE – Mean Absolute Relative Error	●
RR3. Q – Directional Reliability Indicator	●
RR4. RMSRE – Root Mean Square Relative Error	●
RR5. Bias component	●
RR6. Regression component	●
RR7. Disturbance component	●

[5] As defined by ISO 3166-1 country code.
[6] As defined in ESA 2010.
[7] Monetary and financial statistics.
[8] Balance of payments statistics.
[9] Other changes of volume and price.

These indicators are currently computed by SA Unit for annual quality reports purposes and for Key Risk Indicators' (KRI) monitoring.[10] According to this new paradigm, their inclusion in the model of quality indicators represent an opportunity for statistical compilers to regularly, and almost automatically, compute them, anticipating deviations and mitigating risks at source.

The choice of additional items/balance sheet aggregates to be tested is yet to be discussed with each statistical domain.

– Consistency (logical and numerical coherence, including consistency over time, within datasets, across datasets, and comparisons with external data.

Table 5 Consistency indicators

C1:.	First difference of the series between growth rates of change in stocks/transactions/other changes in volume and price (breakdown by instruments, balance sheet items – MFI, or functional categories – BOP)	o
C2:.	Difference between EO[11] series and lower and upper threshold of 3% of current account turnover (only applicable to BOP)	●/o
C3:.	Cross-checks between main balance sheet items in statistical MFI balance sheet information (BSI) and supervisory data in FINREP[12] (only applicable to MFI)	o
C4:.	First difference of the series between goods credits and debits in BOP statistics, and exports and imports in international trade statistics, or between their growth rates (only applicable to BOP)	●/o

In fact, a set of other consistency indicators can be generally defined as "Cn: First difference of the series between prime source (S1) and secondary source (S2)". Whenever this kind of indicator is computed by one statistical domain where S2 becomes S1, the results must be compatible with its "mirror" indicator.

– Timeliness and punctuality (the length of time between its availability and the event it describes; the time lag between the release of data and the target date announced in official release calendar.

Table 6 Timeliness and punctuality indicators

TP1. Punctuality of time schedule of effective publication (in days)	●
TP2. Time lag between the end of reference period and the date of the first/final results (in days)	●

Combining indicator TP2. for timeliness with the number of BP*stat*[13] consultations by statistical domain highlights the relevance of their statistics for users, given the time lag for publication.

– Accessibility (the availability of statistical information to the user, including data and metadata accessibility, and assistance to users.

[10] Revision indicators are based on the report by the joint ECB DG-S/European Commission (Eurostat) Task Force on Quality. See VIOLETTA, D. and AGUILAR, C. P. *Quantitative quality indicators for statistics – an application to euro area balance of payment.* ECB Occasional paper series No. 54, 2006.

[11] Net errors and omissions.

[12] Prudential reporting requirement of financial information enshrined in Implementing Regulation (EU) No. 680/2014, published by the European Banking Authority (EBA).

[13] The "BP*stat*" is a dissemination service of Banco de Portugal that provides statistical information (data and metadata) organised in domains and allowing for both time series and multidimensional exploration.

Table 7 Accessibility indicators

A1. Number of series disseminated (in BP*stat*)	●
A2. Number of statistical press releases disseminated	●
A3. Number of media content related with statistics published on website (videos, infographics, explainers)	o
A4. Total BP*stat* consultations by statistical domain	●
A5. Number of published news by statistical domain	●
A6. Number of requests for information or clarifications answered, by statistical domain	●
A7. Quality assessment punctuation from data users satisfaction surveys	o

BP*stat* will soon be substituted by a dedicated website (a statistics portal) that will allow for greater user-friendliness and interactivity with the users. At that time, indicator A7. should be implemented with a reasonable periodicity.

What are the next steps? Indicators should be exhaustively characterised, tolerance intervals should be defined and main sources must be selected and prepared for regular computation. Comparable indicators, across statistical domains, should be distinguished from those which are specifically related to one domain. In addition, data owners, responsible for updating indicators, must be assigned.

