STATISTIKA

STATISTICS AND ECONOMY JOURNAL

VOL. **49** (1) 2012

CZECH STATISTICAL OFFICE

EDITOR-IN-CHIEF

Stanislava Hronová
Vice-Rector, University of Economics, Prague
Prague, Czech Republic

EDITORIAL BOARD

Iva Ritschelová
President, Czech Statistical Office
Prague, Czech Republic

Marie Bohatá
Vice-President, Eurostat
Luxembourg

Ľudmila Benkovičová
President, Statistical Office of the Slovak Republic
Bratislava, Slovak Republic

Roderich Egeler
President, German Federal Statistical Office
Wiesbaden, Germany

Richard Hindls
Rector, University of Economics, Prague
Prague, Czech Republic

Gejza Dohnal
President of the Czech Statistical Society
Head of the Centre for Quality and Production Reliability,
Czech Technical University in Prague
Prague, Czech Republic

Štěpán Jurajda
Director, CERGE-EI: Center for Economic Research
and Graduate Education — Economics Institute
Prague, Czech Republic

Vladimír Tomšík
Vice-Governor, Czech National Bank
Prague, Czech Republic

Jana Jurečková
Prof., Department of Probability and Mathematical
Statistics, Charles University in Prague
Prague, Czech Republic

Jaromír Antoch
Prof., Department of Probability and Mathematical
Statistics, Charles University in Prague
Prague, Czech Republic

Martin Mandel
Prof., Department of Monetary Theory and Policy,
University of Economics, Prague
Prague, Czech Republic

František Cvengroš
Head of the Macroeconomic Predictions Unit,
Financial Policy Department,
Ministry of Finance of the Czech Republic
Prague, Czech Republic

Josef Plandor
Department of Analysis and Statistics,
Ministry of Industry and Trade of the Czech Republic
Prague, Czech Republic

Petr Zahradník
EU Office, Česká spořitelna, a.s.
Prague, Czech Republic

Kamil Janáček
Board Member, Czech National Bank
Prague, Czech Republic

Petr Vojtíšek
Deputy Director, Monetary and Statistics Department,
Czech National Bank
Prague, Czech Republic

Milan Terek
Prof., Department of Statistics,
University of Economics in Bratislava
Bratislava, Slovak Republic

Cesare Costantino
Research Director,
Italian National Institute of Statistics
Rome, Italy

Walenty Ostasiewicz
Prof., Department of Statistics,
Wroclaw University of Economics
Wroclaw, Poland

ASSOCIATE EDITORS

Jakub Fischer
Vice-Rector, University of Economics, Prague
Prague, Czech Republic

Luboš Marek
Dean of the Faculty of Informatics and Statistics,
University of Economics, Prague
Prague, Czech Republic

Marek Rojíček
Head-Manager of the Macroeconomic Statistics Section,
Czech Statistical Office
Prague, Czech Republic

Hana Řezanková
Prof., Department of Statistics and Probability,
University of Economics, Prague
Prague, Czech Republic

MANAGING EDITOR

Jiří Novotný
Czech Statistical Office
Prague, Czech Republic

FOREWORD

Dear readers,
It is a great pleasure for me to introduce the first 2012 issue of *Statistika, Statistics and Economy Journal*. The Czech version of this journal published by the Czech Statistical Office has a long history starting back in 1964. This issue opens the second year of *Statistika* provided in English version only.

Development of statistical science along with improvement of tools supporting the routine statistical work in the recent years significantly improve the value of statistics as a tool supporting decision making. Synergy of methodological and application-focused papers makes the journal an appropriate platform enabling national statistical and research institutions present and support further progress and development in the field.

The new English language format has opened the new horizons for our periodical. On the one hand, the journal has managed to maintain the majority of the former readers and authors. On the other hand, the new international character of the journal enables the constant acquisition of more and more new readers and authors from all over the world. The electronic open-access version of the journal also greatly contributes to these processes.

This year we plan to continue improving our periodical by extending its reader base as well as attracting new authors. I am also pleased to introduce the new visual style of the journal designed in line with the new corporate identity of the Czech Statistical Office. We also offer an updated version of our website (www.czso.cz/statistika_journal). I believe that these changes will find appreciation and popularity among the readers' community.

I wish the journal a lot of inspired readers and plenty of creative authors. I also hope that the range of articles we offer is useful both for your everyday work and professional growth.

Iva Ritschelová
President of the Czech Statistical Office

CONTENTS

ANALYSES

METHODOLOGIES

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 a professional journal included in the list of peer-reviewed scientific periodicals published in the Czech Republic and rated by the methodology of the Council for Research, Development and Innovation of the Government of the Czech Republic. Since 2011 is Statistika published quarterly in English only.

Publisher

The Czech Statistical Office is an official national statistical institution of the Czech Republic. The Office 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

Ownership Principle in the Foreign Trade Statistics: Czech Approach

Marek Rojíček[1] **|** *Czech Statistical Office, Prague, Czech Republic*
Tereza Košťáková[2] **|** *Czech Statistical Office, Prague, Czech Republic*
Jaroslav Sixta[3] **|** *Czech Statistical Office, Prague, Czech Republic*

Abstract

The Czech Republic is a small open economy, which is vitally dependent on its export performance. In the period after EU accession the intensity of international cooperation grew rapidly in all the Central European countries, which is mostly the result of the huge FDI inflow at the beginning of the decade. In this context one significant problem for the Czech Republic and some other countries of the region appeared: the valuation of the trade flows based on the cross-border measuring overestimates the country's trade balance in comparison with its value added created. This is the case of trade declared by non-resident units, which is more and more common within the European Union. This phenomenon is even enhanced by strategic geographical location of the Czech Republic, which is an important factor explaining why a lot of this "quasi-transit" trade is being operated. The revision of the foreign trade data, which aim is to follow more consistently the ownership approach, significantly changes the picture of the Czech economy, specifically the role of external demand to the economic growth.

Keywords	***JEL code***
Globalisation, foreign trade statistics, balance of payments, quasi-transit, commodity flows	*F10, F15, F23*

INTRODUCTION

The surplus of the balance of trade according to the foreign trade statistics in the Czech Republic has been gradually increasing since joining the EU in 2004. This trend coincided to the effect of the rapid growth of foreign direct investment to manufacturing sector in the preceding years. However, such a development was in contrast to the balance of payment. Moreover, a growing discrepancy has been observed between supply- and use-side of certain commodities during the compilation of the supply and use tables mainly due to exports and imports from the foreign trade statistics. Exports even exceeded production in some of these commodities. So it seemed that exports were overestimated and imports underestimated or both

[1] Na padesátém 81, 100 82 Prague 10, Czech Republic. E-mail: marek.rojicek@czso.cz.
[2] Na padesátém 81, 100 82 Prague 10, Czech Republic. E-mail: tereza.kostakova@czso.cz.
[3] Na padesátém 81, 100 82 Prague 10, Czech Republic. E-mail: jaroslav.sixta@czso.cz.

exports and imports in the foreign trade statistics far exceeded the real economic inputs and outputs of the domestic economy.

The alleged positive balance was actually caused by the value added generated by non-residents and as such it cannot be included in the value added of domestic economy. To this end a new national concept of foreign trade in the Czech Republic was drawn up and corresponding methodology of adjustment of traditional foreign trade data was developed. This phenomenon can be associated with two different but complementary issues. First, an increasing influence of non-residents over the flows of goods across the borders of the Czech Republic and, secondly, an increasing number of movements of goods across the national borders without changing ownership (mainly due to the convenient location of the Czech Republic and sufficient storage facilities that encourage extensive flows of goods across the borders that can be considered only as re-export or quasi-transit trade).

This article describes the separation between foreign trade statistics and change of ownership principle within the EU due to the VAT registered non-residents and introduces the Czech approach to follow the concept of change of ownership related to exports and imports in National Accounts and Balance of Payment.

1 DEFINITION OF FOREIGN TRADE

There are two main approaches to capture commodity transactions in international trade. One is based on the principle of movement of goods across the borders, which is consistent with traditional Foreign Trade Statistics (FTS), the other is based on the change of ownership principle and is consistent with standards on Balance of Payment (BoP) and National Accounts (NA).[4] The cross-border movements used to be considered as an acceptable proxy for the change of ownership. However, globalization in trade[5] led to the separation of these concepts as it extended the variety of transactions when movements of goods are not followed by the change of ownership.

So far, most European countries have considered this separation to be related solely to the trade with non-EU countries (so-called quasi-transit). However, this issue has to be extended also to the trade within the EU as the system of collecting data (Intrastat) instructs not only residents but also non-residents to report their transactions across the borders of domestic economy to its national statistics. This results in inclusion of non-resident transactions in exports and imports of any domestic economy according to the compilation rules of the FTS.

Table 1 Definitions of certain transactions in foreign trade relations

Transactions	Description
Simple transit trade	Transactions in goods which cross the reporting economy on the way to their final destination. They are excluded from the FTS, BoP and NA of the reporting economy.
Re-export	Transactions in goods which are imported into the reporting economy by a resident and then re-exported. Re-exports imply a change in ownership and are included in the FTS, BoP and NA of the reporting economy.
Merchanting	Purchases of goods by a resident of the reporting economy from a non-resident and the subsequent resale of the same goods to another non-resident unless the goods entered the reporting economy.
Quasi-transit trade	Transactions in goods which are imported into the reporting country by a non- resident, and then re-exported to a third country within the same economic union (a variant being the case in which they are imported into the country and, later, sold to a resident there, sometimes at a much higher price, without significant change to the goods and without the involvement of any resident to whom the value added reflecting the increase in price might be attributed).

Source: UNECE, 2010

4 More information can be found in Hronová, Hindls, Fischer, Sixta, 2009. See also IMF, 1993.

5 Discussion about the statistical impacts of globalisation can be found in Fischer, 2007.

The international trade traditionally occurs when delivery of goods from country A to country B is associated with a change of ownership. However, there are also transactions that are associated either solely with movement of goods or only with the change of ownership that has to be treated differently and can have a different impact on macroeconomic statistics (see Table 1). Simple transit trade, quasi-transit trade and re-exports have a common element: in all three cases the domestic supply of goods in the compiling economy is not increased, even if the goods are physically present there. Merchanting is fundamentally different from transit and quasi-transit trade and re-exports, in that the merchanted goods are not physically present in the compiling economy. It is however relevant to this discussion because it is a possible cause of the increase in value of the goods between their import and their export or sale to a final user in the importing country.

2 NON-RESIDENTS' TRANSACTIONS IN INTRASTAT

Intrastat is closely related to the system of value added tax (VAT) in the EU. All VAT registered entities in a country A (above the threshold) are obliged to report their transactions across the borders of the country A to Intrastat in the country A. However, VAT registered entities are not only residents of the country A.

According to the VAT legislation harmonized across the EU, non-resident traders are obliged to register for VAT in any country where they realized any taxable transaction. These taxable transactions include supply of goods (e.g. sales of goods on internal national market or dispatch of goods to other member states and also any transfer of own goods for business purposes across the borders to the country) or the intra-EU acquisition of goods (also any transfer of goods for business purposes across the borders from the country). In all these cases non-resident traders have to register for the VAT and consequently they become respondents to Intrastat in the country where they are not seated and do not have even any physical representation (in tax terminology: "VAT-only").

The reasons behind the business transactions carried out by non-residents are summarized in Table 2. Most of these transactions take place between related companies and the motivation can be of a different nature. There can be also logistical reasons, when the country has a geographically strategic location and serves as an import / export gateway to other countries (mainly countries at the external frontier of the EU, but also Central European countries like the Czech Republic). But it may also involve processing operations and strategy of multinational companies in the distribution market. However, most of these transactions are motivated by the cost reduction and tax optimization.

As for the Czech Republic, two prevailing issues concerning non-resident activities essential for the FTS are recorded. Firstly, there are significant flows of goods imported to the Czech Republic by non-residents that are re-exported without any change of ownership to resident (Figure 1). The core of these transactions is the same as in case of quasi-transit (Table 1) even though they are related mostly to the trade

Table 2 Types of business activities and motivations for transactions carried out by non-resident units

Activities	Motivation
Distribution activities — — rental of warehouses, logistic operations, purchasing, import / export, domestic sales	Logistics
Sales Channels — — "Export / Import Gateway" (e.g. from the West to the East of Europe or vice versa)	Internal / cost reduction Tax benefits
Inward processing — — import / export, purchase processing services at home	Cost reduction
Mediation between residents — — from residents to purchase processing, sale to residents (no imports)	Mastering market / agreements between foreign companies

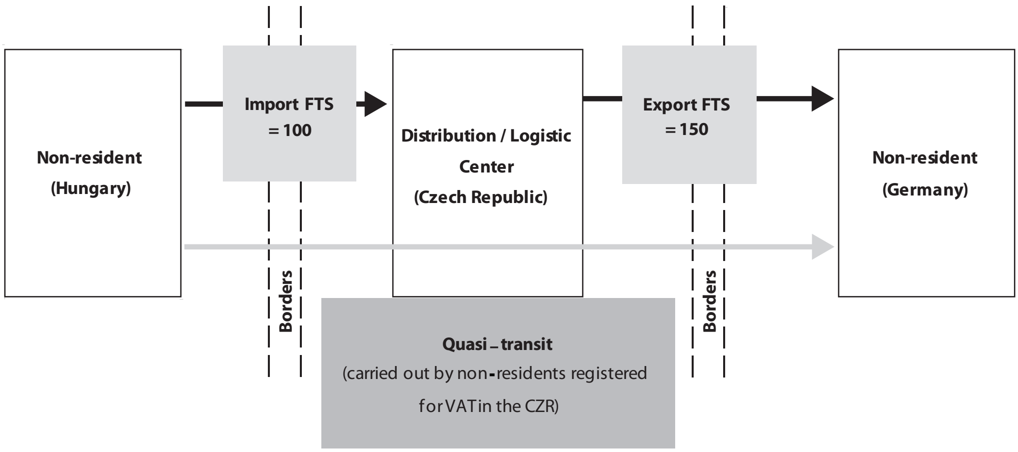
Source: Author's construction

within the EU. As they are not carried out by residents they should not be included in the exports and imports according to the change of ownership principle.

Secondly, there are significant flows of goods across the borders reported by non-residents that are related to their activities on the internal national market: their imports are sold to residents and their exports come from domestic production. Anyway, the value of imports and exports via non-residents reported to the FTS can differ greatly from the value of transactions between them and residents (Figure 2).

In both cases, the balance of exports and imports declared by the FTS is influenced and thus must be adjusted for the value added generated by non-residents if it is to be corresponding to the change of ownership principle.

Figure 1 Illustration of the impact of 'internal' quasi-transit carried out by non-residents on the trade balance



Source: Author's construction

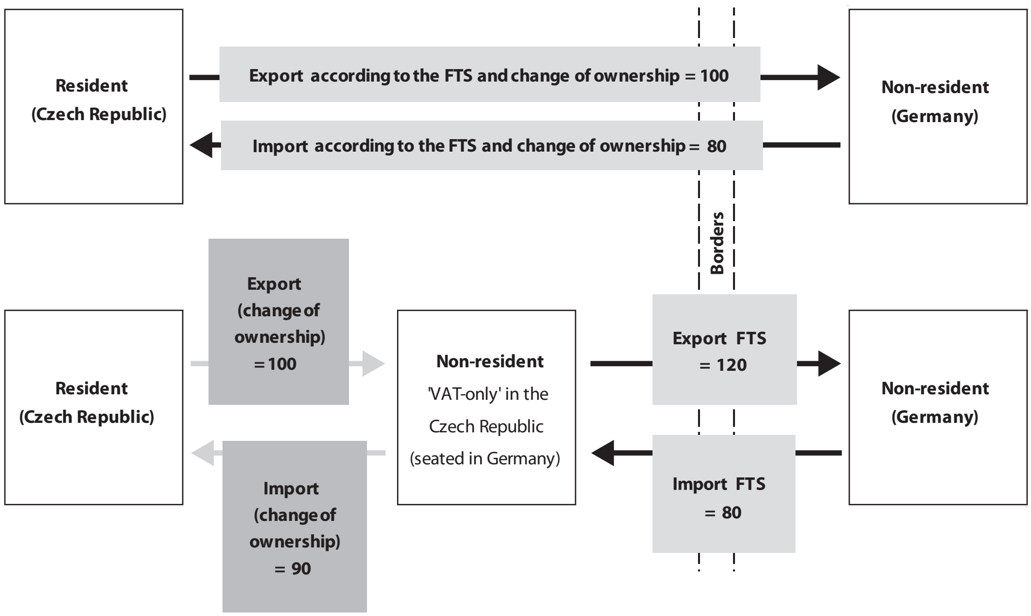
Figure 1: According to the FTS domestic economy imported goods for 100 from Hungary and exported it for 150 to Germany. It seems that the balance of exports and imports of the Czech economy shows surplus (50). Moreover, domestic country shows considerable volumes of imports and exports regardless domestic production or domestic final uses. However, according to the change of ownership there is no import and export because the change of ownership between resident and non-resident did not occur. These transactions should not be recorded as imports and exports in BoP and NA. If the same transaction was carried out by residents of the country, it would be a classical re-export and the value of mediation services (50) will be the value added of domestic traders.

A typical example of quasi-transit is so-called "Rotterdam effect"[6] as described e.g. by the Netherlands or UK (see HM Revenue & Customs, 2005). Increasingly, there can be observed quasi-transit operations

[6] The 'Rotterdam effect' means that a foreign trade transaction is reported for EU statistics first as the imports from a non-EU country to the EU Member State where the goods crossed the EU border and were released to free circulation. This statistical record is part of Extrastat. The following movement of the goods from this EU Member State to the EU Member State which is the final real importing country is then recorded as a dispatch (export) and arrival (import) between these two EU Member States within Intrastat. The 'Rotterdam effect' exists as well for Community exports, but to a lesser extent. The 'Rotterdam effect' inflates the exports and imports of the EU Member States which are exposed to this phenomenon (see HM Revenue & Customs, 2005).

also within European Union (as in the example above). This effect was described by Hungary (see UN-ECE, 2010) and independently this problem has been identified also in the Czech Republic. Unlike the "pure quasi-transit", where the goods do not change its nature in the "transit" economy, the problem of valuation is wider and is related to all cases, where the goods are traded via non-residents (even if the commodities imported are further processed and new products are produced).

Figure 2 The impact of trading carried out by non-residents on the trade balance



Source: Author's construction

Figure 2: 'Direct trade' carried out by residents across the borders can be considered as exports and imports in both cross-border and change of ownership principles (for the first example see Figure 2). The balance of trade shows surplus of 20 which is entirely related to residents' activities (Export = 100 minus Import 80).

However, there is a significant volume of trade in goods carried out 'indirectly' by non-residents across the borders (second example in Figure 2). Unlike the example at Figure 1 the goods traded by non-residents become either final use (in case of imports) or come from domestic production (in case of exports). Non-resident reports to the FTS export of 120 and import of 80 even though the change in ownership between resident and non-resident occurred within the borders for significantly different price (purchase by non-resident for 100 and sale by non-resident for 90).

The balance of trade of the Czech Republic according to the FTS shows alleged surplus of 60 (40 plus the surplus from the direct trade by residents for 20). However, according to the change of ownership principle the balance of trade of the Czech economy amount to 30 (10 for purchase minus sale by non-resident on internal market plus 20 for direct trade by residents).

The surplus according to the FTS includes also the value added generated by non-residents and thus for the purpose of BoP and NA it should be excluded from the value added of the Czech economy. Simultaneously, the value added achieved by residents trading with non-residents on the internal market should be included.

The impact of both examples (shown in Figures 2 and 3) on the trade balance in the FTS and the volume of trade in the FTS depends on its share of the transactions carried out and reported by non-residents in the domestic economy.

3 IMPACT OF NON-RESIDENTS' TRADING ON STATISTICS

Generally, there is serious effect of the trading via non-residents on the consistency between supply and use side in the economy. For some commodity groups exports exceed the production or the imports exceed domestic uses. In this case commodity balancing process within supply- and use-tables is very difficult as the data sources are considerably inconsistent (see Eurostat, 2002).