Further possible developments may include the definition of harmonized rules to generically compare the overall quality of each statistical system. The quality assessment exercise should complement automatic results with casuistic analysis.

Finally, the Statistics Department Board, the SA Unit and the staff of statistical domains should reflect together over the evolution of indicators, and contribute to improve the model.

CONCLUSION – REWARD

Like a carrot and stick method, a good quality indicators' model offers several rewards to data users and providers, to the Statistics Department Board, to intermediate managers and staff.

To data users and providers, this model shows that:

- A robust set of indicators can quickly assess users' changing demand for information and help to deliver tailor-made statistics;
- Successful tests to external consistency might reveal new opportunities to integrate and merge information from different sources, hence reducing the reporting burden to data providers.

To the Statistics Department Board, quality indicators represent:

- A valuable management tool to keep track of key performance indicators (KPI) and key risk indicators (KRI);
- A way to raise awareness to their need to intervene and implement new tools and procedures;
- A key management information when prioritising the investment in IT solutions (software, hardware) and in specialised human resources (training).

From the perspective of statistical domains, a new routine will only be embraced if managers and staff believe in the reward. Their benefits can be listed as follows:

- Intermediate managers will have tools to evaluate the quality of their statistical system and the effort put by their teams;
- Statistical domains are encouraged to continuously monitor the quality of their outputs, rather than answering to quality reports' results on an annual basis;

- The staff becomes able to compare their work with their peers;
- In order to anticipate opportunities for improvement, staff might feel encouraged to reroute the calculation of indicators from published data to acquisition/production databases;
- It ultimately makes processes more efficient, leaving the staff with more time to focus on their core business, data analysis and research.

High quality statistical systems provide for more focused and motivated statisticians craving for high quality statistics.

References

AGOSTINHO, A. AND MIGUEL, A. How to increase quality in the Central Banks statistical business process? The experience of Banco de Portugal. *European Conference on Quality in Official Statistics,* Madrid, Spain, 2016 (June).

AGOSTINHO, A. AND MIGUEL, A. Quality indicators and quality measurement to foster and enhance cooperation between users and producers. *European Conference on Quality in Official Statistics*, Vienna, Austria, 2014 (June).

DEFAYS, D. (EUROSTAT) AND LALIBERTÉ, L. (IMF) *Extract of: Comparison of the IMF's Data Quality Assessment Framework (DQAF) and European Statistical System Quality Approaches – An Update"*. 2006.

DUHIGG, C. *The Power of Habit.* 1st Ed. Random House Publishing Group, USA, 2012.

ECB. *ECB Statistics Quality Framework (SQF).* 2008.

ECB. *Public commitment on European Statistics by the ESCB* [online]. [cit. April 2018]. <https://www.ecb.europa.eu/stats/ecb_statistics/governance_and_quality_framework/html/escb_public_commitment_on_european_statistics.en.html>.

EUROSTAT. *ESS handbook for quality reports* [online]. 2014 Ed. [cit. April 2018]. <https://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-GQ-15-003>.

EUROSTAT. *European statistics code of practice (for the national and community statistical authorities* [online]. 2017 revised Ed. [cit. April 2018]. <https://ec.europa.eu/eurostat/en/web/products-catalogues/-/KS-02-18-142>.

EUROSTAT. *Handbook on Data Quality Assessment Methods and Tools* [online]. 2007. [cit. April 2018]. <https://ec.europa.eu/eurostat/documents/64157/4373903/05-Handbook-on-data-quality-assessment-methods-and-tools.pdf/c8bbb146-4d59-4a69-b7c4-218c43952214>.

EUROSTAT. *Standard Quality Indicators – Quality in Statistics (Seventh meeting)* [online]. 2005. [cit. April 2018]. <https://circabc.europa.eu/webdav/CircaBC/ESTAT/prodcom/Library/25-Quality/quality_reports/reference_documents/STANDARD_QUALITY_INDICATORS.pdf>.

INTERNATIONAL MONETARY FUND. *Data Quality Assessment Framework*. 2012.

UN STATISTICS DIVISION. Committee for the Coordination of Statistical Activities *"Principles Governing International Statistical Activities* [online]. 2016. [cit. April 2018]. <https://unstats.un.org/unsd/ccsa/principles_stat_activities>.

VIOLETTA, D. AND AGUILAR, C. P. *Quantitative quality indicators for statistics – an application to euro area balance of payment.* ECB Occasional paper series No. 54, 2006.

Implementation of Quality Management System in the NSI of Serbia – Success Stories and Future Plans

Nataša Cvetković [1] | *Statistical Office of the Republic of Serbia, Belgrade, Serbia*
Miodrag Cerovina [2] | *Statistical Office of the Republic of Serbia, Belgrade, Serbia*

Abstract

During the last several years the Statistical Office of the Republic of Serbia (SORS) made significant efforts in order to introduce quality management system. Information and knowledge about different quality issues are continuously being spread through the intranet portal on quality. User and staff satisfaction surveys have already being implemented for a number of years. Documentation system for quality management, based on ISO 9001 standard, has been introduced, as well as system for producing reference metadata and quality reports (RZSMETA). Standardization of the production process started in the SORS many years ago, in connection with introduction of Integrated Survey Technology (IST). Completely designed and developed in the SORS, IST has become regionally recognized data integration concept supporting several phases of statistical production process (build, collect, process and analyse). It is a challenge now how to efficiently use these comparative advantages in order to build general metadata system as a precondition for further improvement of statistical process and product quality.[3]

Keywords

Quality management, metadata, standardization, IST

JEL code

C19

INTRODUCTION – ORIENTATION TOWARDS QUALITY AND FIRST INITIATIVES

The Statistical Office of the Republic of Serbia (SORS) pursues a long-time strategic goal to produce high quality official statistical results and disseminate them impartially, independently and timely, making them available simultaneously to all users. This comes from the national *Law on Official Statistics*, which specifies the main goal of the official statistics and its basic functioning principles.

[1] Statistical Office of the Republic of Serbia, Milana Rakića 5, Belgrade, Serbia. Corresponding author: e-mail: natasa.cvetkovic@stat.gov.rs, phone: (+381)112412922.
[2] Statistical Office of the Republic of Serbia, Milana Rakića 5, Belgrade, Serbia. Corresponding author: e-mail: miodrag.cerovina@stat.gov.rs, phone: (+381)112412922.
[3] This paper is based on contribution at the *European Conference on Quality in Official Statistics (Q2018)* in June 2018 in Krakow, Poland.

The *mission* of the official statistics in the Republic of Serbia supports our aim to provide relevant, impartial, reliable, timely and internationally comparable statistical indicators. Coordination of authorised producers of official statistics and their active international statistical cooperation should lead to the final result where published official statistical results meet the needs of decision-makers, researchers and other users and provide basis for monitoring and directing of policies in economic and social fields as well as policies concerning the process of accession of the Republic of Serbia to the European Union. Data collection, processing and dissemination rely on inevitable use of methodological and organizational knowledge, statistical standards, modern technologies, protection of statistical confidentiality, optimal use of resources, burden on respondents at a reasonable level and accessibility of data to users under the same conditions.

Reaching a high level of harmonization with international statistical standards, high level of quality of published data and acquired confidence of data providers and users by building professional and infrastructural capacities, adopting and applying the best statistical practices are stated as *vision* of the SORS.

Starting from this strategic orientation the SORS committed itself to implement quality management framework and establish comprehensive quality management system to be able to continuously monitor, evaluate and improve the quality of its statistical products and processes as well as level of overall performance. Moreover, the main aim of the SORS is to become an organisation with quality grounded culture in order to reach excellent level of performance.