Another problem arises regarding consistency of the current and financial account balance. The balance of payments is based on the monitoring of transactions between resident and non-resident entities, both in real terms (current account) and financial transactions (financial account). As for the trade carried out by residents the balance of real transactions (foreign trade) will be reflected in financial transactions, namely the balance of receivables and liabilities to non-residents. If the balance of foreign trade is carried out by non-resident units, residents' financial claims on non-residents do not arise and there is a disproportion between the current and financial account balance.

Consider the following very common situation where a Czech company (resident) sells to its parent company goods at a fixed price. The parent company (registered for VAT only in the CR) then exports the goods and reports to statisticians an entirely different value (usually higher) at which the goods are sold on Western markets. At first sight it seems that the Czech economy gains high export prices, but subsidiary (resident) has significantly lower yields. At macro level there is a disproportion between the current and financial account balance, the (value of) movement of goods is higher than money transfers.

After the EU accession in 2004 the system of foreign trade statistics based on customs declarations was replaced for the transactions within the EU by the Intrastat. The structure of data and rules for their declarations are consistent with international manuals of merchandise statistics (IMTS) and are strictly regulated by EU Regulations (data reported to Eurostat). It is nonetheless allowed to adjust data according to national specifics (called 'national concept'). One of them is "quasi-transit" trade, which was generally considered to be the problem related to the trade between non-EU and EU countries at the external EU border (above mentioned "Rotterdam effect").

The first time when the problems with inconsistency of macroeconomic aggregates in the Czech economy appeared was during the balancing process of commodity flows for year 2007, carried out in 2009. Export of certain commodities many times exceeded their domestic production (see Table 3). This can be described by the following model example (names of the companies and data are fictional):

The company of "Global Toys", registered in the Great Britain, is the owner of the Czech toy producer "Czech Toys". This manufacturer produces toys for CZK 5 million and exports them (to the EU countries) through its parent company, which due to this transaction had to registered for VAT in the Czech Republic. Simultaneously, this parent company imports toys from Poland (at the value of CZK 7 million), which are only packed in the CR and then forwarded to the markets in the EU. The overall sales value of the toys exported from the Czech Republic accounts for CZK 16 million.

Company "Global Toys", VAT-only in the CR, reports imports of toys at the value of CZK 7 million to Intrastat. At the same time, it declares "dispatch of goods to other Member State" (export) at the amount of CZK 16 million in Intrastat. In its VAT tax form the company states "received taxable transactions of goods in the CR" at the amount of CZK 5 million (purchased from the company of "Czech Toys"). Therefore value added generated by this non-resident is equal to CZK 16 – 7 – 5 = 4 million (export minus import minus purchase in the CR). The balance of trade according to the cross-border FTS shows the surplus of CZK 9 million. However, 4 million of the surplus belongs to non-resident.

Table 3 Difference between exports and output in 2007 in the Czech Republic (CZK mil.)

CPA		Export FTS[1)]	Output[2)]	Import for inward processing	Difference (Output[2)] — Export[1)])	Ration Export[1)]— Output[2)]
Total		**2 479 234**	**7 446 771**	**117 244**	**–159 695**	
	Including					
182	Other wearing	21 764	18 480	1 361	**–3 284**	**1.18**
193	Footwear	7 484	3 473	348	**–4 011**	**2.15**
245	Glycerol; soap	21 175	21 163	451	**–12**	**1.00**
246	Other chemical prod.	16 615	13 216	323	**–3 399**	**1.26**
274	Basic metals	29 708	23 332	2 858	**–6 376**	**1.27**
300	Office machinery	188 461	128 107	714	**–60 354**	**1.47**
321	Electronic valves	36 205	31 838	3 376	**–4 367**	**1.14**
322	TV and ratio transmitters	41 605	20 647	59	**–20 958**	**2.02**
323	TV and ratio receivers	91 395	76 620	511	**–14 775**	**1.19**
365	Games and toys	22 749	9 876	1 579	**–12 873**	**2.30**

1) Export without Import of goods for Inward processing, 2) Output before compilation of supply–use tables.
Source: Czech Statistical Office

This problem began to be evident also on the quarterly national accounts and balance of payments data in 2009, because of the sharp increase in the year on year surplus in trade balance, without corresponding development in domestic value added and foreign claims. This imbalance, and a solution suggested by the Czech Statistical Office was reported along with the publication of GDP data in March 2010.

In the next twelve months in close cooperation between the Czech Statistical Office and the Czech National Bank a new approach to the foreign trade transactions has been developed called 'national concept'. It follows the change of ownership principle and allows more realistic look at the transactions with the rest of the world and the structure of the Czech economy (see Rojíček, Košťáková, Sixta, 2010 and 2011).

4 POSSIBLE SOLUTIONS OF ADJUSTMENT OF FTS

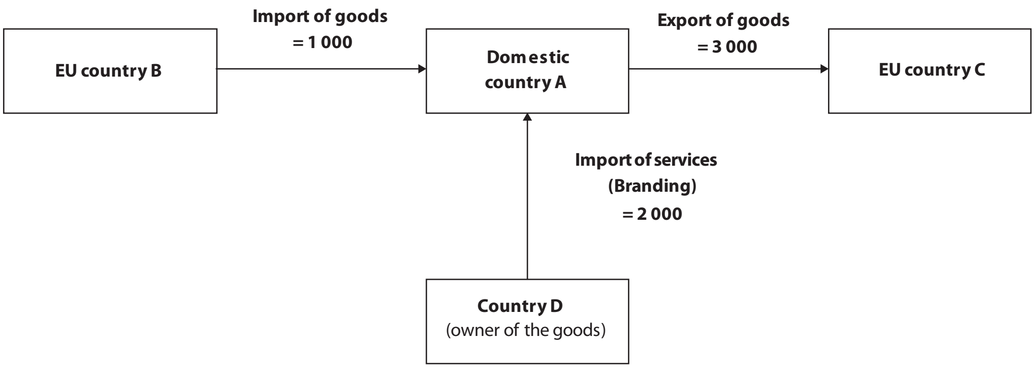
There are two possible ways to solve capturing the inconsistency between the FTS and change of ownership principle within current statistical system: either to impute the difference to the import of services (item called 'branding',[7] see Figure 3) or to adjust data on trade in goods to follow the change of ownership principle. The former solution was applied temporarily to the Czech National Accounts and Balance of Payment at the beginning, when the range of inconsistency had not been thoroughly analyzed.

The EU prefers the above-mentioned approach (branding), which ensures the consistency with the community concept of the FTS (FTS data remain unchanged), and the value adjustment is made in the balance of services. However, when the difference is caused not only by selling of imported goods in an internal market or exporting of purchased goods by non-residents but also by goods merely imported-exported by non-residents through the territory of a member state without a change of ownership to resident methodically more correct and for analytical purposes preferable would be the adjustment of the FTS data. Moreover, when the difference shows a significant impact on the balance of trade and

7 Item reflects price differences in the turnover of foreign trade caused by internal cross-border transactions of multinational companies registered as VAT payer in the exporting country (see CNB, 2010).

the value of goods traded as it has in the Czech Republic, the imputation of the difference to the services would deform the whole picture of foreign trade in services. For all these reasons the adjustment of data on trade in goods is preferred by the CZSO and CNB.[8]

Figure 3 Branding and quasi-transit trade



Source: Author's construction

Figure 3: As for the country D where non-resident is seated, the transaction is captured as 'merchanting' (as an export of trading services).

In March 2011, the CZSO published data on foreign trade for the years 2009 and 2010 according to the national concept for the first time. Data on exports and imports according to the FTS is from now on labelled as "cross-border statistics". Since that data on foreign trade in national concept became an integrated part of monthly issued press releases alongside the cross-border statistics data. During the year 2011 foreign trade in national concept replaced formerly used FTS data in the National Accounts and the Balance of Payment in the Czech Republic.

As there was a parallel revision of the trade in services within the revision on National Accounts in the year 2011 (revision of years 1995–2009) more than half of the impact on the current account balance was offset. The change in trade in services consisted mainly of increasing the so-called direct trade costs associated with import and export of goods and removing the "branding" item from the balance of services (as the phenomenon was now treated in goods, not service balance). The total negative impact on the BoP current account balance was about 1.7% of GDP.

So far the national concept can provide data on export, import and the balance of trade yet with some breakdown limitations. This results from the nature of the methodology, because data are first calculated at the macro level and the structure is modelled using cross-border statistics. The largest relative differences between national concept and cross-borders statistics data occur in computers, electrical equipment and other machines, which is also the most involved in global production chains.

5 NATIONAL CONCEPT OF THE FOREIGN TRADE IN THE CZECH REPUBLIC — METHODOLOGY

The adjustment of FTS-exports and imports according to the national concept is divided into two stages. At the Stage 1 — balance of foreign trade in national concept is estimated (regarding the data of non-

[8] This approach is preferred also by Belgium, where FTS data for non-residents are adjusted using information from VAT files (see NBB, 2009–2011).

residents in FTS and using VAT declarations). At Stage 2, assuming that the balance from Stage 1 remains unchanged, the total value of exports and imports is estimated, partly according to the adjusted exports and imports from Stage 1, partly (in commodities CPA 26, CPA 27, CPA 28) on the basis of the production statistics.

Stage 1

The aim of this stage is to estimate the balance of trade in national concept, in other words, to adjust the balance of trade of the FTS in relation to the change of ownership concept. The commodity balances are also estimated.

There are adjusted only exports and imports declared by non-residents at Stage 1. Transactions reported by residents are not a subject matter of the adjustment. Non-residents in the FTS — Intrastat (trade within EU) are distinguished by their specific Tax ID number (beginning "CZ68" with nine digits). To be identified as a non-resident unit they have to meet also other necessary requirements: 1) do not have Czech ID number of economic unit, 2) do not have any affiliate in the Czech Republic and 3) do not pay income tax here. Non-residents in the FTS — Extrastat (trade with countries outside the EU) are distinguished by their specific EORI[9] number which is unique for each entity within the EU, however, can take a various shape.

In general, the total value of exports (and imports) of non-residents is replaced by the total value of purchases (or sales) of non-residents in the Czech Republic according to their VAT declarations, which are identified as well as in FTS — Intrastat (specific VAT number).

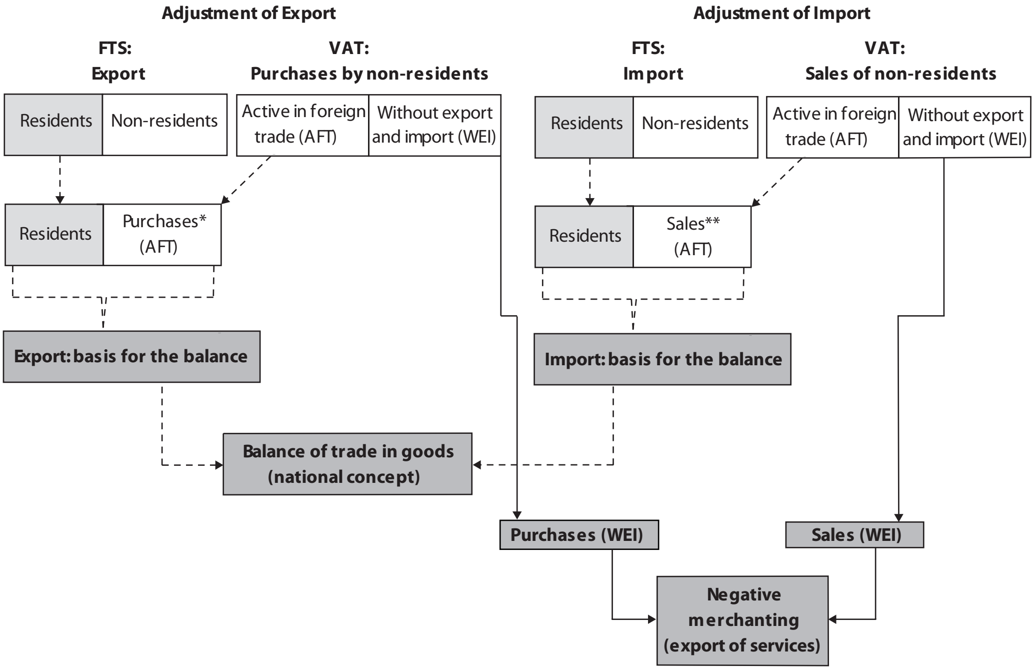
Unfortunately, as each unit identifies itself differently in each data source (Intrastat and VAT-declarations vs. Extrastat) so far it has not been possible to interlink each non-resident individually in all data sources. As a consequence, the computation of foreign trade in national concept is computed in total (for all non-residents together) instead of approaching each non-resident individually (see Figure 5).

However, there is one exception of inclusion of all non-residents' transactions from the VAT declarations to the adjustment of trade in goods. The sales and purchases of those non-residents that do not carry out any (or almost any) export and import according to their VAT declarations (see Figure 5, Purchases and Sales — WEI) are taken aside and the difference between their sales and their purchases in the Czech Republic is considered as import of intermediation services and is therefore added negatively to export of services (as it is similar to the concept of merchanting).

As for the computation of import in national concept, the value of goods flowing into the Czech Republic across the borders declared by non-residents (imports according to the FTS) is substituted by the value of sales in the Czech Republic by non-residents that take part in foreign trade (these sales are imports according to the national concept as a change of ownership from non-resident to resident occurs). The value of domestic sales is based on realized taxable supplies by non-residents with a place of supply in the Czech Republic. These sales in VAT declarations, however, could include not only goods but also some services provided by non-residents. However, it is impossible to identify these services directly in VAT declarations so they are estimated and excluded subsequently. The estimation of the services provided by non-residents in the Czech Republic that can be declared in their VAT statements is based on the statistical survey of import and export of services held by the Czech Statistical Office (ZO 1-04). These services are related mainly to real estate in the Czech Republic or to cultural, sporting or educational gatherings. The impact of the adjustment for the services is about 1% of the total value of sales.

[9] EORI = Economic Operator Registration and Identification.

Figure 5 Estimation of the trade balance in national concept (Stage 1)



* Purchases including inward proccessing services ordered by non-residents registered for VAT.
** Sales after exclusion of services carried out by non-residents registered for VAT.

Source: Author's construction

As for the computation of export in national concept, the value of goods flowing out of the Czech Republic across the borders declared by non-residents (exports according to the FTS) is substituted by the value of purchases of non-residents in the Czech Republic that take part in foreign trade (the purchases are exports according to the national concept as a change of ownership from resident to non-resident occurs). The value of domestic purchases is based on received taxable supplies by non-residents with a place of supply in the Czech Republic. The value of domestic purchases of non-residents does not include the goods sent for inward processing by VAT registered non-residents in the Czech Republic (and ordered services). In order to follow the methodology that requires inclusion of these transactions into the foreign trade aggregates (ESA95), the difference between goods exported after inward processing (declared by non-residents in the FTS) and goods imported for inward processing in the Czech Republic (declared by non-residents in the FTS) is added to the value of purchases of non-residents. The impact of the adjustment of the goods sent for inward processing by VAT-registered non-residents is less than 0.3% of the total value of purchases.

The commodity breakdown of adjusted imports of non-residents that are the basis for the estimation of commodity balances is identical to the commodity breakdown of imports of non-residents according to the FTS. The commodity breakdown of adjusted exports of non-residents that are the basis for the estimation of commodity balances is slightly different from the breakdown according to the FTS due to the commodity balance of the inward processing ordered by VAT-registered non-residents.

The adjusted exports and imports of non-residents (in fact, exports and imports of residents carried out across the borders by non-residents) are added to the exports and imports of residents according to the FTS and these aggregates are the basis for the estimation of the total balance and commodity balances of foreign trade in national concept.

Stage 2
Regarding the long-term observed inconsistency between the value of exports and the output of certain commodities due to quasi-transit the estimation of the total value of transactions between residents and non-residents according to the national concept in the Czech Republic is as important as the estimation of the balance and must be made in relation to the output performance of the domestic economy. As the inconsistency was observed at the commodity level it is necessary to make the estimation also at the commodity level, especially for those commodities that are most influenced by non-resident transactions across the borders (computers, electronic devices, its parts, other machines etc). The correspondence between the output and estimated export is provided by the balancing process of supply- and use-tables.

For the commodity groups CPA 26, CPA 27 and CPA 28 the estimation is based on the residents' production and the share of direct and indirect export in the domestic production. Additionally, also the import for inward processing in these commodities must be added to such estimated export in order to obtain the total value of exports of the commodities (as it is also imputed to the output). The total value of imports in these commodity groups are subsequently computed according to the total value of exports provided unchanged balances of these commodities (obtained at Stage 1). In other words, the adjustment is done on both sides equally so the total balance of trade and commodity balances remain unchanged (from Stage 1).

This calculation is processed at the 2-digit CPA level because data at more detailed CPA levels shows significant inconsistency between classification used in the FTS (KN8), production and industry statistics (PRODCOM).

The difference between the value of exports according to the national concept (based on output performance) and the value of exports according to the movements across the borders (FTS) in these commodity groups have increased significantly in recent years (see Table 4). This indicates the growing separation of transactions according to the FTS and the real output performance of the domestic economy.

Table 4 Ratio between the value of exports in national concept and exports in FTS in commodity groups CPA26, CPA27, CPA28

	2004	2005	2006	2007	2008	2009	2010
CPA26	0.86	0.84	0.73	0.65	0.60	0.66	0.59
CPA27	0.80	0.78	0.78	0.72	0.70	0.76	0.70
CPA28	0.72	0.71	0.71	0.64	0.64	0.68	0.63

Source: Czech Statistical Office

The ratio shown in Table 4 is used for the estimation of exports in the year following the balancing of supply and use tables. That means that the ratio computed during the balancing process of preliminary supply and use tables for year T is used for monthly computed exports in year T+1 and T+2 until the balancing of preliminary supply and use for year T+1 occurs.

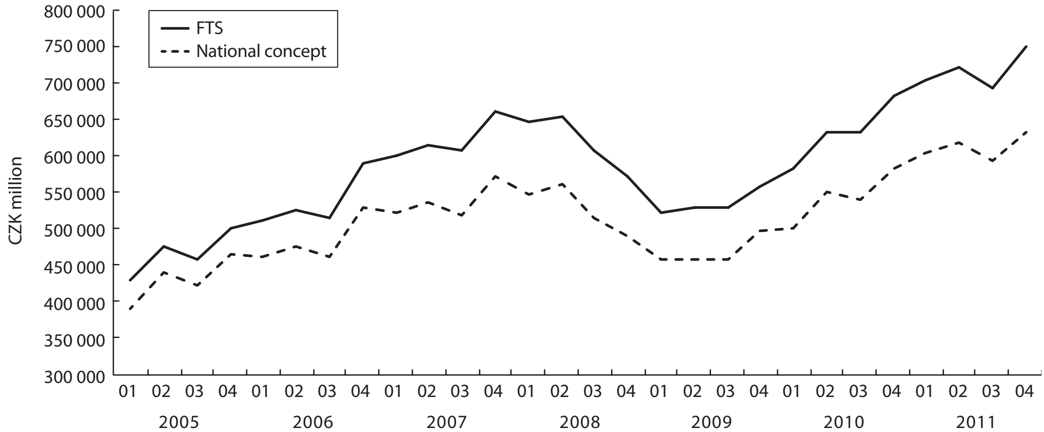
The total value of export and import of commodities other than CPA 26–28 are estimated at Stage 1 (adjusted value by sales and purchases in internal market) as these commodity groups do not indicates significant imbalance caused by quasi-transit through the storage and logistics centres in the Czech Republic. The total value of exports (imports) in national concept is given as a sum of all commodity exports (imports).

6 IMPACT OF ADJUSTMENT OF FTS TO NATIONAL CONCEPT IN THE CZECH REPUBLIC

The difference between the two methodologies has been increasing since 2005 when data is available. In the year 2010 the difference amounted to 14.1% of the exports of goods and 9.0% on the import side (see Figures 6 and 7). The impact on the balance was CZK –142 billion resulting in balance of CZK –21 billion according to the national concept (instead of surplus CZK 121 billion in the FTS) — see Figure 7.