More concrete quality-related actions and initiatives in the SORS started in 2011 by adoption of two key strategic documents on quality: *Declaration on Quality*, where European Statistics Code of Practice was put in the heart of quality framework of the SORS, and *Quality Policy.*

At the same time, organisational changes in 2010 took into consideration the need for more intensive work on quality management activities. The *Group for coordination of regional departments, quality management and elections* has been established. Since this organizational unit did not have the capacity to start work on implementation of quality management system, by the end of 2011 the *Working group for Quality management* was formed. It had mandate to act as a coordination body for systematic work on introduction of quality management system according to the SORS' adopted strategic goals and objectives. In order to create synergy of multi-disciplinary work it gathered experienced employees with different background that covered all relevant specialities. This group started a systematic work on introducing the quality management framework, methods and tools, harmonized with the ESS recommendation, with policy of continuous improvement.

Members of the Working group for Quality management have participated in numerous workshops, seminars and study visits and studied and analysed available quality management related documents recommended by EU as well as other countries' best practices. Some activities recommended by the Light Peer Review team (2011) were initiated by this group. Work on quality became more transparent by promoting the SORS quality orientation on the website where main strategic documents on quality could be found, as well as the results of user satisfaction surveys. *Intranet portal for quality management* dedicated to all aspects of quality management was established in 2014. The idea was to maintain and update portal in order to keep informed all employees about the available documentation, results and initiatives of the Working group for quality management.

Enthusiasm and hard work of this group resulted in significant achievements, like establishment of quality management framework, introduction of user and staff satisfaction surveys, development of quality management documentation system, as well as reference metadata and quality reporting system (RZSMETA). By achieving all these results the group was recognised by top management as an important strength of the SORS to set up quality management framework, methods and tools in systematic and harmonised way.

1 MEASURING USER AND STAFF SATISFACTION

Striving to provide high quality products and services to meet users' needs, the SORS have introduced *user satisfaction survey* as a useful tool to identify and analyse users' habits, attitudes and needs, in order to raise the quality of products and services in line with their satisfaction level. The survey had been introduced in 2010, and then was repeatedly conducted in 2013, 2015 and 2017. It is now conducted regularly as a web survey in two-year periodicity, and the results are made public on the SORS website. Apart from this, SORS is also monitoring and analysing comments that users leave through the website survey application *"What do you think about our website?",* social networking channels or through the official e-mail address *stat@stat.gov.rs.* All the information on user satisfaction are used to improve the quality of data and services provided to users in the segments where users are less satisfied. After each user satisfaction survey an action plan is prepared. It includes planned activities regarding, for example, improvement of communication with users, their trainings and increase of statistical literacy, upgrading of the website and dissemination database, providing users with reference metadata, etc.

Likewise, having in mind that staff competence, professionalism and motivation are important elements upon which the quality of statistical products and services depend, SORS introduced *staff satisfaction survey* in order to collect the information on their opinion and attitude regarding various issues. SORS had launched this survey in 2011 for the first time, then conducted it again in 2014 and since then it has been carried out as a regular survey in two-year periodicity. Based on the results of the survey analysis, action plan for improving staff satisfaction is prepared. It is made for the two-year period and includes activities on improving professional development and motivation of employees, internal communication, work environment, etc.

2 QUALITY MANAGEMENT DOCUMENTATION SYSTEM

Determining the need to standardise processes and ensure that (at least) critical processes are performed consistently the same way by all employees, Working group for quality management developed the *Quality management documentation system*. Processes described and presented in documentation have to be uncomplicated, understandable and repeatable, respecting"keep it simple" principle. Implementation of this system started in April 2016. It is based on ISO 9001 standard, and provides standardised and updated documentation (procedures and guidelines) in electronic form to the whole staff. This system is available on the SORS intranet portal, as a set of documents, and it is foreseen to develop an application for dealing with this documentation in the future.