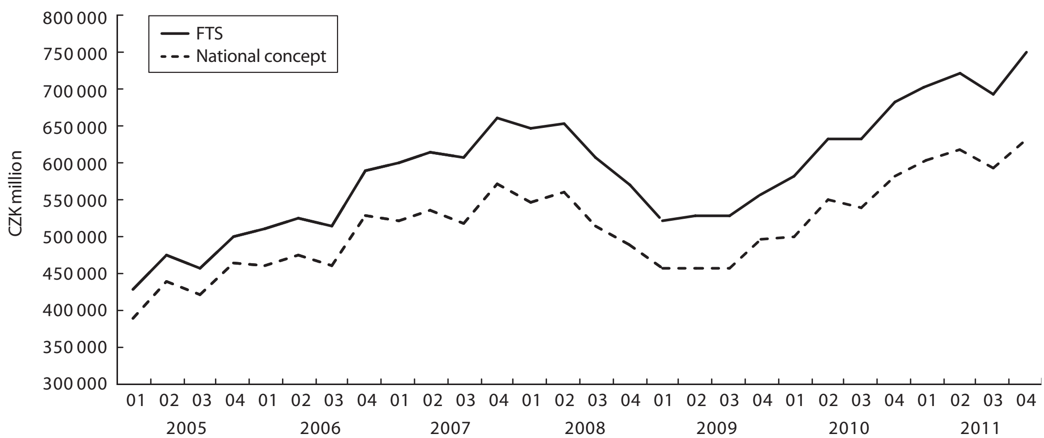
In 2011, the relative adjustment in exports and imports was alike, however, the adjustment of the balance increased to CZK –174 billion which was more than 90% of the surplus according to the FTS (CZK 192 billion).

Figure 6 Export in national concept and FTS in the Czech Republic (FOB)



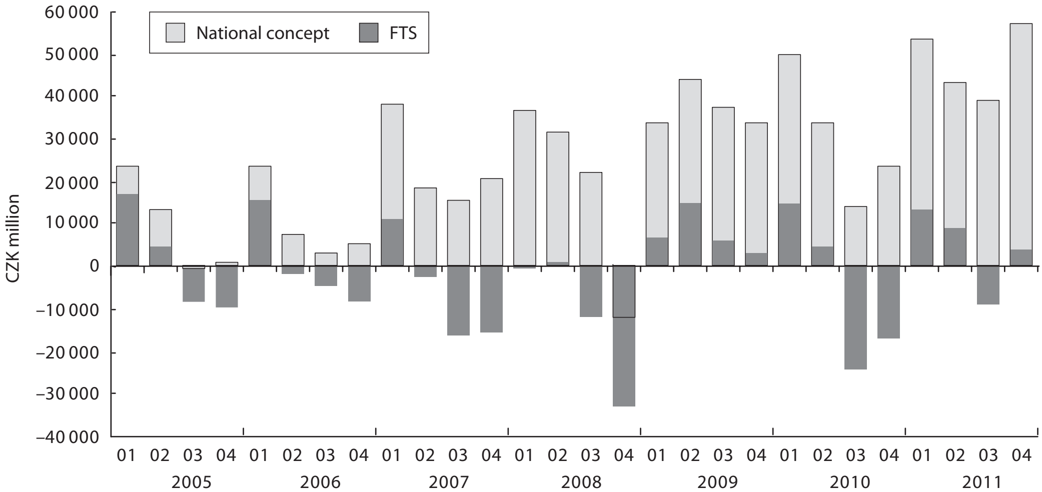
Source: Czech Statistical Office

Figure 7 Import in national concept and FTS in the Czech Republic (CIF)



Source: Czech Statistical Office

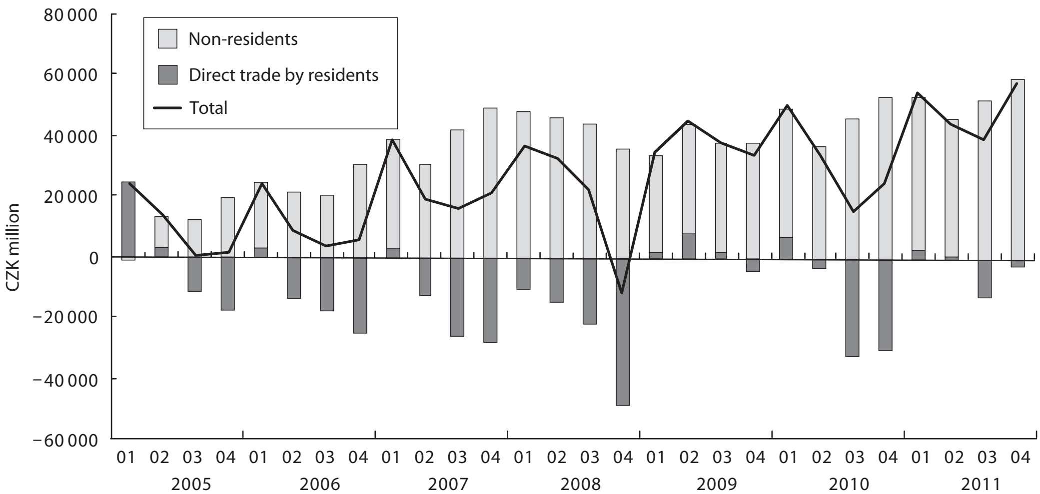
Figure 8 Balance of trade in national concept and FTS in the Czech Republic (CIF / FOB)



Source: Czech Statistical Office

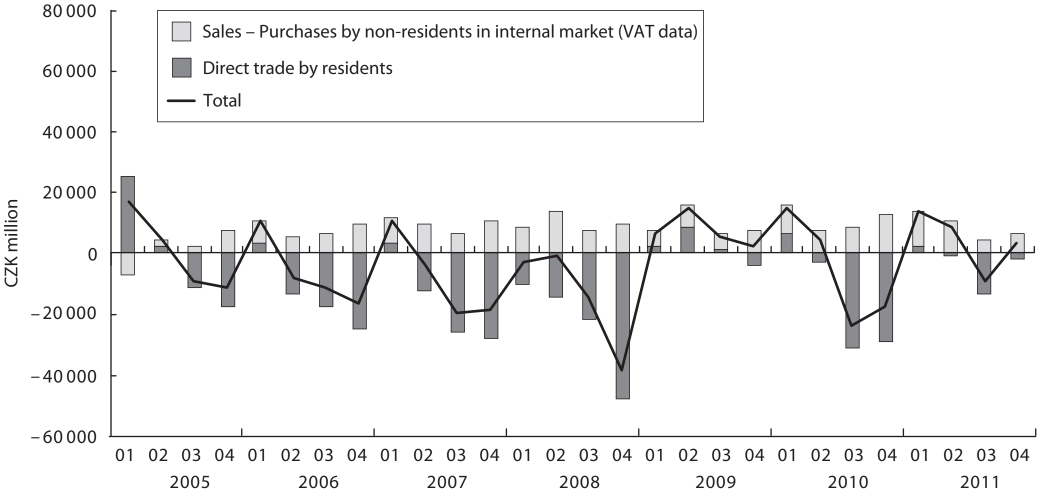
Although the contribution of non-residents to the total exports was 23% in 2011 (19% on imports), in fact they are creating the whole trade surplus (see Figures 8 and 9). On the other hand, resident's trading resulted in deficit (with exception of the year 2009, when the oil prices sharply decreased).

Figure 9 Balance of trade in FTS: non-residents and direct trade by residents (FOB / CIF)



Source: Czech Statistical Office

Figure 10 Balance of trade in national concept: non-residents' transactions in internal trade and direct trade by residents (FOB / CIF)



Source: Czech Statistical Office

Table A3 (see the Annex) illustrates that almost two thirds of the difference of trade balance between cross-border statistics and national concept are in commodity groups CPA 26–28. These are the commodities that represent most of the trade across the borders by non-residents in the Czech Republic and are influenced in the FTS greatly by quasi-transit transactions related to the storage facilities in the country. The differences in all other commodity groups are related only to the transactions in the internal market. Commodities that are not traded by non-residents across the borders are not adjusted (e.g. coal, crude petroleum and natural gas).

CONCLUSION

The Czech Republic is a small open economy, which is vitally dependent on its export performance. In the period after EU accession the intensity of international cooperation grew rapidly in all the Central European countries, which is mostly the result of the huge FDI inflow at the beginning of the decade. In this context one significant problem for the Czech Republic and some other countries of the region appeared: the valuation of the trade flows based on the cross-border measuring overestimates the country's trade balance in comparison with its value added created. This is the case of trade declared by non-resident units, which is more and more common within the European Union. This phenomenon is even enhanced by the strategic geographical location of the Czech Republic, which is an important factor why a lot of this "quasi-transit" trade is being operated. The overvaluation of the trade balance is concentrated in several commodity groups, among them especially computers and electric equipment are significant. The revision of the foreign trade data, whose aim was to follow more consistently the ownership approach, significantly changed the picture of the Czech economy, specifically the role of external demand to the economic growth. It had also an impact on the structure of the input-output tables, especially the division of the domestic and foreign part of the supply and use matrices.

The national concept of foreign trade based on the change of ownership principle is consistent with the methodology of Balance of Payments and National accounting. While in the global context most attention is devoted to the problem of recording "processing" operations, for countries within the EU the problem of quasi-transit trade and the role of non-resident units seems to be very topical. We expect that in the next years this issue has to be seriously discussed. Supply and use tables serve as an important tool in this process. The next efforts will focus on the improvements of linkage between Custom declarations and Intrastat and, moreover, the knowledge of connection between resident enterprises and VAT-registered non-residents. One of the tools for improving quality and detail of foreign trade data is comparison to industrial statistics (surveys on production and direct and indirect exports).

References

CNB. *Balance of Payments Report 2010* [online]. Prague: Czech National Bank, 2010. <http://www.cnb.cz/miranda2/ /export/sites/www.cnb.cz/en/statistics/bop_stat/bop_publications/bop_reports/bop_2010_en.pdf>.

EUROSTAT. *The ESA 95 Input-Output Manual.* Luxembourg: Eurostat, 2002.

FISCHER, J.: *Globalization (its Impact on Statistical Measurements).* 93rd DGINS Conference Discussion Paper, Budapest, 2007.

HM Revenue & Customs. *Analysis of Asymmetries in Intra-Community Trade Statistics with Particular Regard to the Impact of the Rotterdam and Antwerp Effects.* Edicom Report. HM Revenue & Customs, December 2005.

HRONOVÁ, S., FISCHER, J., HINDLS, R., SIXTA, J.: N*árodní účetnictví. Nástroj popisu globální ekonomiky* (National Accounts. Tool for Describtion of the Global Economy). Prague: C. H. Beck, 2009.

IMF. *Balance of Payments Manual,* 5th edition. IMF: Washington, 1993.

NBB: *Statistique du Commerce Extérieur Bulletin Mensuel* (Foreign Trade Statistics Monthly Bulletin) [online]. Bank Nationale de Belgique, 2009–2011. <http://www.nbb.be>.

OECD: *Measuring Globalization: Handbook on Economic Globalization Indicators.* Paris: OECD, 2005. ISBN 92-64-10808-4.

OECD: *Staying Competitive in the Global Economy. Compendium of Studies on Global Value Chains.* Paris: OECD, 2008.

ROJÍČEK, M.: *Globalisation Effects on the Trade Flows: Czech Experience.* 19th International Input-Output Conference, Egypt, Alexandria, 2011.

ROJÍČEK, M.: *Globalizační aspekty zahraničního obchodu v ČR* (Globalization Aspects of the Foreign Trade in the CR). Working Paper, CVKS, 2011.

ROJÍČEK, M., KOŠŤÁKOVÁ, T., SIXTA, J.: *Zahraniční obchod v pojetí platební bilance a národních účtů. Metodické změny v roce 2011* (Foreign Trade in the Concept of the Balance of Payments and National Accounts. Methodological Changes in 2011). Presentation, Workshop at the Czech Statistical Office and the Czech National Bank, Prague, 18.10.2010.

ROJÍČEK, M., KOŠŤÁKOVÁ, T., SIXTA, J.: *Zahraniční obchod v národním pojetí. Metodické změny a údaje za roky 2009–2010* (Foreign Trade in the National Concept. Methodological Changes and Data for the Years 2009–2010). Presentation, Workshop at the Czech Statistical Office, Prague, 7.3.2011.

SIXTA, J., FISCHER, J.: Perspectives of the Measurement of Growth. Poprad, Slovakia, 30.8.2007–31.8.2007. In *AMSE 2007* [CD-ROM]. Banská Bystrica: Universita Mateje Bela, 2007, pp. 190–193.

UNECE: *Impact of Globalisation on National Accounts: Practical Guidance (draft).* UNECE, 2010.

ANNEX

Table A1 Commodity structure of export of goods in FTS (residents and non-residents) and national concept in 2010 (CZK million)

Code	Commodity group	Cross-border statistics					National concept	
		Total	Including		Commodity stucture (%)		Adjust-ment	Total
			Residents	Non-residents	Residents	Non-residents		
Total		**2 532 797**	**1 980 347**	**552 449**	**100**	**100**	**–357 954**	**2 174 842**
	Including							
01	Products of agriculture	**23 382**	22 775	607	1.2	0.1	–48	**23 335**
05	Coal and lignite	**20 889**	20 889	0	1.1	0.0	0	**20 889**
06	Crude petroleum, natural gas	**12 929**	12 929	0	0.7	0.0	0	**12 929**
10	Food products	**62 278**	59 096	3 182	3.0	0.6	–256	**62 024**
12	Tobacco products	**7 762**	3 389	4 373	0.2	0.8	–321	**7 441**
13	Textiles	**42 655**	36 456	6 199	1.8	1.1	–468	**42 187**
14	Wearing apparel	**24 397**	13 629	10 768	0.7	1.9	–934	**23 463**
15	Leather and related products	**12 991**	7 110	5 881	0.4	1.1	–524	**12 467**
17	Paper and paper products	**39 465**	33 213	6 252	1.7	1.1	–727	**38 739**
19	Coke and refined petroleum	**30 107**	29 909	198	1.5	0.0	–16	**30 090**
20	Chemicals, chemical products	**117 177**	96 213	20 964	4.9	3.8	–2 120	**115 059**
21	Basic pharmaceutical products	**32 889**	25 798	7 091	1.3	1.3	–550	**32 340**
22	Rubber and plastics products	**118 430**	107 502	10 928	5.4	2.0	–769	**117 661**
23	Other non-metallic mineral pr.	**54 401**	49 779	4 622	2.5	0.8	–339	**54 060**
24	Basic metals	**113 876**	108 809	5 067	5.5	0.9	–458	**113 419**
25	Fabricated metal products	**139 355**	130 148	9 207	6.6	1.7	–1 754	**137 603**
26	Computer, electronic, optical pr.	**427 915**	165 667	262 248	8.4	47.5	–173 437	**254 478**
27	Electrical equipment	**215 262**	165 591	49 671	8.4	9.0	–65 505	**149 757**
28	Machinery and equipment n.e.c.	**281 622**	223 666	57 956	11.3	10.5	–103 237	**178 385**
29	Motor vehicles, trailers	**463 767**	430 020	33 747	21.7	6.1	–2 487	**461 281**
30	Other transport equipment	**31 359**	29 876	1 483	1.5	0.3	–204	**31 155**
32	Other manufactured goods	**67 030**	35 461	31 569	1.8	5.7	–2 766	**64 263**
35	Electricity, gas, steam, air cond.	**27 756**	27 756	0	1.4	0.0	0	**27 756**
58	Publishing services	**33 610**	18 533	15 077	0.9	2.7	–632	**32 978**

Note: Figures can differ from the published data due to different rounding.
Source: Czech Statistical Office

Table A2 Commodity structure of import of goods in FTS (residents and non-residents) and national concept in 2010 (CZK million)

Code	Commodity group	Cross-border statistics					National concept	
		Total	Including		Commodity stucture (%)		Adjust-ment	Total
			Residents	Non-residents	Residents	Non-residents		
Total		**2 411 556**	**2 038 334**	**373 223**	**100**	**100**	**–216 149**	**2 195 421**
	Including							
01	Products of agriculture	**40 795**	37 152	3 643	1.8	1.0	956	**41 750**
05	Coal and lignite	**5 761**	5 761	0	0.3	0.0	0	**5 761**
06	Crude petroleum, natural gas	**161 835**	161 835	0	7.9	0.0	0	**161 835**
10	Food products	**93 018**	88 776	4 242	4.4	1.1	1 106	**94 123**
12	Tobacco products	**3 501**	3 501	0	0.2	0.0	0	**3 501**
13	Textiles	**36 685**	34 776	1 909	1.7	0.5	501	**37 188**
14	Wearing apparel	**34 878**	28 176	6 702	1.4	1.8	1 709	**36 587**
15	Leather and related products	**21 555**	17 390	4 165	0.9	1.1	1 107	**22 661**
17	Paper and paper products	**45 180**	38 667	6 513	1.9	1.7	1 709	**46 889**
19	Coke and refined petroleum	**41 135**	40 957	178	2.0	0.0	44	**41 180**
20	Chemicals, chemical products	**167 688**	149 487	18 201	7.3	4.9	4 717	**172 405**
21	Basic pharmaceutical products	**73 462**	67 132	6 330	3.3	1.7	1 692	**75 155**
22	Rubber and plastics products	**112 203**	97 671	14 532	4.8	3.9	3 810	**116 013**
23	Other non-metallic mineral pr.	**33 528**	31 848	1 680	1.6	0.5	443	**33 971**
24	Basic metals	**181 568**	164 981	16 587	8.1	4.4	4 431	**186 000**
25	Fabricated metal products	**99 643**	93 296	6 347	4.6	1.7	1 568	**101 210**
26	Computer, electronic, optical pr.	**473 776**	274 319	199 457	13.5	53.4	–101 760	**372 016**
27	Electrical equipment	**166 167**	140 252	25 915	6.9	6.9	–53 985	**112 182**
28	Machinery and equipment n.e.c.	**196 897**	171 880	25 017	8.4	6.7	–92 613	**104 284**
29	Motor vehicles, trailers	**227 727**	214 748	12 979	10.5	3.5	3 397	**231 123**
30	Other transport equipment	**24 015**	22 074	1 941	1.1	0.5	599	**24 613**
32	Other manufactured goods	**50 486**	42 299	8 187	2.1	2.2	2 086	**52 572**
35	Electricity, gas, steam, air cond.	**20 843**	20 843	0	1.0	0.0	0	**20 842**
58	Publishing services	**17 711**	14 276	3 435	0.7	0.9	975	**18 687**

Note: Figures can differ from the published data due to different rounding.
Source: Czech Statistical Office

Table A3 Commodity structure of balance of trade in FTS (residents and non-residents) and national concept in 2010 (CZK million)

Code	Commodity group	Cross-border statistics			National concept	
		Total	Including		Adjustment	Total
			Residents	Non-residents		
Total		**121 239**	**–57 987**	**179 226**	**–141 818**	**–20 579**
	Including					
01	Products of agriculture	**–17 413**	–14 377	–3 036	–1 002	**–18 415**
05	Coal and lignite	**15 128**	15 128	0	0	**15 128**
06	Crude petroleum, natural gas	**–148 906**	–148 906	0	0	**–148 906**
10	Food products	**–30 740**	–29 680	–1 060	–1 359	**–32 099**
12	Tobacco products	**4 261**	–112	4 373	–321	**3 940**
13	Textiles	**5 970**	1 680	4 290	–971	**4 999**
14	Wearing apparel	**–10 481**	–14 547	4 066	–2 643	**–13 124**
15	Leather and related products	**–8 564**	–10 280	1 716	–1 630	**–10 194**
17	Paper and paper products	**–5 715**	–5 454	–261	–2 435	**–8 150**
19	Coke and refined petroleum	**–11 028**	–11 048	20	–62	**–11 090**
20	Chemicals, chemical products	**–50 511**	–53 274	2 763	–6 835	**–57 346**
21	Basic pharmaceutical products	**–40 573**	–41 334	761	–2 242	**–42 815**
22	Rubber and plastics products	**6 227**	9 831	–3 604	–4 579	**1 648**
23	Other non-metallic mineral pr.	**20 873**	17 931	2 942	–784	**20 089**
24	Basic metals	**–67 692**	–56 172	–11 520	–4 889	**–72 581**
25	Fabricated metal products	**39 712**	36 852	2 860	–3 319	**36 393**
26	Computer, electronic, optical pr.	**–45 861**	–108 652	62 791	–71 677	**–117 538**
27	Electrical equipment	**49 095**	25 339	23 756	–11 520	**37 575**
28	Machinery and equipment n.e.c.	**84 725**	51 786	32 939	–10 624	**74 101**
29	Motor vehicles, trailers	**236 040**	215 272	20 768	–5 882	**230 158**
30	Other transport equipment	**7 344**	7 802	–458	–802	**6 542**
32	Other manufactured goods	**16 544**	–6 838	23 382	–4 853	**11 691**
35	Electricity, gas, steam, air cond.	**6 913**	6 913	0	1	**6 914**
58	Publishing services	**15 899**	4 257	11 642	–1 608	**14 291**

Note: Figures can differ from the published data due to different rounding.
Source: Czech Statistical Office

Private Rate of Return on Human Capital Investment in the Czech Republic: Differences by Study Fields[1]

Savina Finardi[2] | *University of Economics, Prague, Czech Republic*
Jakub Fischer[3] | *University of Economics, Prague, Czech Republic*
Petr Mazouch[4] | *University of Economics, Prague, Czech Republic*

Abstract

The paper is focused on approaches to the measurement of the returns of private investments on human capital in the Czech Republic. In the last ten years, there is observed a significant increase in number of students at Higher Education Institutions (HEIs) and an increasing number of HEIs graduates is also expected in the Czech Republic in forthcoming years. Using data from the research project "REFLEX", from the Czech Statistical Office and from EUROSTUDENT IV survey, the paper provides the methodology and the experimental computations of the rates of return on private investment in the tertiary education broken down by study fields.