3 REFERENCE METADATA AND QUALITY REPORTING SYSTEM (RZSMETA)

Quality measurement and quality reporting is an integral part of SORS quality policy. Product quality should be regularly assessed and validated in order to provide an input for continuous improvement. Taking this into account, SORS made a significant effort during the last few years in order to develop a reference metadata system – RZSMETA. The main goal was to ensure the production of updated user and producer oriented quality reports that will be used for monitoring product quality. *RZSMETA system* is developed according to Single Integrated Metadata Structure v2.0 (SIMS v2.0) and its underlying reporting structures ESMS 2.0 and ESQRS 2.0, in order to meet EU requirements. The RZSMETA system consists of a repository of metadata, that is a SQLSERVER database, and a web application that enables its users to fill-in their own metadata. RZSMETA application is developed as a user-friendly application with efficient functionalities necessary for producing and disseminating user and producer oriented quality reports. User oriented quality reports (reference metadata) are made available to the public on the official SORS website, and producer oriented quality reports are published on the SORS intranet portal and will be used for monitoring product quality over time. The created files can be automatically downloaded in the SDMX format and sent to Eurostat through ESS Metadata Handler application.

Supporting documentation related to preparation of user and producer oriented quality reports was compiled and it is available on the SORS intranet portal. The Working group for quality management has organized trainings for quality reporting and use of RZSMETA application for all survey statisticians who will be responsible for the preparation of related quality documents.

Deployment of the RZSMETA is considered as SORS high priority in order to provide a tool for posting user oriented quality reports on the website. That will contribute to higher transparency and quality of the statistical activities and raise confidence and trust in the statistics provided by SORS as well as a better understanding of statistics in general.

4 OPERATIONAL METADATA: INTEGRATED SYSTEM OF DATA PROCESSING (IST)

In order to process data of over 300 surveys, the SORS have been using standalone applications that were developed by using different platforms (VB, C++, Java etc.) for many years. The data were stored in different places, from the servers to the local machines, in various formats, from relational databases to text files. In order to avoid this practice and improve delivering high quality statistical products and services at reduced cost, the SORS have developed IST that is consisted of a .net application and a simple metabase (only six tables). These two equally important components deal with a large number of different databases containing individual and aggregated data. IST.net application is an interpreter reading data from the IST metabase that contains descriptions of each application. Based on the read data, the application generates and in real time executes each stage of statistical research or any other project, from data entry to data tabulation for more than 90% of SORS surveys. IST have provided full integration of data and IT standardization as comparative advantages contributing to the introduction of QMS.

5 LINKING EVERYTHING: INTRODUCTION OF GENERAL METADATA (GM)

In parallel with developing of a reference metadata system, it became clear that we were lacking statistical semantics for the most central metadata pillars. This was a potential source of misunderstandings among people working in different subject matter areas or on different types of tasks. Furthermore, this could lead to misunderstandings and decrease of efficiency and quality when exchanging or updating data and metadata in SORS metadata or metadata-driven systems such as RZSMETA system for reference metadata and quality reporting, processing environment (IST), dissemination database etc. Therefore, the SORS have started to build system of *general metadata* which is harmonized, simple and in transparent manner available for reuse and search. Work on codifying and linking of all components (Methods, Instruments, Documentation, etc.) to the Researches (Processes) as a central pillar of GM, is in progress.

General idea is to minimize everyday (interactive) need for maintenance and to maximize automatically updating from the already existing satellite systems like Electronic Library, Statistical Plan, Employees DB and Dissemination DB (machine-to-machine communication). Connection of GM with other satellite systems implies their redesign and use of standardized internal ID codes in order to integrate into centralized statistical information system. Statistical plan represents operationalization of goals of official statistics. As a legal act, the plan clarifies roles and responsibilities of all producers but, at the same time, it is a precious source for general metadata database.

CONCLUSION

Innovative practice of the SORS that could benefit other countries considers two successful stories: the concept and functionalities of the *Integrated System of Data Processing (IST)* and *RZSMETA system for reference metadata and quality reporting*. However, full synergetic effects of two above-mentioned successful stories can be achieved only after their linkage to the General Metadata (GM). Regarding IST, it means mapping GSBPM phases, sub-phases and activities, along with adding a standardized logical layer that consolidates operational metadata according to GSIM in terms of units, variables and classifications

that are used in data processing. It will have twofold benefit: enforcing and reviving GM (ensuring their harmonised use/reuse) and creation of preconditions for various standardized quality reports/indicators (i.e. on correction, validation, transformation) – directly from statistical processing. Regarding RZSMETA, connection to GM means level up in standardized reference metadata descriptions (including standard methodological documents), linking indicators with their constituent variables etc.