Keywords	***JEL code***
Human capital, rates of return, higher education, private investment	*I23*

INTRODUCTION

It is very popular to analyse the returns on human capital. Investments in the human capital can be assessed from different points of view. We can see these investments from the point of view of an individual: he has some opportunity costs (due to the postponing of starting his entrance to the labour market as well as the direct costs of the education such as tuitions, living costs in the university town, transport charges and so on). On the other hand, more educated person have higher wages, lower risk of unemployment,

[1] This article has been written under the support of the Internal Grant Agency of the University of Economics, Prague, project No. IGS/MF/28/2010, "Economic and Demographic Aspects of Tertiary Education".
[2] Nám W. Churchilla 4, 130 67 Prague 3, Czech Republic. Corresponding author: e-mail: savina@finardi.cz, phone: (+420) 723 009 080.
[3] Nám W. Churchilla 4, 130 67 Prague 3, Czech Republic.
[4] Nám W. Churchilla 4, 130 67 Prague 3, Czech Republic.

higher retire pension, longer life expectancy etc. These costs and benefits could be divided to two main groups — economic and non-economic.

From the point of view of the society, we can consider at the cost side public expenditures on education, opportunity costs such as lower tax revenues, and at the benefit side higher tax revenues in future, lower unemployment, higher gross value added, gross domestic product and so on.

1 DATA AND METODOLOGY

We use three different data sources: data from REFLEX survey, data on wages from the Czech Statistical Office (CSO) and finally the data from survey EUROSTUDENT IV. CSO data provide us information on level of wages depending on age, attained education and study field. REFLEX survey collected data on tertiary-educated persons after finishing their studies (from year 2000 till 2003) and then after 5 years (mainly 2005–2006). Most of graduates finished their studies in 2001 and 2002 and therefore we consider years 2001 and 2006 as basic years for our research. In total, REFLEX survey collected 6 794 responses, which means about 23% response rate. 17% of the sample are bachelor graduates and 82% master graduates, 57% are women and 43% men. The structure of graduates by study field is following: 27% economic, law and human sciences, 23% technical, 19% pedagogical. 90% of graduates studied in full-time study programs. Introduction and methodology of REFLEX survey has also information about age structure and regional structure of the Czech HEIs.

The third data source is the EUROSTUDENT survey, which has been realized during year 2009 and contains 8 386 responses (60% response rate). Only 7 166 observations related to full-time students are analyzed. 6 885 students studied at public HEIs, 281 students at private ones. According to the study cycle, the structure is following: 64.6% bachelor students, 17.9% long master students, 17.1% short master students and 0.3% Ph.D. students. The EEUROSTUDENT survey is a part of the international and periodic project, which includes all developed European countries (EU, EEA, Croatia and Turkey).[5]

For our experimental computations, we consider the differences between the net wages of the tertiary-educated person and the net wages of the upper-secondary-educated person with the General Certificate of Education (GCE) at the side of benefits and the EUROSTUDENT estimation of costs of studies at the tertiary stage as costs. We do not consider the risk of unemployment and the consequent losses at the side of benefits (construction of a probabilistic model is a complicated task due to the necessity of solving the problem of different wages and the probability of unemployment among the different age groups), but we consider the differences at the level of the retirement pensions; it is necessary to say that the regressive model is used in the process of computation of retirement pensions. The pensions from the Czech pension scheme (the Pay-as-You-Go model is currently used) depend on the number of years of the productive activity and on the amount of wages paid, but the amount of wages paid is reduced for the highest levels. We used the model for computing of the internal rate of return after 50 years of working. We do not consider the non-economic cost and benefits of the tertiary education (such as a better health, lower rate of criminality, etc.).

We compare wage development of two hypothetical individuals, who decided about their future in 1996. They have completed their upper secondary studies and one of them attends the 5-year tertiary education level and the second one joins the labour market. We use the data from the Czech Statistical Office on the distribution of the wages by age and by the highest level of education (two-dimensional cross table), published for years 2001 and 2006 as well. From these tables, we can use the levels of wages for an upper-secondary-educated person after 5-year-practice (he finished his secondary studies in 1996 and has 5 years of practice in 2001) and the starting wages for tertiary-educated person in 2001. We can

[5] More information about Eurostudent survey at: <http://www.eurostudent.eu>.

also estimate future development of the wages of both individuals using the longitudinal analysis of the wages, which depend on the age and on the education level as well. This estimation was recomputed by the newer data from the Czech Statistical Office about level of wages in years 2005 and 2010. All previous computations (Finardi, Fischer, Mazouch, 2008a) were recounted due to the financial and economic crisis, which affected the Czech economy and led to a lower level of wages in comparison with assumptions made in 2007.

The data on development of wages (with respect to age and educational profile as well) are shown in Tables 1, 2 and 3. Table 1 shows data from REFLEX survey, which includes data on wages of HEIs graduates. These data have been used for the first six-year prognosis of graduates' nominal wages (2001–2006). The average annual growth differs across different study fields.

Table 1 Wage Development Between 2001 and 2006, tertiary-educated (REFLEX survey data)

Study Fields	Monthly Wage (CZK)		Average Annual Growth (2001–2006)
	2006	2001	
Natural Sciences	29 790	14 812	1.1500
Technical	29 898	14 932	1.1490
Agricultural	21 755	11 715	1.1318
Medical	28 072	12 007	1.1851
Economic	32 530	15 854	1.1546
Human Sciences	25 234	13 492	1.1334
Pedagogical	21 855	11 572	1.1356

Source: REFLEX survey, own calculation

Table 2 shows the development of wages of upper-secondary-educated persons in the Czech Republic in the previous years. Average growth between years 2005 and 2010 has slowed down compared to years 2001 and 2006 and there is a change in wages of employees in the highest age categories: the age group of 65 and more has lower monthly wage than age group 60–64 years. Table 3 includes data on wage development of tertiary-educated persons between the same time periods as in Table 2. In both tables, we can observe that at the beginning of career the index of average annual growth being higher than

Table 2 Wages Development Between 2010 and 2005, upper-secondary-educated

2010		2005		Average Annual Growth (2005–2010)
Age Group	Monthly Wage (CZK)	Age Group	Monthly Wage (CZK)	Nominal Wages
25–29	25 657	20–24	17 026	1.08547
30–34	29 551	25–29	21 804	1.06269
35–39	30 405	30–34	23 885	1.04946
40–44	28 876	35–39	22 888	1.04758
45–49	28 579	40–44	22 771	1.04648
50–54	28 208	45–49	22 859	1.04295
55–59	28 825	50–54	23 408	1.04251
60–64	30 973	55–59	24 137	1.05114
65 and more	26 466	60–64	25 211	1.00976

Source: Czech Statistical Office, own calculation

Table 3 Wages Development Between 2010 and 2005, tertiary-educated

2010		2005		Average Annual Growth (2005–2010)
Age Group	**Monthly Wage (CZK)**	**Age Group**	**Monthly Wage (CZK)**	**Nominal Wages**
30–34	46 119	25–29	27 774	1.10675
35–39	56 749	30–34	39 690	1.07413
40–44	56 853	35–39	42 170	1.06157
45–49	52 646	40–44	39 608	1.05856
50–54	49 969	45–49	39 234	1.04956
55–59	48 722	50–54	39 401	1.04338
60–64	50 966	55–59	39 384	1.05291
65 and more	46 824	60–64	40 403	1.02994

Source: Czech Statistical Office, own calculation

in the forthcoming years. This feature is the same for upper-secondary-educated and form tertiary-educated employees. This is caused by the starting position in the labour market. During the first years after graduation the nominal growth of wages is quicker than after 10 and more years of working experiences.

Table 4 Rate of Inflation (in %)

2010	1.5	**2007**	2.8
2009	1.0	**2006**	2.5
2008	6.3	**2005**	1.9

Source: Czech Statistical Office

Table 5 Real Growth of Wages Between 2005–2010, upper-secondary-educated

Age Group	Index of Average Growth in Nominal Wages	Average Inflation Rate (6 years)	Average Risk-Free Interest Rate Index (6 years)	Real Growth of Wages RI (2005–2010)	Real Growth of Wages RFIR (2005–2010)
25–29	1.08547	1.022381	1.040597	1.061707	1.043122
30–34	1.06269	1.022381	1.040597	1.039426	1.021231
35–39	1.04946	1.022381	1.040597	1.026486	1.008517
40–44	1.04758	1.022381	1.040597	1.024647	1.006710
45–49	1.04648	1.022381	1.040597	1.023571	1.005653
50–54	1.04295	1.022381	1.040597	1.020118	1.002261
55–59	1.04251	1.022381	1.040597	1.019688	1.001838
60–64	1.05114	1.022381	1.040597	1.028129	1.010131
65 and more	1.00976	1.022381	1.040597	0.987655	0.970366

Note: RI — Inflation Rate, RFIR — Risk-Free Interest Rate.
Source: Czech Statistical Office, Ministry of Finance of the Czech Republic, own calculation

We can compare the index of average growth of wages from 2005 till 2010 with the index of average rate of inflation measured by the CPI (Consumer Price Index); the index of average rate of inflation is 1.022381 (see Table 4). The index of average rate of inflation (1.022381) is significantly lower than the index of average risk-free interest rate from 2005 till 2010 (1.040597). The real growth in wages of upper-secondary-educated persons is showed in table 5 and for tertiary-educated persons in Table 6. The indices of average nominal growth of wages divided into age groups were compared with inflation rate (CPI) and with risk-free interest rate index. Both indices mentioned above are computed as a geometric mean from 2005 till 2010.

Table 6 Real Growth of Wages Between 2005–2010, tertiary-educated

Age Group	Index of Average Growth in Nominal Wages	Average Inflation Rate (6 years)	Average Risk-Free Interest Rate Index (6 years)	Real Growth of Wages RI (2005–2010)	Real Growth of Wages RFIR (2005–2010)
30–34	1.10675	1.022381	1.040597	1.082518	1.063569
35–39	1.07413	1.022381	1.040597	1.050612	1.032221
40–44	1.06157	1.022381	1.040597	1.038334	1.020157
45–49	1.05856	1.022381	1.040597	1.035389	1.017264
50–54	1.04956	1.022381	1.040597	1.026584	1.008614
55–59	1.04338	1.022381	1.040597	1.020541	1.002676
60–64	1.05291	1.022381	1.040597	1.02986	1.011834
65 and more	1.02994	1.022381	1.040597	1.007391	0.989756

Note: RI – Rate of Inflation, RFIR – Risk-Free Interest Rate.
Source: Czech Statistical Office, Ministry of Finance of the Czech Republic, own calculation

Table 7 Risk-free Interest Rate (in %)

2011	4.00	**2006**	3.78
2010	3.71	**2005**	3.51
2009	4.67	**2004**	4.75
2008	4.55	**2003**	4.12
2007	4.28	**2002**	4.94

Source: Ministry of Finance of the Czech Republic, Czech National Bank

Table 8 Average Costs and Incomes of Public HEISs Students

Year	Average Costs of Public HEISs Students (CZK / per month)	Average Incomes of Public HEIs Students (CZK / per month)
2009	8 448	6 748
2008	7 947	6 348
2007	8 170	6 175
2006	7 971	6 025
2005	7 312	5 527
2004	7 113	5 377
2003	7 106	5 371
2002	6 580	4 973
2001	6 284	4 750
2000	6 049	4 572
1999	5 924	4 478
1998	5 352	4 045
1997	4 932	3 728
1996	4 533	3 427

Source: Eurostudent IV, Czech Statistical Office, own calculation

All the computations are in nominal values, but if we want to express the nominal wages in real value, we will use the risk-free interest rate.[6] Table 7 includes interest rates of middle-term and long-term bonds issued by the Czech National Bank. The annual average risk-free interest rate is 1.042056. Also the risk-free interest rate is lower than 5 years before — 1.05869 (Finardi, Fischer, Mazouch, 2008a).

For all the estimations and computations we consider tax conditions of year 2011: flat tax rate 15%, concept of so-called super-gross wage, social insurance rate paid by employer 34% and paid by employee 11%, tax credit for tax payer 1 970 CZK per month (23 640 CZK per year). Also the tax credit was discounted by the index of average growth of wages between years 1996 and 2010. The monthly average wage in year 1996 was 11 069 CZK and monthly average wage in year 2010 was 26 881 CZK; the final index of the growth of wages is 1.040295.

[6] Risk-free interest rate is mainly used for computations, which are connected with the public sector. Rate of inflation is mostly used for computations of purchasing power of consumers.

From EUROSTUDENT survey, we consider incomes and costs of public HEIs students only. The monthly average income of these students was 6 748 CZK in 2009. We discount the average income by the CPI from 1996 till 2009, because we estimated private rate of return for students of public HEIs, which began their studies mostly in year 1996 and finished studies in year 2000. The same method was used for costs of public HEIs students. Table 8 shows average incomes and costs of students per month. Total average costs from 1996 till 2000 reaches 321 478 CZK and total average incomes 242 988 CZK.

For the estimation of private returns on human capital we use the method of discount factor (Maříková, Mařík, 2007). In the first step we compute a Wage Premium (WP) for the HEIs graduates. In the second step we could finally estimate the discount factor:

$$WP = \sum \frac{W_{te} - W_{se} - C_{he}}{(1 + DF)^i}, \qquad (1)$$

where:

WP is Wage Premium,
ΣW_{te} is a sum of nominal wages of tertiary-educated graduates for their working cycle,
ΣW_{se} is a sum of nominal wages of secondary-educated graduates for their working cycle,
ΣC_{he} is a sum of costs on studies of tertiary-educated graduates,
DF is discount factor,
i is a duration of cycle.

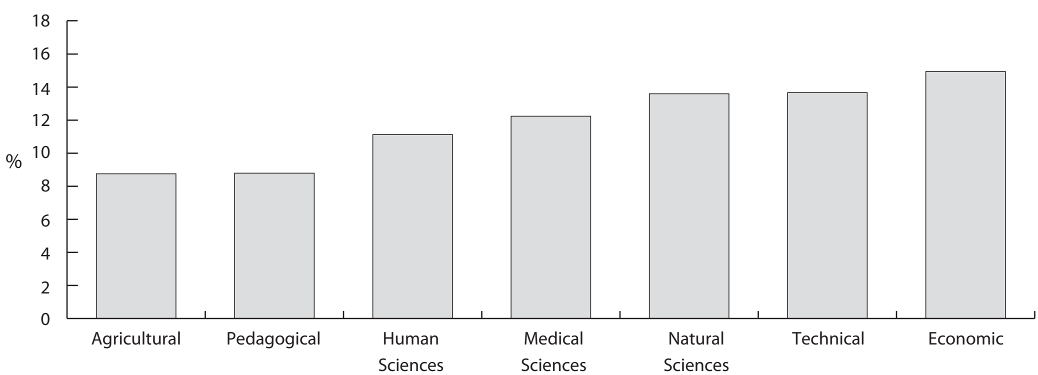
The equation (1) is solved for unknown DF, when $WP = 0$.

2 RESULTS AND INTERNATIONAL COMPARISON

Figure 1 shows the private rate of return on human capital broken down by study fields. The lowest rate of return is for agricultural studies graduates and the highest rate is for economic studies graduates. This is not surprising, because the labour market is still very "hungry" for economists.

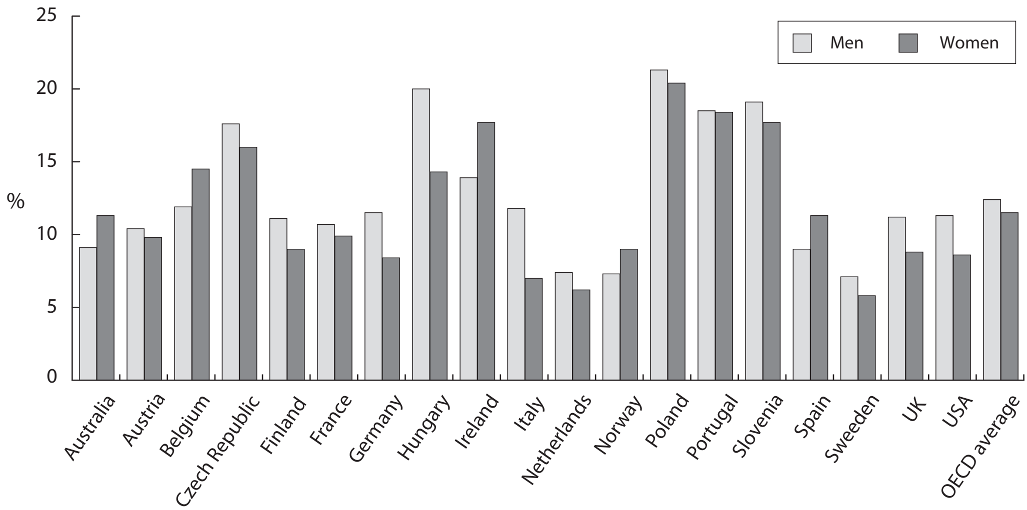
Figure 2 describes private rate of return computed by OECD experts on education in the annual report called Education at a Glance 2011. The rates were computed on the data from year 2007. It is obvious that the Czech Republic is above the OECD average and this is mainly caused by the fact that there are no tuition fees. The private rate of return is mainly depends on the level of wages and wage premium for

Figure 1 Private Rate of Return on Human Capital in the Czech Republic according to Study Fields



Source: Own calculation

Figure 2 Private Rate of Return on Human Capital in the OECD Countries (tertiary education)



Source: Education at a Glance 2011 (OECD)

tertiary-educated persons and secondly on the tuition fees (if collected), see Figure 2. Only in 5 countries (Australia, Belgium, Ireland, Norway and Spain) are rates of return higher for women than for men. In Australia and Belgium taxes for men are higher than for women. Belgium and Spain have progressive tax rates of personal income tax. In Ireland special tax deductions are applied for married couples, because it is a country with a high share of religious residents.

CONCLUSION

Computing and estimating private and also public rates of return is very important for a future discussion about proposals of tuition fees in the Czech Republic. In our further research, we plan to include in our model tuition fees, progressive personal income taxes and estimate the rates of return for men, women and different regions of the Czech Republic. Estimations would help to set optimal rates of tuition fees for different faculties, study programmes or study fields.

The Ministry of Education, Youth and Sports of the Czech Republic presented a reform of a tertiary education system including tuition fees; therefore it is necessary to have detailed information about future incomes of HEIs graduates. So far, no research made an estimation of wage premium for different study fields in the Czech Republic.

References

CENTER OF EDUCATIONAL POLICY. *Final Report of the REFLEX Survey in the Czech Republic* [online]. Prague: Center of Educational Policy, September 2007. [cit. 3.2.2012]. <http://www.strediskovzdelavacipolitiky.info/download/ZavZprReflex.pdf>.