A new web-based application solution for compilation of *Statistical plan* (programme) will integrate and coordinate all producers of official statistics within yearly operationalization of statistical goals. Speaking about other authorised producers of official statistics, the idea is to continue work on their education concerning various quality issues and to include them in more concrete quality related activities.

Further focus will be to identify and describe all processes in the statistical production chain, using GSBPM as a tool, and to set mandatory documentation necessary to provide consistent performance of processes. This would lead to higher standardisation of processes, and enable monitoring of their quality in a unique way, including implementation of quality reviews and audits as planned and regular activities.

Identifying and description of processes would also create preconditions for the introduction of a *risk management* system as a system for identification, assessment and prioritization of risks followed by a coordinated and cost-effective use of resources to minimize, monitor and control the probability and/or impact of adverse events. Process analysis and quality assessments are suitable "tools" for identifying risks and control activities.

Management and coordination of all successful stories described in this paper (on-going development programmes, projects and initiatives related to quality, metadata and dissemination) are responsibilities of new department established in the beginning of 2018 by decision of the SORS director general. That decision is remarkably contributing to the further integration of mentioned achievements as ultimate precondition for future development of quality management system. Furthermore, if this organizational change will be followed by joint efforts and understanding of the whole institution, then the SORS strategic and policy goals (towards monitoring, evaluating and improving product and process quality) together with high level of performance, will be achieved.

References

STATISTICAL OFFICE OF THE REPUBLIC OF SERBIA. Official Statistics Law. *Official Journal of the RS*, Belgrade, 2009, No. 104/09, 24/11.

STATISTICAL OFFICE OF THE REPUBLIC OF SERBIA. Strategy of official statistics development in the Republic of Serbia over the period 2016–2020. *Official Journal of the RS*, Belgrade, 2015, No. 55/15.

STATISTICAL OFFICE OF THE REPUBLIC OF SERBIA. *Quality Policy* [online]. Belgrade, 2017. <http://www.stat.gov.rs/media/2485/quality-policy.pdf>.

EUROSTAT. *European Statistics Code of Practice.* 2017.

EUROSTAT. *Single Integrated Metadata Structure v 2.0 (SIMS v2.0) and its underlying reporting structures the ESS Quality and Reference Metadata Reporting Standards ESMS 2.0 and ESQRS 2.0.* December 2015.

EUROSTAT. *ESS handbook for quality reports.* 2014 Ed.

Recent Publications and Events

New publications of the Czech Statistical Office

Agricultural Producer Price Indices 2018. Prague: CZSO, 2019.
Czech Republic in Figures 2018. Prague: CZSO, 2018.
Direct Public Support for Research and Development in the Czech Republic 2017. Prague: CZSO, 2018.
Food Consumption in 2017. Prague: CZSO, 2018.
Foreigners in the Czech Republic 2018. Prague: CZSO, 2018.
Generation, Recovery and Disposal of Waste for the period 2017. Prague: CZSO, 2018.
Indicators of Social and Economic Development of the Czech Republic 2000–3rd quarter 2018. Prague: CZSO, 2018.
Industrial Producer Price Indices 2018. Prague: CZSO, 2019.
Information Economy in Figures 2018. Prague: CZSO, 2018.
Small Lexicon of Municipalities of the Czech Republic 2018. Prague: CZSO, 2018.
Životní podmínky v ČR 2017 (Living Conditions in the CR 2017). Prague: CZSO, 2018.

Other selected publications

Development of the Basic Living Standard Indicators in the Czech Republic 1993–2017. Prague: MoLSA, 2018.
EUROSTAT. *Key figures on Europe.* 2018 Ed. Luxembourg: Publication Office of the European Union, 2018.
EUROSTAT. *Sustainable development in the European Union.* 2018 Ed. Luxembourg: Publication Office of the European Union, 2018.
HENDL, J. AND REMR, J. *Metody výzkumu a evaluace* (Research and Evaluation Methods). Prague: Portal, 2017.

Conferences

The ***62nd ISI World Statistics Congress*** will take place in **Kuala Lumpur, Malaysia, from 18th to 23rd August 2019**. More information available at: <*http://www.isi2019.org*>.

The ***22nd AMSE 2019 Conference*** will be held in **Nižná, Slovakia, from 28th August to 1st September 2019**. The conference is held under the auspices of the President of the Czech Statistical Office and of the President of the Statistical Office of the Slovak Republic and is dedicated to the 100th anniversary of statistics in Czechoslovakia (*https://www.czso.cz/csu/czso/history_of_czech_statistics_after_1918*). More information available at: <*http://www.amse-conference.eu*>.

Papers
We publish articles focused at theoretical and applied statistics, mathematical and statistical methods, conception of official (state) statistics, statistical education, applied economics and econometrics, economic, social and environmental analyses, economic indicators, social and environmental issues in terms of statistics or economics, and regional development issues.

The journal of *Statistika* has the following sections:

The *Analyses* section publishes high quality, complex, and advanced analyses based on the official statistics data focused on economic, environmental, and social spheres. Papers shall have up to 12 000 words or up to twenty (20) 1.5-spaced pages.

Discussion brings the opportunity to openly discuss the current or more general statistical or economic issues, in short what the authors would like to contribute to the scientific debate. Contribution shall have up to 6 000 words or up to 10 1.5-spaced pages.

In the *Methodology* section gives space for the discussion on potential approaches to the statistical description of social, economic, and environmental phenomena, development of indicators, estimation issues, etc. Papers shall have up to 12 000 words or up to twenty (20) 1.5-spaced pages.

Consultation contains papers focused primarily on new perspectives or innovative approaches in statistics or economics about which the authors would like to inform the professional public. Consultation shall have up to 6 000 words or up to 10 1.5-spaced pages.

The *Book Review* section brings reviews of recent books in the fieled of the official statistics. Reviews shall have up to 600 words or one (1) 1.5-spaced page.

The *Information* section includes informative (descriptive) texts, information on latest publications (issued not only by the CZSO), recent and upcoming scientific conferences. Recommended range of information is 6 000 words or up to 10 1.5-spaced pages.

Language
The submission language is English only. Authors are expected to refer to a native language speaker in case they are not sure of language quality of their papers.

Recommended Paper Structure
Title (e.g. On Laconic and Informative Titles) — Authors and Contacts — Abstract (max. 160 words) — Keywords (max. 6 words / phrases) — JEL classification code — Introduction — … — Conclusion — Annex — Acknowledgments — References — Tables and Figures

Authors and Contacts
Rudolf Novak*, Institution Name, Street, City, Country
Jonathan Davis, Institution Name, Street, City, Country
* Corresponding author: e-mail: rudolf.novak@domainname.cz, phone: (+420) 111 222 333

Main Text Format
Times 12 (main text), 1.5 spacing between lines. Page numbers in the lower right-hand corner. *Italics* can be used in the text if necessary. *Do not* use **bold** or underline in the text. Paper parts numbering: 1, 1.1, 1.2, etc.

Headings
1 FIRST-LEVEL HEADING (Times New Roman 12, bold)
1.1 Second-level heading (Times New Roman 12, bold)
1.1.1 Third-level heading (Times New Roman 12, bold italic)

Footnotes
Footnotes should be used sparingly. Do not use endnotes. Do not use footnotes for citing references.

References in the Text
Place reference in the text enclosing authors' names and the year of the reference, e.g. "White (2009) points out that…", "… recent literature (Atkinson et Black, 2010a, 2010b, 2011; Chase et al., 2011, pp. 12–14) conclude…". Note the use of alphabetical order. Include page numbers if appropriate.