FINARDI, S., FISCHER, J., MAZOUCH, P. About Possibilities How to Measure the Investments into the Education by Study Fields. In *New Trends — New Ideas,* 2008a, p. 11.

FISCHER, J., FINARDI, S., MAZOUCH, P. Current Value of Money in Relation to Public Investment Return On Human Capital. In *International Days of Statistics and Economics,* 2008b, pp. 1–7.

HECKMAN, J. J., LOCHNER, L. J., TODD, P. E. *Earnings Functions and Rates of Return* [online]. January 2008. [cit. 3.2.2012]. <http://ftp.iza.org/dp3310.pdf>.

HECKMAN, J. J., LOCHNER, L. J., TODD, P. E. *Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond* [online]. August 2005. [cit.3.2.2012]. <http://ftp.iza.org/dp1700.pdf>.

MAŘÍKOVÁ, P., MAŘÍK, M. *Discount Rate for Business Valuation.* Prague: Oeconomica Publishing, 2007.

MAZOUCH, P., FISCHER, J. *Human Capital — Measuing, Context, Prognosis.* Prague: C. H. Beck, 2011.

OECD. *Education at a Glance 2011* [online]. OECD, 2011. [cit. 3.2.2012]. <http://www.oecd.org/dataoecd/61/2/48631582.pdf>.

PSACHAROPOULOS, G. *The Profitability of Investment in Education: Concepts and Methods* [online]. December 1995. [cit. 3.2.2012]. <http://www.c3l.uni-oldenburg.de/cde/econ/readings/psacharo.pdf>.

RADOVÁ, J., DVOŘÁK, P., MÁLEK, J. *Financial Mathematics for Everybody.* Prague: Grada Publishing, 2009.

ŠIROKÝ, J. *Taxes in the European Union.* Prague: Linde, 2010.

URBÁNEK, V. *Financing of Higher Education.* Prague: Oeconomica Publishing, 2007.

Evaluation of Economic Education from Graduates' Point of View

Renáta Kunstová[1] | *University of Economics, Prague, Czech Republic*
Hana Řezanková[2] | *University of Economics, Prague, Czech Republic*

Abstract

The labour market expects graduates with certain levels of competences, which reflect the quality of education. In this paper we present the results of the analyses of selected indicators concerning of the education quality obtained on the basis of answers of graduates of the University of Economics, Prague. The graduates were addressed four or five years after graduation within large REFLEX surveys realized in 2006 and 2010. We compare competence levels acquired by graduates with competence levels required by employers; both types of levels were evaluated by graduates. We investigate dependency, agreement and similarity of acquired and required competence levels by different coefficients and we compare their values in the 2006 and 2010 surveys.

Keywords

Higher education, graduates, education evaluation, acquired competence levels, required competence levels

JEL code

A23, I21, J24

INTRODUCTION

One of many reasons, why universities should pay attention to evaluation of education quality, is investigation of reform implications. The Bologna declaration, signed on June 1999 by ministers in charge of higher education, started reforms of higher education in many European countries. In accordance with these reforms, since the beginning of the 2000s, study programmes have been converted into a three-cycle structure of higher education (bachelor's, master's and doctoral degrees) with the uniform European credit transfer and accumulation system (ECTS). It is obviously useful to investigate whether the transformation of higher education influenced the acquired knowledge, skills and abilities of graduates.

From the partial conclusions published in OECD (2010) it appears that the Czech Republic is a country with a low level of tuition fees, with no financial or other barriers to entry to higher education, with the most significant increasing number of students admitted to universities, and unemploy-

[1] Nám. W. Churchilla 4, 130 67 Prague 3, Czech Republic. Corresponding author: e-mail: renata.kunstova@vse.cz, phone: (+420) 224 095 418.
[2] Nám. W. Churchilla 4, 130 67 Prague 3, Czech Republic.

ment is generally still low. According to the data published by the Czech Statistical Office (2011), the total number of students studying at the universities has almost doubled during the past ten years. Because universities are a part of the labour market, and we can assume that this market will soon be saturated, universities have to be prepared for the competitiveness of their graduates (European Commission, 2003). According to Koucký, Zelenka (2010), the unemployment rate of university students under the age of 30 years moves on a long-term basis in the range from 50% to 90% of the total unemployment rate in the Czech Republic. Although the unemployment rate of university students is below the national average, their number increased in the last four years also as a result of the economic crisis. For example, Koucký, Zelenka (2010) has published, that unemployment rate of university students increased from 1.5% in 2008 to 2.4% in 2010. The Czech Republic has started to tackle the unnaturally high number of universities graduates, but, according to Doucek et al. (2011), it is a long-term process. Universities have to deal with the quality of their students to be competitive in the saturated labour market.

In this paper we analyse the data collected by the REFLEX surveys in 2006 and 2010, in which the Czech Republic participated. We deal with data obtained from graduates of five faculties of the University of Economics, Prague, and we focus on selected indicators of education quality. We are interested in the evaluation in relation to employers and further professions of graduates, and in the competence levels acquired by graduates and required by employers.

1 SURVEYS OF HIGHER EDUCATION QUALITY

Different data collections concerning higher education have been realized in the last few years. The series of books Education at a Glance published from 1998 with the latest in 2011 (OECD, 2011) provides internationally comparable data on education. Each publication focuses on four main topics: education levels and student numbers, the economic and social benefits of education, paying for education and the school environment. Educational systems are compared mainly by means of quantitative indicators.

For the evaluation of education it is very important to know how well students are prepared for various professions. However, the measuring of this aspect is very difficult. A wide variety of potential indicators of a graduate's performance was reviewed by Hartnett and Willingham (1979). Emphasis was placed on problems with the selection of indicators and their definitions. The available results of other surveys are primarily oriented towards single evaluations of the data (Mason, 2001, Archer, Davidson, 2008).

Several surveys focusing on acquired competences of graduates have been realized in the past decade. They followed the CHEERS[3] project (Careers after graduation from Higher Educational institutions — a European Research Study), which was realized in the years 1998–2000 in twelve countries and concerned graduates from the 1994.

An important European project was REFLEX[4] (Research into Employment and professional FLEXibility), which was realized in the years 2004–2007 in 16 countries. It concerned graduates from 2001 and 2002. Besides European countries (including non-EU members), Japan participated in this project. One of the objectives was a qualitative study on graduate competences in the knowledge society. The structure of the questionnaire was unified for all countries (the questionnaire was translated into the native language in each country). General results from the international point of view were published, for instance, by Allen, Van Derveldend (2007) and Petersson (2007).

---

3 Available at: <http://www.uni-kassel.de/incher/cheers/index.ghk>.

4 Available at: <http://www.fdewb.unimaas.nl/roa/reflex>.

The PROFLEX[5] project (with the title "Flexible Professional in the Knowledge Society: New Demands on Higher Education in Latin America"), was undertaken using some parts of the REFLEX methodology in six countries of Latin America. The HEGESCO[6] project (Higher Education as a Generator of Strategic Competences) addressed the needs of the main groups of higher education stakeholders who were interested in the employability of graduates. It was based on qualitative interviews among employers and higher education institutions from five partner countries. It was realized in eight countries of southern and eastern Europe.

The DECOWE Network[7] (Development of Competencies in the World of Work and Education), was established after the conference held in September 2009 at the University of Ljubljana. The main purpose of this initiative is to promote relevant research, researchers, teaching and governmental projects, initiatives and events related to improvements in educational and employability policies, studies related to development of competences in different environments and establishment of learning recognition and qualification frameworks.

The second REFLEX project, with the title "Employability and graduates' labour market success", was realized in the Czech Republic in 2010. Similar surveys were also being undertaken in Austria and Germany at that time. General results of the REFLEX 2006 and 2010 surveys related to the Czech Republic were published, for example, by Kalousková (2006), Kalousková, Vojtěch (2008), Koucký, Lepič (2008) and Koucký, Zelenka (2010).

2 CHARACTERIZATION OF ANALYSED DATA FILES

In the Czech Republic, projects REFLEX 2006 and REFLEX 2010 were coordinated by the Education Policy Centre at the Charles University in Prague. Selection of the graduates was designed as regional where individual faculties corresponded to regions. For the survey, the Education Policy Centre (EPC) determined numbers of graduates for individual faculties and individual years.

The technique of the graduate selection was rather complicated. The number of addressed graduates was specified on the basis of the number of the faculty's graduates. It was determined as a percentage of graduates, from 33% (each third graduate from the alphabetic list was asked) to 100% (all graduates were asked). The EPC assumed that 20% of questionnaires were fulfilled enough with using the possibility of searching for other contacts in case if a postal address was invalid (e.g. by e-mail address). However, a graduate degree was not taken into consideration.

We analysed the data relating to five faculties of the University of Economics in Prague, which participated in both surveys. The numbers of received fulfilled questionnaires desired by the EPC were achieved in case of these faculties. In accordance with the rules of use and publishing results of the REFLEX project we do not mention the names of these faculties.

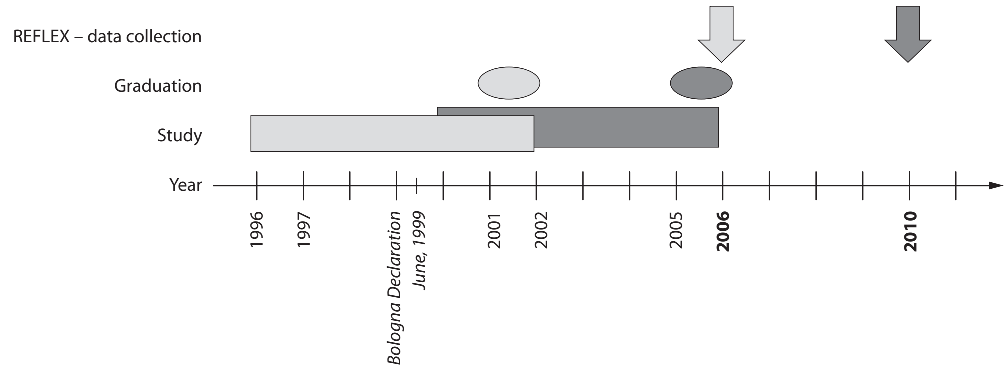
The analysed data sample includes only graduates with a master's degree because they account for the major part of all graduates in investigated periods. The numbers of these respondents were 412 in 2006 and 506 in 2010. The continuity of data collection to the end of the study is shown in Figure 1. Graduates from ISCED (International Standard Classification of Education) 5A programmes (UNESCO, 1997), who got their degree in 2001 or 2002 (in the 2006 survey) and 2005 or 2006 (in the 2010 survey), were involved. It means that respondents were addressed four or five years after graduation. Both periods are displayed in the length of the whole study in Figure 1. In the earlier period students completed a five-year study, three-cycle system students achieved the same level of study after completing bachelor's (three years) and master's (two years) degrees of study.

5 Available at: <http://www.encuesta-proflex.org>.

6 Available at: <http://www.hegesco.org/content/view/8/10>.

7 Available at: <http://www.decowe.com>.

Figure 1 The timeline of the REFLEX projects



Source: Own construction, data origin from REFLEX 2006 and REFLEX 2010

The questionnaires used in surveys 2006 and 2010 were divided into several sections which were focused on study and work orientations, evaluation of the educational programme, work experiences before and during higher education, the transition to the labour market, characteristics of the first and current job, characteristics of the occupational and labour market career up to the present, assessment of required and acquired competences, etc. The questionnaires in 2006 and 2010 were not identical, only similar. Some questions were changed and new questions were added in the 2010 questionnaire in accordance with the experiences from the first survey and new circumstances. We focused on questions from a few selected sections in our analysis.

Firstly there is evaluation of study programmes from different aspects, including relationship to employers and further professions of graduates. Five (or four) years after graduation respondents re-evaluated their university studies. They could compare content and demands of study with their employment needs. Respondents judged their study from the following aspects: an overall concept of the study programme, the study programme as a basis for future professional and personal development and satisfaction with the selection of the study programme and the university. The analysed indicators were defined identically in both surveys.

Secondly there is the comparison of graduates' acquired and required levels of competences. Although respondents are employed, their satisfaction with the way the faculty prepared them for their professions may be various. Both acquired and required levels of competences were evaluated by graduates. The indicators of competences were various in the surveys. They differed in the number of competences, their formulations and in the rating scale. In 2006 respondents evaluated competences on a seven-point scale and in 2010 the scale was ten-point. We therefore focused on the analysis of selected competences, whose formulations were similar. Due to comparability of association coefficients in different periods, we recalculated both original scales to a three-point scale.

3 RESULTS OF ANALYSES

In this section we present the results of comparison of selected indicators concerning evaluation of study programmes from the surveys in 2006 and 2010. Besides the percentage distributions of individual categories, different independency tests were applied for investigation of statistical dependency of indicators on a study period. In addition, different measures of dependency, agreement and similarity are used for relationship investigation of acquired and required competence levels. Statistical calculations were performed in the IBM SPSS Statistics and MS Excel systems.

3.1 Evaluation of study programmes

Percentage distribution of the evaluating scale concerning description of the study programme concepts is showed in Table 1. Grey colour indicates statistical dependency of the evaluation on a year according to different independency tests in a contingency table, including the chi-square test and zero tests for asymmetric tau and uncertainty coefficients. Respondents of the 2006 survey underwent a different structure of study from respondents of the 2010 survey, whose studies have already been influenced by the Bologna process. The frequency distribution shows that the study of economic disciplines was regarded as medium and rather demanding in both periods.

Table 1 Percentage distributions for description of study programmes

Description of study programme	Year	1 not at all	2	3	4	5 very high extent
Programme was generally regarded as demanding	2006	1.0	16.3	38.6	36.4	7.7
	2010	1.4	18.4	42.1	31.8	6.3
Employers are familiar with the content of programme	2006	7.4	33.7	29.5	24.6	4.9
	2010	6.5	30.2	25.3	32.6	5.3
There was freedom in composing your own programme	2006	3.9	14.5	27.8	35.1	18.7
	2010	2.2	17.8	32.2	38.3	9.5
Programme had a broad focus	2006	0.2	14.7	28.7	47.3	9.1
	2010	0.6	16.4	22.5	49.0	11.5
Programme was vocationally orientated	2006	2.7	22.7	34.0	33.5	7.1
	2010	4.2	26.5	30.8	32.0	6.5
Programme was academically prestigious	2006	4.4	34.2	30.8	22.2	8.4
	2010	4.3	28.1	27.3	32.4	7.9

Source: Own calculation, data origin from REFLEX 2006 and REFLEX 2010

Employers were more familiar with the content of the programme in the later period. However, the survey does not tell us whether faculties have better public relations or employers searched for such information.

The smaller degree of freedom in composing one's own programme in the 2010 survey is related to the division into bachelor's and master's studies; we can consider the same reason for the changes in the broad focus of the study programme and in vocational orientation. Academic prestige had relatively low ratings in the 2006 survey. In the later period this prestige changed significantly in favour of higher evaluating levels.

Published statistics of employment and unemployment provide initial information about the employability of graduates in the labour market. However, although respondents are employed, their satisfaction with the way the faculty prepared them for their professions may be various. Results in Table 2 provide a comparison of percentage distributions concerning preparedness for future professions (grey colour indicates statistical dependency of the evaluation on a year at 5% significance level). Most of the frequencies of higher levels are lower in the 2010 survey, but the obtained values are positive in general.

With the exception of one indicator, more than 70% of the answers were at the middle or higher levels in both surveys. The evaluation of "good basis for respondent's personal development" was the best. Only the last indicator "development of entrepreneurial skills" was evaluated more by lower categories.

Respondents of both surveys answered similarly to the question of whether they would choose the same study programme at the university on the basis of their current opinions (see Figure 2). Sixty-one

Table 2 Percentage distributions for evaluation of study programmes in relation to future professions

Study programme was a good basis for	Year	1 not at all	2	3	4	5 very high extent
Starting work	2006	6.1	9.8	24.9	35.0	24.2
	2010	8.1	16.8	30.6	31.2	13.2
Further learning on the job	2006	5.2	11.9	26.6	43.2	13.2
	2010	5.7	17.2	30.6	31.4	15.0
Performing current work tasks	2006	5.2	17.0	35.2	30.3	12.3
	2010	6.9	20.2	29.4	33.0	10.5
Future career	2006	4.4	9.3	32.2	40.0	14.0
	2010	4.7	20.9	29.6	33.2	11.5
Your personal development	2006	1.7	7.6	23.4	44.6	22.7
	2010	3.2	10.3	28.3	40.5	17.8
Development of entrepreneurial skills	2006	25.1	29.6	23.4	18.4	3.5
	2010	25.1	31.0	24.7	15.0	4.2

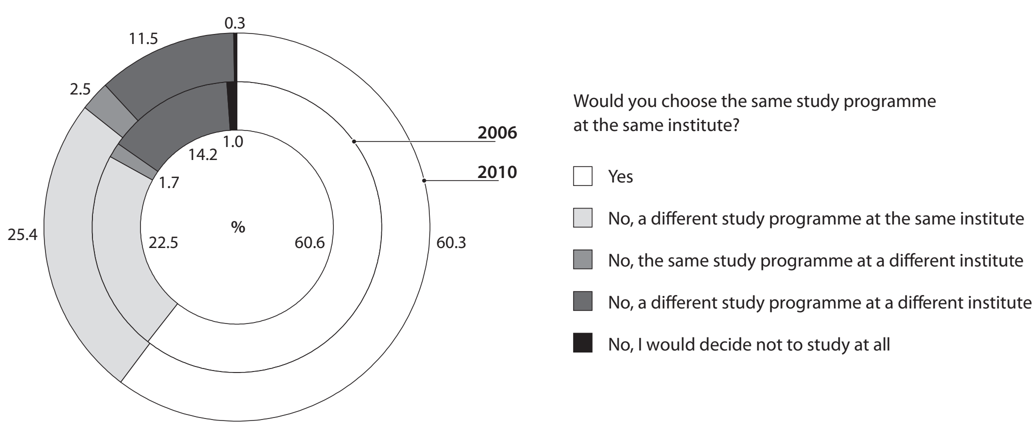
Source: Own calculation, data origin from REFLEX 2006 and REFLEX 2010

percent of respondents would choose the same programme at the same university. Over twenty percent of respondents would change the study programme but not the university. The number of respondents who would change the study programme and the university was higher in 2006.

3.2 Evaluation of acquired and required competence levels

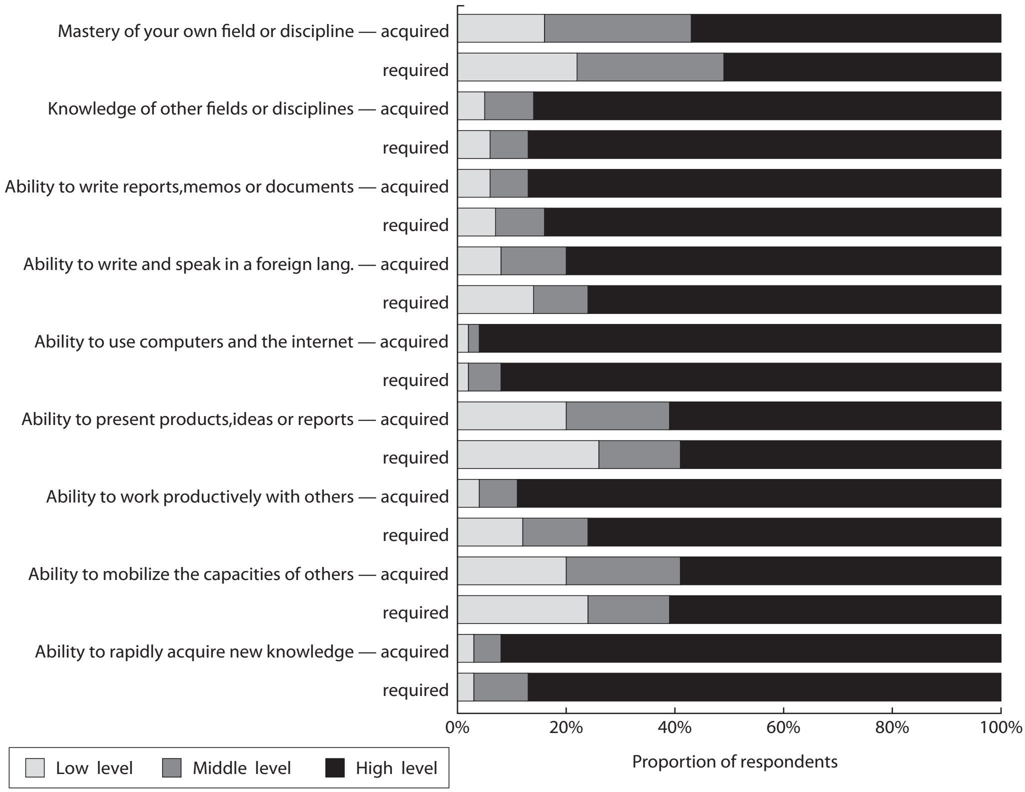
Figures 3 and 4 present the comparison of acquired and required competence levels in both periods. We can see that the level of an acquired competence is almost always higher than the level of a required competence. The comparison of investigated periods shows that results are better from the 2010 survey. In this year respondents evaluated the level of acquired competences always as higher and with a greater difference compared with the level of required competences.