List of References
Arrange list of references alphabetically. Use the following reference styles: [for a book] HICKS, J. *Value and Capital: An inquiry into some fundamental principles of economic theory.* 1st Ed. Oxford: Clarendon Press, 1939. [for chapter in an edited book] DASGUPTA, P. et al. Intergenerational Equity, Social Discount Rates and Global Warming. In: PORTNEY, P. AND WEYANT, J., eds. *Discounting and Intergenerational Equity.* Washington, D.C.: Resources for the Future, 1999. [for a journal] HRONOVÁ, S., HINDLS, R., ČABLA, A. Conjunctural Evolution of the Czech Economy. *Statistika: Statistics and Economy Journal,* 2011, 3 (September), pp. 4–17. [for an online source] CZECH COAL. *Annual Report and Financial Statement 2007* [online]. Prague: Czech Coal, 2008. [cit. 20.9.2008]. <http://www.czechcoal.cz/cs/ur/zprava/ur2007cz.pdf>.

Tables
Provide each table on a separate page. Indicate position of the table by placing in the text "insert Table 1 about here". Number tables in the order of appearance Table 1, Table 2, etc. Each table should be titled (e.g. Table 1 Self-explanatory title). Refer to tables using their numbers (e.g. see Table 1, Table A1 in the Annex). Try to break one large table into several smaller tables, whenever possible. Separate thousands with a space (e.g. 1 528 000) and decimal points with a dot (e.g. 1.0). Specify the data source below the tables.

Figures
Figure is any graphical object other than table. Attach each figure as a separate file. Indicate position of the figure by placing in the text "insert Figure 1 about here". Number figures in the order of appearance Figure 1, Figure 2, etc. Each figure should be titled (e.g. Figure 1 Self-explanatory title). Refer to figures using their numbers (e.g. see Figure 1, Figure A1 in the Annex).

Figures should be accompanied by the *.xls, *.xlsx table with the source data. Please provide cartograms in the vector format. Other graphic objects should be provided in *.tif, *.jpg, *.eps formats. Do not supply low-resolution files optimized for the screen use. Specify the source below the figures.

Formulas
Formulas should be prepared in formula editor in the same text format (Times 12) as the main text and numbered.

Paper Submission
Please email your papers in *.doc, *.docx or *.pdf formats to statistika.journal@czso.cz. All papers are subject to double-blind peer review procedure. Articles for the review process are accepted continuously and may contain tables and figures in the text (for final graphical typesetting must be supplied separately as specified in the instructions above). Please be informed about our Publication Ethics rules (i.e. authors responsibilities) published at: <http://www.czso.cz/statistika_journal>.

Managing Editor: Jiří Novotný
phone: (+420) 274 054 299
fax: (+420) 274 052 133
e-mail: statistika.journal@czso.cz
web: www.czso.cz/statistika_journal
address: Czech Statistical Office | Na padesátém 81 | 100 82 Prague 10 | Czech Republic

Subscription price (4 issues yearly)
CZK 372 (incl. postage) for the Czech Republic,
EUR 117 or USD 174 for other countries.
Printed copies can be bought at the Publications Shop of the Czech Statistical Office (CZK 66 per copy).
address: Na padesátém 81 | 100 82 Prague 10 | Czech Republic

Subscriptions and orders
MYRIS TRADE, s. r. o.
P. O. BOX 2 | 142 01 Prague 4 | Czech Republic
phone: (+420) 234 035 200,
fax: (+420) 234 035 207
e-mail: myris@myris.cz

Design: Toman Design
Layout: Ondřej Pazdera
Typesetting: Václav Adam
Print: Czech Statistical Office

99th year of the series of professional statistics and economy journals of the State Statistical Service in the Czech Republic: ***Statistika*** (since 1964), ***Statistika a kontrola*** (1962–1963), ***Statistický obzor*** (1931–1961) and ***Československý statistický věstník*** (1920–1930).

Published by the Czech Statistical Office
ISSN 1804-8765 (Online)
ISSN 0322-788X (Print)
Reg. MK CR E 4684