Figure 2 Percentage distributions of responses to the choice of a study programme (in %)



Source: Own construction, data origin from REFLEX 2006 and REFLEX 2010

Figure 3 REFLEX 2006 – graduates' acquired and required levels of selected competences (in %)



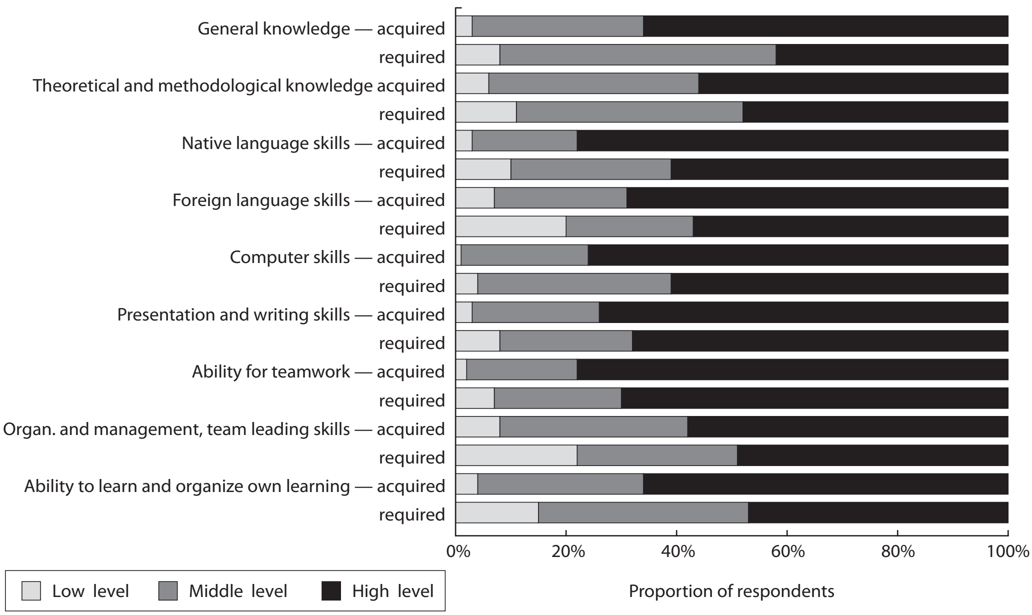
Source: Own construction, data origin from REFLEX 2006

The biggest difference between acquired and required competence levels was in "ability to work productively with others" (in 2006) and in "general knowledge" (in 2010). On the other hand, the smallest difference was in "knowledge of other fields or disciplines" (in 2006) and in "ability for teamwork" (in 2010).

Further, we investigated dependency, agreement and similarity of acquired and required competence levels. We applied Kendall's tau-b as a measure of dependency, Cohen's kappa as a measure of agreement and the cosine measure for investigation of similarity. Computational formulae and properties of these measures are described, for example, by Pecáková (2011), Řezanková (2011) and Řezanková et al. (2009). The obtained values are in Table 3. They need to be considered with the relationships to percentage distributions presented in Figures 3 and 4. If the highest category predominated, then the relationship between levels of acquired and required competences is more important. In all cases, dependency and agreement were statistically significant at 1% significance level.

In the 2006 survey the highest dependency, agreement and also similarity were in the case of "ability to use computer and the Internet". Higher values are related to dominance of the third category. In the 2010 survey, the value of Kendall's tau-b was the highest for "organization and management, team leading skills", whereas Cohen's kappa and the cosine measure were the highest for "ability for teamwork". In this case the relationship between levels of acquired and required competences is more important because the proportion of the third category is higher.

Figure 4 REFLEX 2010 – graduates' acquired and required levels of selected competences (in %)



Source: Own construction, data origin from REFLEX 2010

Table 3 Evaluation of the relationships between acquired and required competence levels

Competence	Year	Tau-b	Kappa	Cosine measure
Mastery of your own field or discipline	2006	0.385	0.315	0.983
General knowledge	2010	0.409	0.336	0.969
Knowledge of other fields or disciplines	2006	0.349	0.294	0.936
Theoretical and methodological knowledge	2010	0.435	0.369	0.966
Ability to write reports, memos or documents	2006	0.485	0.473	0.984
Native language skills	2010	0.509	0.441	0.973
Ability to write and speak in a foreign language	2006	0.357	0.315	0.969
Foreign language skills	2010	0.413	0.307	0.955
Ability to use computers and the internet	2006	0.573	0.507	0.996
Computer skills	2010	0.419	0.376	0.977
Ability to present products, ideas or reports to an audience	2006	0.408	0.333	0.938
Presentation and writing skills	2010	0.500	0.431	0.978
Ability to work productively with others	2006	0.297	0.214	0.969
Ability for teamwork	2010	0.566	0.484	0.983
Ability to mobilize the capacities of others	2006	0.465	0.347	0.950
Organization and management, team leading skills	2010	0.574	0.447	0.963
Ability to rapidly acquire new knowledge	2006	0.338	0.291	0.986
Ability to learn and organize own learning	2010	0.457	0.340	0.962

Source: Own calculation, data origin from REFLEX 2006 and REFLEX 2010

CONCLUSION

Employability of graduates is one of the general criteria of the universities evaluation. It is not possible to make the simple conclusions that the smaller the unemployment of graduates is, the better their study was. In this paper we focused on the opinions of graduates and their retrospective evaluation of completed study programmes. We analysed the answers of two graduate groups, which differed in their study period. The respondents graduated at the same economic faculties and they were addressed four or five years after graduation. Groups differed in the structure of the study programme, which had changed between the two investigated periods of studies.

Employers were more familiar with the content of a programme in the later period (in the 2010 survey). Less freedom in the composing of graduates' study programmes was found in this period. Academic prestige had relatively low ratings in the 2006 survey; in the later period this prestige increased significantly. But in both periods the study of economic disciplines was mostly regarded as either middling or rather demanding.

The respondents evaluated their study programmes in relation to future professions very well. The evaluation of "good basis for respondent's personal development" was the best. Only the indicator "development of entrepreneurial skills" was evaluated worse.

From the general graduates' point of view, they almost always evaluated their competences at a better level than their employer required. The investigation showed better results for the later period, when the level of all acquired competences was, in graduates' evaluations, better than the level of required competences and the difference between levels was higher.

ACKNOWLEDGEMENT

This paper was prepared with institutional support for the conceptual development of science and research at the Faculty of Informatics and Statistics, University of Economics, Prague, project No. IP 400040.

References

ALLEN, J., VAN DERVELDEND, R. *The Flexible Professional in the Knowledge Society: General Results of the REFLEX project* [online]. 2007. [cit. 13.1.2012]. <http://www.fdewb.unimaas.nl/roa/reflex/>.

ARCHER, W., DAVIDSON, J. *Graduate Employability: What do Employers Think and Want?* [online]. London: The Council for Industry and Higher Education, 2008. [cit. 13.1.2012]. <http://ec.europa.eu/education/higher-education/doc1263_en.htm>.

CZSO. *Ženy a muži v datech 2011. Kapitola 3: Vzdělávání* (Women and Men in Data 2011. Chapter 3: Education) [online]. Prague: Czech Statistical Office, 2011. [cit. 13.2.2012]. <http://www.czso.cz/csu/2011edicniplan.nsf/p/1417-11>.

DOUCEK, P., KUNSTOVÁ, R., MARYŠKA, M. Do We Have Enough ICT Specialists in the Period of eDependency? In *Creating Solutions for the Individual, Organisations and Society Proceedings.* Bled: University of Maribor, 2011, pp. 1–17.

EUROPEAN COMMISSION. *The Role of Universities in the Europe of Knowledge, Communication from the Commission, COM/2003/0058* [online]. Brussels: European Commission, 2003. [cit. 13.1.2012]. <http://europa.eu/legislation_summaries/education_training_youth/lifelong_learning/c11067_en.htm>.

HARTNETT, R. T., WILLINGHAM, W. W. *The Criterion Problem: What Measure of Success in Graduate Education?, GRE Board Research Report GREB No. 77-4R* [online]. 1979 [cit. 13.1.2012]. <www.ets.org/Media/Research/pdf/GREB-77-04R.pdf>.

KALOUSKOVÁ, P. *Potřeby zaměstnavatelů a připravenost absolventů škol — šetření v terciární sféře* (Needs of Employers and Readiness of Graduates — a Survey in the Tertiary Sector) [online]. Prague: National Institution of Technical and

Vocational Education (NUOV), 2006. [cit. 13.1.2012].
<http://www.nuov.cz/potreby-zamestnavatelu-a-pripravenost-absolventu-skol>.

KALOUSKOVÁ, P., VOJTĚCH, J. *Potřeby zaměstnavatelů a připravenost absolventů škol — souhrnný pohled* (Needs of Employers and Readiness of Graduates — Summary) [online]. Prague: National Institution of Technical and Vocational Education (NUOV), 2008. [cit. 13.1.2012]. <http://www.nuov.cz/potreby-zamestnavatelu-a-pripravenost-absolventu-skol-2>.

KOUCKÝ, J., LEPIČ, M. *Vývoj kvalifikačních požadavků na pracovním trhu v ČR a v zahraničí* (Development of Skills Requirements in the Labor Market in the CR and Abroad) [online]. Prague: National Institution of Technical and Vocational Education (NUOV), 2008. [cit. 13.1.2012]. <http://www.nuov.cz/vyvoj-kvalifikacnich-pozadavku-na-pracovnim-trhu-v-cr-a-v>.

KOUCKÝ, J., ZELENKA, M. *Postavení vysokoškoláků a uplatnění absolventů vysokých škol na pracovním trhu* (The Position of University Graduates and their Use in the Labor Market) [online]. Prague: Education Policy Centre (EPC), Charles University, 2010. [cit. 13.1.2012]. <http://www.strediskovzdelavacipolitiky.info/>.

MASON, G. The Mix of Graduates and Intermediate-Level Skills in Britain: What Should the Balance Be? *Journal of Education and Work,* 2001, 1, pp. 5–27.

OECD. *Education at a Glance 2010. OECD Indicators* [online]. Paris: OECD, 2010. [cit. 13.1.2012]. <http://www.oecd-ilibrary.org/education/education-at-a-glance-2010_eag-2010-en>.

OECD. *Education at a Glance 2011. OECD Indicators* [online]. Paris: OECD, 2011. [cit. 14.1.2012]. <http://www.oecd-ilibrary.org/education/education-at-a-glance_19991487>.

PECÁKOVÁ, I. *Statistika v terénních průzkumech* (Statistics in Field Surveys). 2nd ed. Prague: Professional Publishing, 2011.

PETERSSON, K. *Graduates from Higher Education in Europe* [online]. 2007. [cit. 13.1.2012]. <http://www.fdewb.unimaas.nl/roa/reflex/documents%20public/publications/REFLEX_Sweden.pdf>.

ŘEZANKOVÁ, H. *Analýza dat z dotazníkových šetření* (Analysis of Data from Surveys). 3rd ed. Prague: Professional Publishing, 2011.

ŘEZANKOVÁ, H., HÚSEK, D., SNÁŠEL, V. *Shluková analýza dat* (Cluster Analysis of Data). 2nd ed. Prague: Professional Publishing, 2009.

UNESCO. *International Standard Classification of Education* [online]. UNESCO, 1997. [cit. 13.1.2012]. <http://www.uis.unesco.org/Education/Pages/international-standard-classification-of-education.aspx>.

The Use of the Sentiment Economic Indicator for GDP Short-term Forecasting: Evidence from EU Economies

Vít Pošta[1] | *Ministry of Finance of the Czech Republic, Prague, Czech Republic*
Zdeněk Pikhart[2] | *Ministry of Finance of the Czech Republic, Prague, Czech Republic*

Abstract

The paper presents a quantitative analysis of the possibilities of Sentiment Economic Indicator based on the joint harmonized EU programme of business and consumer surveys to forecast quarterly GDP growths as a result of the publication lag of the data on GDP. We construct ARMAX models in some cases augmented by the GARCH models to capture the relationship between quarterly changes in GDP and the Sentiment Economic Indicator. The models show some forecasting power of the indicator for approximately half the sample. We show that only for some of the models the forecasting power of the ARMAX / GARCH models actually beats that of a simple ARMA model. We also show that the turbulences in 2007–2008 had a detrimental impact on the relationship between the Indicator and GDP. With the use of the results of rolling forecasts we run a panel regression to test whether or not the forecast errors are dependent on the magnitude of the quarterly changes in GDP. In the applied sample we have found out that the forecasting errors are not dependent on this factor.

Keywords	***JEL code***
Business cycle, forecasting error analysis, short-term GDP forecasting, sentiment economic indicator	*C22, E32, E37*

INTRODUCTION

The paper follows up on the discussion on short-term GDP forecasting, which, as we show shortly, has seen several contributions from the point of view of the Czech economy in the last few years.

Typically, the analyses focus on employing time series of data from the real economy or business and consumer surveys to construct composite indicators, which might hopefully possess the ability to forecast GDP or output gap from a short-term perspective. The issues related to the construction of composite

[1] Letenská 15, 118 10 Prague, Czech Republic. Corresponding author: e-mail: Vit.Posta@mfcr.cz, phone: (+420)257042725.
[2] Letenská 15, 118 10 Prague, Czech Republic.

leading indicators were discussed by Czesaný and Jeřábková (2009a) with the application in Czesaný and Jeřábková (2009b) where GDP is taken as a reference series to capture the cyclical behavior of the economy. Pošta and Valenta (2011) introduced the practice of how composite leading indicators are constructed at the Ministry of Finance where only data from business survey are used. As opposed to Czesaný and Jeřábková (2009b) output gap is used as a reference series. Both papers share a common feature: they do not use the indicators for quantitative output gap forecasting. Svatoň (2011) constructs several composite leading indicators based on data both on real and financial economy and confidence survey. He performs both qualitative and quantitative analysis. Arnoštová et al. (2011) also present a quantitative analysis in an attempt to assess forecasting capacity of several econometric models using especially data on real and financial economy. Benda and Růžička (2007) develop short-term forcasting methods based on the leading indicator approach. They use a set of econometric models (PCA, SURE) that provide estimates of GDP growth for the Czech economy for a co-incident quarter and a few quarters ahead. Their tests show relatively accurate forecasts of GDP fluctuations in the short run. Angelini et al (2008) exploit timely monthly releases of sentiment indicators to compute early estimates of current quarter GDP in the euro area. They also show that survey data and other soft informations are valuable for now-casting GDP.

Adamovicz and Walczyk (2011) examine business cycle in a new EU member states by analysing gross value added and economic sentiment indicator. They observe progressive synchonization of cyclical fluctuations between old and new EU member states. Only higher intensity of cyclical changes has been observed in new EU member.

Gelper and Croux (2009) compare the ESI (further info below) with more sophisticated aggregation schemes that are based on two statistical methods – dynamic factor analysis and partial least squares. The partial least squares method outperforms the other methods, but the ad hoc way of construction ESI can be fully competitive with statistical principles.

Giannone, Reichlin and Simoneli (2009) show that aggregate surveys can produce an accurate early estimate of GDP.

In this paper we focus on the relationship between confidence indicators published by the authorities (the indicators used in the paper will be specified below) and quarterly changes in GDP. The goal of the paper is not a construction of leading indicators in the right sense of the word but rather the examination of the possible use of the fact that the indicators for a given period of time are published sooner than national account data for the same period. It follows that such a publication lag of national accounts behind confidence indicators might be used for forecasting purposes; in this case for a backward forecast – backward in the sense that it is the past which is forecasted, yet unpublished though. We use confidence indicators published at the EU level, therefore, we extend the analysis to basically all EU economies (exceptions due to insufficient data are mentioned below).

We show that in approaximately half the sample it is possible to build a simple model that gives statistically relevant results; i.e. is statistically significant, shows significant forecasting power and stability over time. After the models are built and their forecasting power examined, we investigate the sensitivity of forecasting errors to the magnitude of quarterly changes in GDP. In other words, we examine to which degree the forecasting power of the models is influenced by the instability of the dependent variable. A panel is set up and by means of regression we show that in the sample considered in the paper the sensitivity of forecasting errors to quarterly changes in GDP is statistically insignificant. However, we also estimate the models only up to 2006 and show the forecasting power of the models was higher in most cases.

The paper is divided into three parts: first, the econometrical model and data and its properties are discussed, second, we present the results in the form of the estimated models and their characteristics and also the results of the sensitivity test. Finally, we conclude the key findings.

1 METHODOLOGY AND DATA

1.1 Methodology

To assess the forecasting capacity of the sentiment indicators, simple models are constructed. Each model uses as a starting point a regression between quarterly changes in GDP and the sentiment indicator. As is shown below in the results of the paper, in most cases the diagnostics render the results of such regression tests irrelevant as high autocorrelation between residuals and remaining heteroskedasticity in the residuals are present. As a first step we impose ARMA structure on the original regression model, i.e.:

$$GDP_t = \alpha + \beta IND_t + \varepsilon_t \,, \tag{1}$$

$$\varepsilon_t = \rho\varepsilon_{t-1} + \eta_t + \theta\eta_{t-1}, \tag{2}$$

where *GDP* denotes quarterly changes in GDP, α is an intercept, β is a regression coefficient, *IND* denotes sentiment indicator, ε is the residuals of the regression equation and η is the residuals of the ARMA equation, ρ and θ are coefficients of the ARMA equation. The additional ARMA structure is presented as ARMA (1,1) in (2) as no higher lags were used (see below).

As one can see below some models with the structure described by (1) and (2) still showed remaining heteroskedasticity in the residuals. Therefore we used ARCH / GARCH model:

$$h_t = \gamma + \delta_1\varepsilon_{t-1}^2 + \delta_2 h_{t-1} \,, \tag{3}$$

where h is the variance of the residuals from (1), γ is a constant and $\delta_{1,2}$ are estimated coefficients of the GARCH equation. In most cases the problem with remaining heteroskedasticity is solved by the simple ARCH / GARCH model.

To assess the relation between forecasting errors and magnitude of quarterly changes in GDP we run a panel regression estimated by the two-stage least squares as a special case of instrumental variables regression. We give additional information on this part of analysis below.

1.2 Data

We use Economic Sentiment Indicator (ESI) published by the European Commission as a composite confidence indicator. ESI consists of six particular confidence indicators where each of them is composed of three questions. The result is calculated as a simple arithmetic average of the seasonally adjusted balances to specific questions. European Commission manual (2007) informs that business and consumer surveys provide monthly judgements and anticipations concerning diverse facets of economic activity in the different sectors of the economy. Each sector has explicit weight for ESI final compilation: industry (40% weight), services (30%), construction (20%) and retail trade (5%), as well as consumers (5%). The indicators are further standardized according to their mean level and volatility before aggregation. The process of ESI compilation is further described in the mentioned manual. ESI is available on monthly basis so we create simple quarterly average in order to compare it with quarterly GDP. We consider this way as more accurate than decomposition of quarterly GDP to monthly basis through quadratic polynom or any other mathematical method. ESI series are seasonally adjusted by Danties alghoritm described in the European Commission manual (2007). Quarterly seasonally adjusted GDP series are taken directly from Eurostat.

We use as long time series of ESI and GDP as available. For Belgium, Denmark, Germany, France, Italy, Finland, Netherlands and United Kingdom the data are accessible since 1991Q1. For European Union (27 member states), Bulgaria, Czech Republic, Estonia, Spain, Latvia, Lithuania, Luxembourg, Hungary, Austria, Poland, Portugal, Slovenia, Slovak Republic and Sweden the sample starts between 1993Q1 and 1997Q1. And for Greece and Romania the data are available since 2001Q1. The sample ends in 2011Q3 for all. We do not include Malta and Cyprus in the sample as the series available for these two economies are too short. Ireland is excluded as the indicator is not published for this economy at all.

Table 1 Crosscorrelogram between ESI and quarterly GDP growth

Lag / Economy	AT	BE	BG	CZ	DE	DK	EE	EL	ES	EU
0	0.5050	0.2762	0.2705	0.6561	0.4594	0.3531	0.5288	0.5696	0.8970	0.7346
1	0.2931	0.1240	0.1343	0.4409	0.2485	0.1999	0.3987	0.5304	0.8555	0.4644
2	0.1066	–0.0105	0.0568	0.2012	0.0534	0.0709	0.1900	0.4783	0.7634	0.1836
3	–0.0126	–0.0931	–0.0572	0.0348	–0.0884	–0.0493	0.0361	0.3889	0.6534	–0.0051
4	–0.1276	–0.1568	–0.0383	–0.0426	–0.1827	–0.0833	–0.0869	0.3165	0.5355	–0.1221
5	–0.2552	–0.2771	–0.1038	–0.0837	–0.2176	–0.1671	–0.1157	0.3432	0.4299	–0.1972
6	–0.3517	–0.3169	–0.0703	–0.0843	–0.2153	–0.0055	–0.1428	0.3006	0.3500	–0.2232
7	–0.3787	–0.3018	–0.1125	–0.0765	–0.2012	–0.0955	–0.2479	0.2178	0.2976	–0.2068
8	–0.3309	–0.2691	–0.1306	–0.0603	–0.1325	–0.1035	–0.2937	0.1383	0.2527	–0.1746
9	–0.2553	–0.1946	–0.0990	–0.0328	–0.0918	–0.1293	–0.2688	0.0334	0.2307	–0.1431
10	–0.1965	–0.1293	–0.1243	–0.0193	–0.0714	–0.1407	–0.2151	0.0177	0.1977	–0.0977
Lag / Economy	**FI**	**FR**	**HU**	**IT**	**LT**	**LU**	**LV**	**NL**	**PL**	**PT**
0	0.6063	0.5652	0.7475	0.5141	0.4763	0.2841	0.5067	0.5766	0.3009	0.5727
1	0.4124	0.3100	0.5507	0.3154	0.3289	0.1945	0.4235	0.3776	0.2011	0.3984
2	0.2335	0.0662	0.3862	0.0876	0.1396	0.1204	0.2854	0.1803	0.1276	0.2991
3	0.0768	–0.0889	0.2605	–0.0557	0.0218	–0.0074	0.1551	0.0542	0.1424	0.2100
4	0.0066	–0.161	0.1935	–0.1490	–0.1101	0.0082	0.1041	–0.0152	0.0366	0.1590
5	–0.0514	–0.2005	0.2095	–0.2137	–0.2237	0.0033	–0.0636	–0.1029	–0.0555	0.1077
6	–0.1592	–0.2411	0.2145	–0.1904	–0.2303	0.0265	–0.1190	–0.1395	–0.0321	0.1590
7	–0.1082	–0.2363	0.254	–0.1655	–0.2002	0.0533	–0.1735	–0.1715	0.0982	0.1734
8	–0.0890	–0.2204	0.2187	–0.1333	–0.2307	–0.0002	–0.1787	–0.2076	–0.0245	0.1476
9	–0.0901	–0.1621	0.1721	–0.0865	–0.2474	0.0633	–0.1761	–0.2142	–0.0284	0.1433
10	–0.0539	–0.1201	0.0963	–0.0856	–0.1914	–0.0052	–0.1733	–0.2017	–0.0273	0.0976
Lag / Economy	**RO**	**SE**	**SI**	**SK**	**UK**					
0	0.6906	0.4646	0.4230	0.4839	0.6015					
1	0.5165	0.2724	0.1913	0.3656	0.4106					
2	0.3287	0.0455	0.0162	0.1368	0.2158					
3	0.2385	–0.1668	0.0261	0.0807	0.0516					
4	0.1283	–0.3114	–0.0604	0.0658	–0.0498					
5	0.0656	–0.3904	–0.1179	0.0603	–0.1389					
6	–0.0226	–0.3791	–0.1470	–0.0711	–0.1505					
7	–0.1079	–0.3242	–0.1303	–0.0343	–0.0950					
8	–0.1390	–0.2465	–0.0835	–0.0202	–0.0418					
9	–0.1532	–0.2254	–0.1067	–0.0467	–0.0073					
10	–0.0945	–0.1986	–0.1391	0.0014	0.0525					

Notes: AT – Austria, BE – Belgium, BG – Bulgaria, CZ – Czech Republic, DE – Germany, DK – Denmark, EE – Estonia, EL – Greece, ES – Spain, EU – EU27, FI – Finland, FR – France, HU – Hungary, IT – Italy, LT – Lithuania, LU – Luxembourg, LV – Latvia, NL – Netherlands, PL – Poland, PT – Portugal, RO – Romania, SE – Sweden, SI – Slovenia, SK – Slovak Republic, UK – United Kingdom.
Source: Own construction

Table 1 shows crosscorrelogram between ESI and quarterly GDP growth. The goal is to capture correlation of lagging values of ESI and quarterly GDP growth. The correlation between the first lagged value of ESI and quarterly GDP growth is in most countries weaker than correlations at zero lag. Only in case of Spain there is a significant correlation at the first lag. Correlation on further lags generally decline steeply. This implies that the ESI should not be considered as a leading indicator with respect to the reference series, but just for publication lead estimation. Simply we try to estimate last unpublished quarterly GDP due to three month lead of ESI against the release of GDP figures. For instance at the end of March we are able to estimate the first quarter of the respective year.

Table 2 Descriptive statistics for quarterly changes in GDP and ESI

Economy	AT		BE		BG		CZ		DE	
Variable	**GDP**	**Indicator**	**GDP**	**Indicator**	**GDP**	**Indicator**	**GDP**	**Indicator**	**GDP**	**Indicator**
Mean	0.50334	99.38849	0.60626	99.85992	1.0932	101.0267	0.6606	99.97846	0.32185	99.64008
St. dev.	0.174	9.44644	1.44761	9.90138	2.62675	8.62944	0.92579	10.27073	0.86046	9.45436
JB	3.81443	5.73540*	7968.853***	6.48908***	243.0699***	2.28921	65.93284***	6.31443**	193.0991***	4.40877
ADF	−4.54643***	−5.17664***	−7.60870***	−4.70014***	−9.80467***	−2.54299	−3.27821**	−2.72152*	−6.65010***	−4.75715***
Economy	**DK**		**EE**		**EL**		**ES**		**EU**	
Variable	**GDP**	**Indicator**	**GDP**	**Indicator**	**GDP**	**Indicator**	**GDP**	**Indicator**	**GDP**	**Indicator**
Mean	0.39822	100.1151	1.24972	101.9961	0.46612	96.3625	0.63788	101.725	0.46342	101.775
St. dev.	1.27505	10.30605	2.32503	9.08649	1.24811	14.27416	0.60867	9.11274	0.59362	9.12546
JB	0.51448	10.09949***	108.3591***	11.96403	1.17308	4.08061	48.64879	16.08724***	430.9949***	35.83318***
ADF	−10.50847***	−3.49219**	−4.79861***	−2.62703*	−4.52401***	−2.93646	−2.69189*	−3.36834*	−3.28124**	−4.62151***
Economy	**FI**		**FR**		**HU**		**IT**		**LT**	
Variable	**GDP**	**Indicator**	**GDP**	**Indicator**	**GDP**	**Indicator**	**GDP**	**Indicator**	**GDP**	**Indicator**
Mean	0.57304	100.5909	0.39494	99.99127	0.53862	100.3776	0.22818	99.62302	1.29555	101.3358
St. dev.	1.27522	9.70815	0.49898	9.91404	0.86687	10.77312	0.67291	9.84491	2.52831	9.31165
JB	275.1786	7.55500**	71.18633***	4.47011	132.3869***	40.55382***	212.9857***	2.09329	770.3447***	6.26799**
ADF	−6.18702***	−5.20137***	−4.77740***	−4.35765***	−3.37641**	−3.04019**	−5.17599***	−3.72686***	−6.61743***	−2.85700*
Economy	**LU**		**LV**		**NL**		**PL**		**PT**	
Variable	**GDP**	**Indicator**	**GDP**	**Indicator**	**GDP**	**Indicator**	**GDP**	**Indicator**	**GDP**	**Indicator**
Mean	0.91797	100.9564	1.14458	101.5216	0.55382	99.74683	1.10456	99.02402	0.41026	99.05539
St. dev.	1.94012	9.82075	2.64071	9.07031	0.66733	9.99059	1.07451	9.72421	0.8619	10.62156
JB	15.69461***	1.7023	140.1597***	8.54669**	54.54470***	4.54595	200.5571***	2.23399	1.41433	2.96888
ADF	−9.76718***	−4.58983***	−3.18975**	−2.91196**	−5.30342***	−3.79204***	−8.99316***	−2.78849*	−6.15051***	−3.27243*
Economy	**RO**		**SE**		**SI**		**SK**		**UK**	
Variable	**GDP**	**Indicator**	**GDP**	**Indicator**	**GDP**	**Indicator**	**GDP**	**Indicator**	**GDP**	**Indicator**
Mean	0.99218	100.4879	0.70053	102.7693	0.82641	99.7607	1.0147	100.3589	0.55143	100.219
St. dev.	1.23958	9.34202	0.93836	8.41717	1.32223	9.54855	1.96844	9.45547	0.68068	9.70783
JB	18.60828***	6.51712**	358.0559***	2.08377	308.6994***	14.84725***	251.2445***	25.59141***	148.3592***	27.39304***
ADF	−3.15157**	−3.39622*	−5.70936***	−5.76801***	−5.22250***	−3.61689***	−8.21872***	−3.02248**	−3.42824**	−3.996512***

Note: JB is Jarque-Bera statistic under the null of normal distribution. ADF is augmented Dickey-Fuller statistic under the null of unit root. (*, **, *** denote rejection of the null at 10%, 5% and 1% level of significance, respectively).
Source: Own construction

We are not able to run the analysis with flash estimates of GDP as that would require having the actual series of GDP for each flash available. However, the older "versions" of GDP series are not published by Eurostat or national statistical offices. It would not make sense to use flash estimates of GDP together with current GDP series.

Descriptive statistics for quarterly changes in GDP and ESI are reported in Table 2. It shows mean value, standard deviation, Jarque-Bera test of normality of the distribution and augmented Dickey-Fuller unit root test. The series may be considered stationary, which is important information for further analysis.

Table 3 Crosscorrelogram between ESI and quarterly GDP growth

Economy	AT	BE	BG	CZ	DE	DK	EE	EL	ES	EU
C	−1.26698 (0.8901)	−3.388918** (1.56185)	−6.41191** (2.52063)	−3.46152** (1.61983)	−2.50689*** (0.78382)	−4.23072*** (1.14555)	−12.38281*** (1.99813)	−4.90125*** (1.20542)	−5.46661*** (0.41003)	−2.24381* (1.20465)
Indicator	0.01776** (0.00891)	0.03997** (0.01555)	0.07386*** (0.02481)	0.04146*** (0.01504)	0.02910*** (0.00779)	0.04585*** (0.01131)	0.13299*** (0.02141)	0.05497*** (0.01224)	0.06006*** (0.00392)	0.02736** (0.01155)
AR(1)	0.30245** (0.12077)	x	x	0.76575*** (0.12762)	x	−0.25333** (0.10875)	0.23567 (0.14506)	x	0.23292 (0.14576)	0.35333 (0.22360)
MA(1)	0.93069*** (0.037661)	x	−0.37008*** (0.12437)	x	x	x	x	x	x	x
C	x	x	x	0.11292*** (0.02621)	0.31467*** (0.7300)	x	2.71110*** (0.17345)	x	0.05388*** (0.01634)	0.06746*** (0.13168)
ARCH(1)	x	x	x	0.69929*** (0.25594)	0.44078** (0.17394)	x	−0.03035*** (0.01070)	x	0.18619 (0.23694)	0.39433** (0.16531)
GARCH(1)	x	x	x	x	x	x	x	x	x	x
R−sq	0.7019	0.076297	0.187455	0.559372	0.185447	0.191255	0.32727	0.324395	0.81749	0.586497
F−stat	60.4349***	6.607926**	6.344299***	17.77289***	5.91936***	9.222893***	7.29703***	20.16652***	67.18626***	21.27542***
AIC	0.80661	3.534856	4.647843	1.547553	2.205431	3.182087	4.22863	2.956912	0.26	0.673153
JB	0.46563*	8605.571***	348.6318***	1.45999	0.58369	1.45968	223.1742***	5.32046*	1.50437	2.97913
ARCH LM	3.00906*	0.01267	0.8091	0.4812	0.57037	0.41473	0.10408	0.03499	0.15266	0.0065
Q−stat	0.2913	1.0428	0.1379	1.412	0.8286	0.005	0.2074	0.0176	0.227	0.9993
MRSE	0.34472	1.38279	2.34728	0.61405	0.77184	1.14461	1.90503	1.01416	0.25953	0.38167
TC	0.22434	0.60484	0.54656	0.28008	0.55697	0.5733	0.42583	0.46903	0.1497	0.27595
MRSE (ARMA)	0.34738	1.42915	2.51385	0.67443	0.81978	1.24519	2.05482	1.05621	0.31015	0.41984
TC (ARMA)	0.22659	0.64602	0.61331	0.32929	0.61662	0.69463	0.47660	0.51866	0.18132	0.30497
MRSE (2006)	0.29565	1.42915	1.94946	0.34886	0.63595	1.10807	1.69540	0.95395	0.26392	0.23811
TC (2006)	0.21487	0.64602	0.35778	0.16333	0.54908	0.50903	0.39923	0.39822	0.14169	0.17943

Note: Sample ends in 2011Q3 and starts according to the information given in part 2. Dependent variable: quarterly changes in GDP. Independent variable: sentiment indicator ESI (denoted as indicator). C denotes a constant in the regression or GARCH specification. AR(1) stands for autoregressive term with 1 lag. MA(1) stands for moving average term with 1 lag. ARCH(1) stands for squared residuals from the regression delayed by 1 lag. GARCH(1) stands for variance of the residuals from the regression delayed by 1 lag. Estimates of the coefficients with standard errors in parenthesis are given. R-sq denotes the coefficient of determination. F-stat is a statistic of an F-test under the null of slope coefficients equal to 0. AIC is the value of Akaike information criterion. JB is Jarque-Berra statistic for the residuals under the null of normal distribution. ARCH LM test is the Engle's LM statistic under the null of no remaining ARCH in the residuals. Q-stat is the Ljung-Box statistic under the null of no autocorrelation of the residuals. MRSE is a root mean square error. TC is Theil inequality coefficient. MRSE (ARMA) and TC(ARMA) give the statistics for simple ARMA models used as a baseline for comparison with the ARMAX models. MRSE (ARMA) and TC(ARMA) give the statistics for models estimated up to 2006Q4 and used for forecast the GDP up to 2006Q4. (*, **, *** denote rejection of the null at 10%, 5% and 1% level of significance, respectively).
Source: Own construction

2 RESULTS

First, we present the estimated models used for the fore-WWcasting exercise. Referring back to part 1 of the paper, a simple regression model with quarterly changes in GDP as the dependent and ESI as the independent variable was estimated for each economy. Based on the analysis of the residuals, ARMA structure was imposed or, further, ARCH / GARCH specification was used to meet the standard conditions for the behavior of the residuals. Tables 3 to 5 give the results.

The ARCH LM test and Q statistic are given for the first relevant lag of the residuals. We checked the remaining autocorrelation and ARCH up to 5 additional lags, but we do not report the results here.

We were able to use the basic regression model only in four cases: Belgium, Greece, Lithuania and Portugal. Even after the complete procedure we could still find some remaining ARCH in some cases: Austria, Hungary, Portugal (here applying ARMA or GARCH structure did not result in well-behaved residuals) and Romania. In some cases we did not obtain normal residuals.

Table 4 Model Output

Economy	FI	FR	HU	IT	LT	LU	LV	NL	PL	PT
C	−5.24922*** (1.04907)	−2.71905*** (0.68274)	−4.30388*** (0.81815)	−2.37217** (0.97666)	−11.6349*** (2.99716)	−5.79568*** (1.82260)	−19.9478*** (1.97156)	−3.68076*** (0.74801)	−2.04585** (0.85174)	−4.36442*** (0.85875)
Indicator	0.05884*** (0.01031)	0.03092*** (0.00677)	0.04859*** (0.008215)	0.02647*** (0.00979)	0.12771*** (0.02947)	0.06700*** (0.01805)	0.20281*** (0.02056)	0.04118*** (0.00761)	0.03193*** (0.00818)	0.04807*** (0.00860)
AR(1)	x	0.40409*** (0.10621)	0.39961*** (0.13618)	0.38126** (0.16468)	x	−0.31899*** (0.11749)	x	0.36097*** (0.10967)	x	x
MA(1)	x	x	x	x	x	x	x	x	x	x
C	0.57549*** (0.12119)	x	0.00223** (0.00098)	2.25415*** (0.08394)	x	x	2.66883*** (0.51945)	0.00546*** (0.00042)	1.69912*** (0.10234)	x
ARCH(1)	0.45276*** (0.15036)	x	−0.06372*** (0.00481)	x	x	x	0.53861* (0.27527)	−0.08649*** (0.01554)	0.30418*** (0.03772)	x
GARCH(1)	x	x	1.08570*** (0.00141)	−0.32734 (0.21368)	x	x	x	1.05616*** (0.01875)	−1.00808*** (0.01401)	x
R–sq	0.33151	0.451095	0.622797	0.341137	0.226841	0.202123	0.184983	0.379109	0.09025	0.32801
F–stat	12.89365***	32.05061***	18.49226***	7.766487***	18.77729***	7.853087***	4.690663***	9.158839***	1.512842	31.2395***
AIC	2.767692	0.911874	1.510904	1.580785	4.481044	4.013098	4.483328	1.386499	2.534005	2.188474
JB	3.99920	3.30805	2.35082	0.28319	607.131***	3.26329	19.1743***	2.27187	112.8966***	1.10558
ARCH LM	1.99664	0.09952	2.76048*	0.34943	0.14655	1.93726	0.95544	0.54316	0.26389	2.964639*
Q–stat	0.4784	0.7504	0.2766	0.0814	0.1507	0.0051	0.0778	0.0123	0.2871	0.3219
MRSE	1.03539	0.36787	0.52788	0.54165	2.20622	1.71851	2.36586	0.52270	1.01708	0.70117
TC	0.44817	0.31432	0.27472	0.45949	0.48104	0.50537	0.49119	0.34043	0.37836	0.44165
MRSE (ARMA)	1.14451	0.40666	0.45578	0.57669	2.45599	1.85853	2.40969	0.58248	1.02449	0.65853
TC (ARMA)	0.52646	0.36470	0.23144	0.51499	0.58079	0.58981	0.55162	0.39015	0.38179	0.42418
MRSE (2006)	0.77169	0.30652	0.33222	0.47206	1.85347	1.74527	1.89518	0.47530	1.09894	0.70613
TC (2006)	0.36692	0.26114	0.17146	0.43110	0.41892	0.44533	0.42364	0.30989	0.37916	0.41332

Note: Sample ends in 2011Q3 and starts according to the information given in part 2. Dependent variable: quarterly changes in GDP. Independent variable: sentiment indicator ESI (denoted as indicator). C denotes a constant in the regression or GARCH specification. AR(1) stands for autoregressive term with 1 lag. MA(1) stands for moving average term with 1 lag. ARCH(1) stands for squared residuals from the regression delayed by 1 lag. GARCH(1) stands for variance of the residuals from the regression delayed by 1 lag. Estimates of the coefficients with standard errors in parenthesis are given. R-sq denotes the coefficient of determination. F-stat is a statistic of an F-test under the null of slope coefficients equal to 0. AIC is the value of Akaike information criterion. JB is Jarque-Berra statistic for the residuals under the null of normal distribution. ARCH LM test is the Engle's LM statistic under the null of no remaining ARCH in the residuals. Q-stat is the Ljung-Box statistic under the null of no autocorrelation of the residuals. MRSE is a root mean square error. TC is Theil inequality coefficient. MRSE (ARMA) and TC(ARMA) give the statistics for simple ARMA models used as a baseline for comparison with the ARMAX models. MRSE (ARMA) and TC(ARMA) give the statistics for models estimated up to 2006Q4 and used for forecast the GDP up to 2006Q4. (*, **, *** denote rejection of the null at 10%, 5% and 1% level of significance, respectively).

Source: Own construction

Table 5 Model Output

Economy	RO	SE	SI	SK	UK
C	−7.42885*** (2.73306)	−3.55915*** (1.12650)	−7.69975*** (1.63584)	−8.58374*** (1.13051)	−2.02008** (0.98265)
Indicator	0.08364*** (0.02404)	0.04284*** (0.01076)	0.08333*** (0.01659)	0.095521*** (0.01120)	0.02668*** (0.00961)
AR(1)	0.47689*** (0.16605)	x	0.25803* (0.13192)	x	x
MA(1)	x	x	x	-0.56334*** (0.11203)	0.48370*** (0.09113)
C	x	0.06518 (0.08144)	0.17703*** (0.06470)	x	0.11180*** (0.03871)
ARCH(1)	x	0.31997 (0.22946)	0.13463 (0.09410)	x	0.444902* (0.22837)
GARCH(1)	x	0.65991*** (0.25148)	0.64937*** (0.07686)	x	x
R-sq	0.586361	0.180101	0.210332	0.413086	0.526823
F-stat	27.6426***	3.789179***	3.142988**	19.35522***	21.43243***
AIC	2.584985	2.436198	3.08135	3.745541	1.203095
JB	1.31397	35.9437***	32.44813***	51.08552***	1.64484
ARCH LM	4.86597**	0.37678	0.78657	0.294868	0.00026
Q-stat	0.0131	0.003	0.3694	0.0063	0.5124
MRSE	0.82047	0.84391	1.17479	1.49497	0.46536
TC	0.27219	0.40372	0.4417	0.3998	0.28917
MRSE (ARMA)	0.87125	0.86054	1.21811	1.95140	0.43776
TC (ARMA)	0.30006	0.43898	0.47911	0.60626	0.26745
MRSE (2006)	0.59859	0.56677	0.85604	0.96861	0.33376
TC (2006)	0.19022	0.31682	0.33418	0.27179	0.21502

Note: Sample ends in 2011Q3 and starts according to the information given in part 2. Dependent variable: quarterly changes in GDP. Independent variable: sentiment indicator ESI (denoted as indicator). C denotes a constant in the regression or GARCH specification. AR(1) stands for autoregressive term with 1 lag. MA(1) stands for moving average term with 1 lag. ARCH(1) stands for squared residuals from the regression delayed by 1 lag. GARCH(1) stands for variance of the residuals from the regression delayed by 1 lag. Estimates of the coefficients with standard errors in parenthesis are given. R-sq denotes the coefficient of determination. F-stat is a statistic of an F-test under the null of slope coefficients equal to 0. AIC is the value of Akaike information criterion. JB is Jarque-Berra statistic for the residuals under the null of normal distribution. ARCH LM test is the Engle's LM statistic under the null of no remaining ARCH in the residuals. Q-stat is the Ljung-Box statistic under the null of no autocorrelation of the residuals. MRSE is a root mean square error. TC is Theil inequality coefficient. MRSE (ARMA) and TC(ARMA) give the statistics for simple ARMA models used as a baseline for comparison with the ARMAX models. MRSE (ARMA) and TC(ARMA) give the statistics for models estimated up to 2006Q4 and used for forecast the GDP up to 2006Q4. (*, **, *** denote rejection of the null at 10%, 5% and 1% level of significance, respectively).
Source: Own construction

If we take as an arbitrary benchmark value of the coefficient of determination 50%, we see that 7 models meet such a condition: Austria, Czech Republic, Spain, EU27, Hungary, Romania and United Kingdom. However, taking the coefficient of determination as a sole measure of fit is very misleading as it provides no information on the fit in levels. When taking as an arbitrary benchmark value of Theil inequality coefficient 0.4 (which normalizes root mean square error by the sum of the roots of the mean squared values of forecast and actual values of the variable), we obtain 11 satisfactory models: Austria, Czech Republic, Spain, EU27, France, Hungary, Netherlands, Poland, Romania, Slovak Republic and United Kingdom.

In the Annex we present graphical output for the whole sample of economies which compares the actuals with the forecast.

In the next step of the analysis we estimated simple ARMA models for each economy up to 2011Q2 and used it for forecast up to 2011Q3. This serves as a baseline forecast to which the forecast from the ARMAX / GARCH models may be compared.

From Tables 3 to 5 one can see that only in three cases does the simple ARMA model produces better results than the ARMAX / GARCH model. On the other hand, it should be noted that the increase in forecasting power due to ARMAX / GARCH (as compared with simple ARMA) is rather negligible in many cases.

We further estimated the ARMAX / GARCH models only up to 2006Q4 to exclude the effect of the turbulences between 2007 and 2008. Then we used the estimates to produce forecasts up to 2006Q4. The results in the form of MRSE and Theil coefficients are presented in Tables 3 to 5. We stress that the models are not directly comparable in some cases as the ARMA (GARCH) structure needed to be altered for the significantly shorter data sample. We do not present the exact specifications of the "2006" models in the paper. It should be noted that in many cases the model produces much better results than the original one. Thus the turbulences between 2007 and 2008 seem to have a rather strong negative impact on the relationship between ESI and GDP.

To evaluate the models further, we ran a panel regression between the forecast errors and absolute values of quarterly changes in GDP to check the sensitivity of the forecasts to the speed with which the dependent variable changes.

We ran a rolling forecast from 2009Q1 to 2011Q3 to obtain the forecast errors, e.g. by forecast for 2009Q1 we mean that the model was estimated up to 2008Q4 (which means that data for the sentiment indicator, ESI, were available for 2009Q1 at that time) and based on the estimation we forecast the GDP growth for 2009Q1. By comparing the forecast for the particular quarter with the actual quarterly growth of GDP in that quarter, we obtained the forecast errors.

To run such an exercise, it is crucial to set the starting quarter of the rolling procedure. We start the forecast in 2009Q1 as most estimated models as presented in Tables 3 to 5 exhibited significant stability back to that period. By choosing 2009Q1 as a starting point for the exercise, we cut the cross-sample down to 20 economies, i.e. we leave out Greece, Latvia, Poland, Sweden and Slovenia whose models were highly unstable. Next we cut the cross-sample down even more by taking account of the fit of the forecasts, i.e. we apply the arbitrary rule used above – Theil inequality coefficient lower than 0.4. The resulting sample consists of 10 economies: Austria, Czech Republic, Spain, EU27, France, Hungary, Netherlands, Romania, Slovak Republic and United Kingdom.

Table 6 Panel regression

	Panel 1	Panel 2
C	0,46423*** (0,14920)	0,42434*** (0,15102)
GDP	0,16311 (0,18672)	-0,00890 (0,14983)
AR(1)	0,31879*** (0,06170)	0,54590*** (0,08474)
R-sq	0,43	0,56
F-stat	5,85	10,24
DW	2,02	2,05
JB	3411,722***	1601,068***

Note: Sample runs from 2009Q1 to 2011Q3 and across 20 economies in Panel 1 and 10 economies in Panel 2 as described above. Dependent variable: forecast errors based on the rolling forecasts. Independent variable: absolute value of quarterly changes in GDP. Other notation corresponds to that used earlier. (*, **, *** denote rejection of the null at 10%, 5% and 1% level of significance, respectively).
Source: Own construction

Table 6 gives the results of the panel regression. Two-stage least squares were used to obtain the estimates, with a constant and lagged values of independent variable as instruments. Autoregressive term was used to obtain serially uncorrelated residuals. In both cases the forecast errors come out as independent of the absolute value of quarterly changes in GDP.

CONCLUSION

The assessment of the so-called soft indicators as sentiment and confidence survey indicators has become increasingly popular in recent years. This paper presents one of many ways how confidence indicators might be useful for forecasting development of GDP. We used Economic Sentiment Indicator built and published by the European Commission.

First, we constructed a regression model augmented by the ARMA and ARCH / GARCH structure in some cases to capture the relationship between quarterly changes in GDP and ESI. It turned out that the regression models had some forecasting power in roughly half the sample. This showed that universal use of the data cannot be expected.

To assess the forecasting power in more detail we created simple ARMA models for each case and used to produce GDP forecasts. The quality of these forecasts was compared to the quality of the ARMAX forecasts (were ESI is used). In most cases the ARMAX forecasts beat the underlying ARMA forecasts although the difference in quality is rather negligible in many cases.

To capture the effect of the turbulences, which roughly took place between 2007 and 2008, on the forecasting power of the model, we estimated the original (ARMAX) model only up to the fourth quarter of 2006. Then we compared the quality of the forecasts of such a model with that of the original version, which was used for the whole sample. It was shown that the forecasting capacity of the ARMAX / GARCH model was negatively influenced by the turbulences.

Finally, we ran a rolling forecast exercise from 2009Q1 to 2011Q3 and compared the forecast with the actual measured quarterly GDP growth. The sample was divided into two groups according to the stability of the models and Theil inequality coefficient. The first group included 20 countries and the second had 10 members. We conducted a panel regression test between the forecast errors and quarterly GDP changes in absolute value to check the sensitivity of the errors on the variability of the forecasted series. In both cases the forecast errors came out as independent of the absolute values of quarterly changes in GDP. Therefore, it seems that the relationship between ESI and GDP may be exploited in relatively peaceful times while the relationship may be quite distorted when an economy is hit by unexpected shocks.

References

ANGELINI, E., CAMBA-MÉNDEZ, G., GIANNONE, D., RÜNSTLER, G., REICHLIN, L. *Short-term Forecasts of Economic Activity in the Euro Area* [online]. Frankfurt am Main: European Central Bank, 2008. [cit. 5.3.2012]. <http://www.ecb.int/pub/pdf/scpwps/ecbwp949.pdf>.

ADAMOVICZ, E., WALCZYK, K. *Cyclical Fluctuations of Business Activity in the New EU Countries* [online]. Warsaw: Warsaw School of Economics – Research Institute for Economic Development, 2011. [cit. 5.3.2012]. <https://www.ciret.org/files/workshops/moscow/papers/Walczyk_Konrad.pdf>.

ARNOŠTOVÁ, K., HAVRLANT, D., RŮŽIČKA, L., TÓTH, P. Short-Term Forecasting of Czech Quarterly GDP Using Monthly Indicators. In *Finance a Úvěr – Czech Journal of Economics and Finance,* 2011, 6, pp. 566–583.

BENDA, V., RŮŽIČKA, L. *Short-term Forecasting Methods Based on the LEI Approach:* The Case of the Czech Republic [online]. Prague: Czech National Bank, 2007. [cit. 5.3.2012]. <http://www.cnb.cz/miranda2/export/sites/www.cnb.cz/en/research/research_publications/irpn/download/rpn_1_2007.pdf>.

CZESANÝ, S., JEŘÁBKOVÁ, Z. Metoda konstrukce kompozitních indikátorů hospodářského cyklu pro českou ekonomiku (The Method of Construction of Composite Business Cycle Indicators for the Czech Economy). In *Statistika*, 2009, 1, pp. 21–31.

CZESANÝ, S., JEŘÁBKOVÁ, Z. Kompozitní indikátory hospodářského cyklu české ekonomiky (Composite Indicators of the Economic Cycle of the Czech Economy). In *Statistika*. 2009, 3, pp. 256–274.

EC. *A User Manual to the Joint Harmonised EU Programme of Business and Consumer Surveys* [online]. Brussels: European Commission, Directorate for Economic and Financial Affairs, 2007. [cit. 5.3.2012]. <http://ec.europa.eu/economy_finance/db_indicators/surveys/documents/userguide_en.pdf>

GELPER, S., CROUX, C. *On the Construction of the European Economic Sentiment Indicator* [online]. Leuven: Katholieke Universiteit Leuven, 2010. [cit. 5.3.2012]. <http://www.econ.kuleuven.be/public/ndbaf40/esi.pdf>.

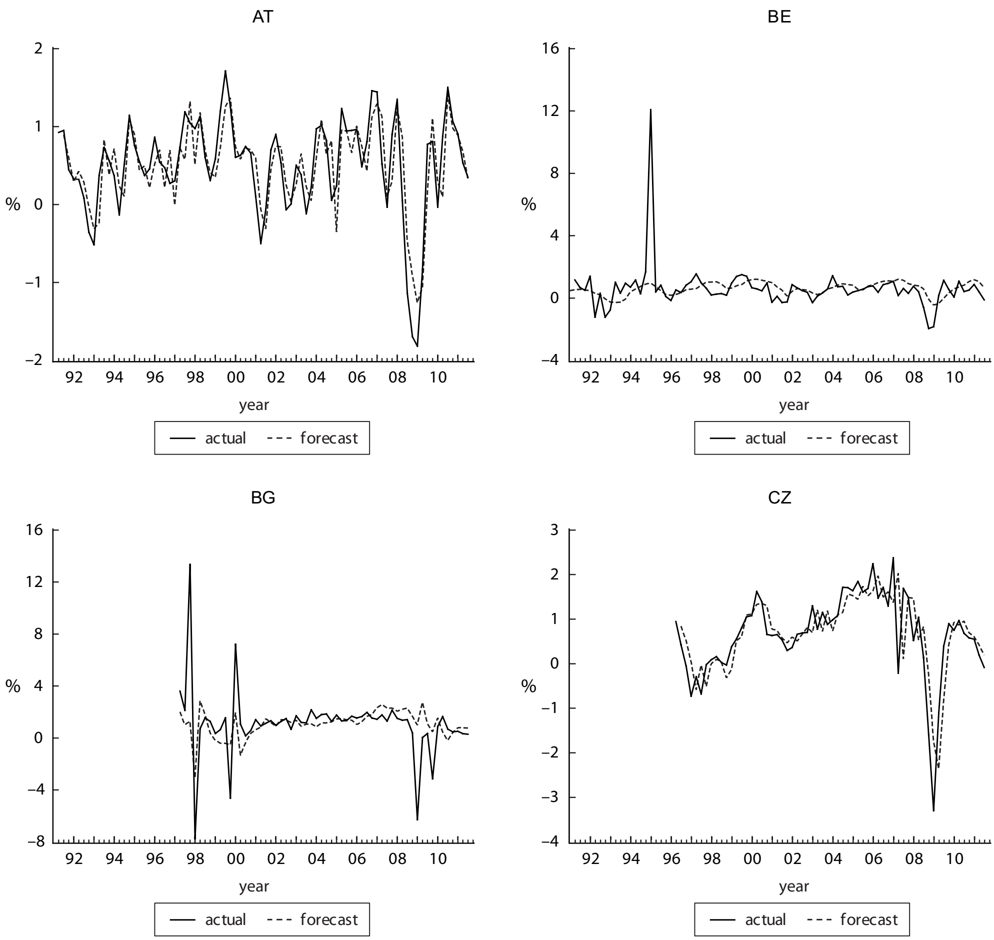
GIANNONE, D., REICHLIN, L, SIMONELLI, S. Nowcasting Euro Area Economic Activity in Real-Time: The Role of Confidence Indicator. *National Institute Economic Review,* 2009, 210, pp. 90–97.

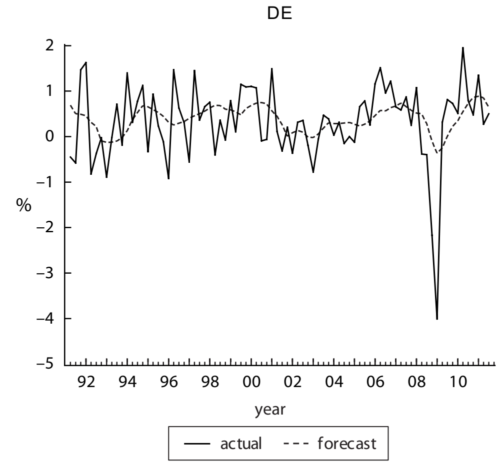
POŠTA, V., VALENTA, V. Composite Leading Indicators Based on Business Survey: Case of the Czech Economy. In *Statistika, Economy and Statistics Journal,* 2011, 1, pp. 12–18.

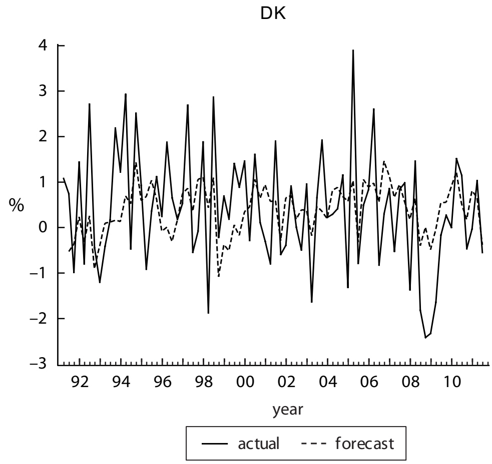
SVATOŇ, P. *Composite Leading Indicators: A Contribution to the Analysis of the Czech Business Cycle* [online]. Prague: Ministry of Finance of the Czech Republic, 2011. <http://www.mfcr.cz/cps/rde/xbcr/mfcr/VS_COMPOSITE_LEADING_INDICATORS-082011.pdf>.

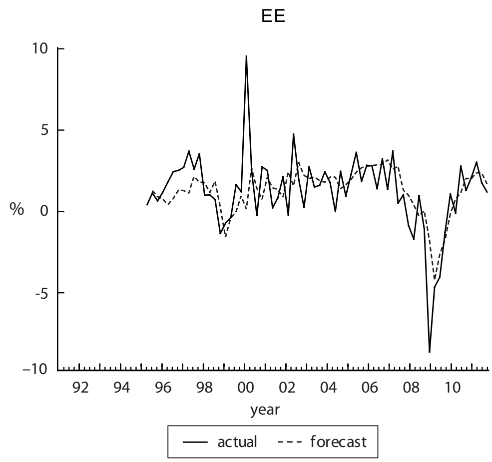
ANNEX

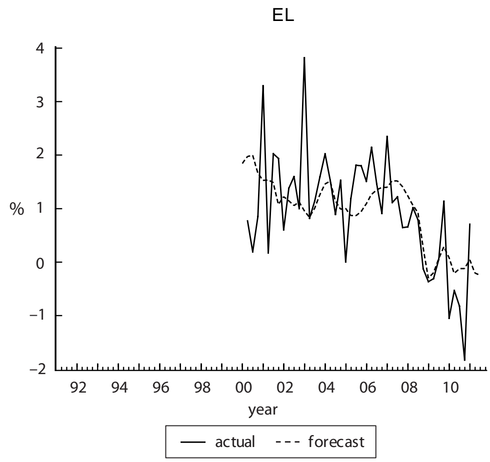
Figures 1 to 25 Actual quarterly changes in GDP vs forecast based on estimation up to 2011Q2 and forecast up to 2011Q3

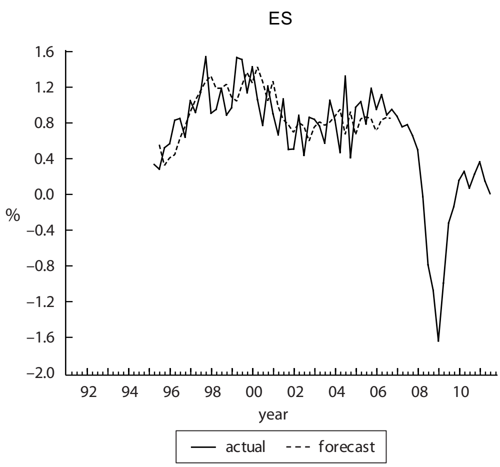



DE
2
1
0
−1
%
−2
−3
−4
−5
92 94 96 98 00 02 04 06 08 10
year
actual
forecast

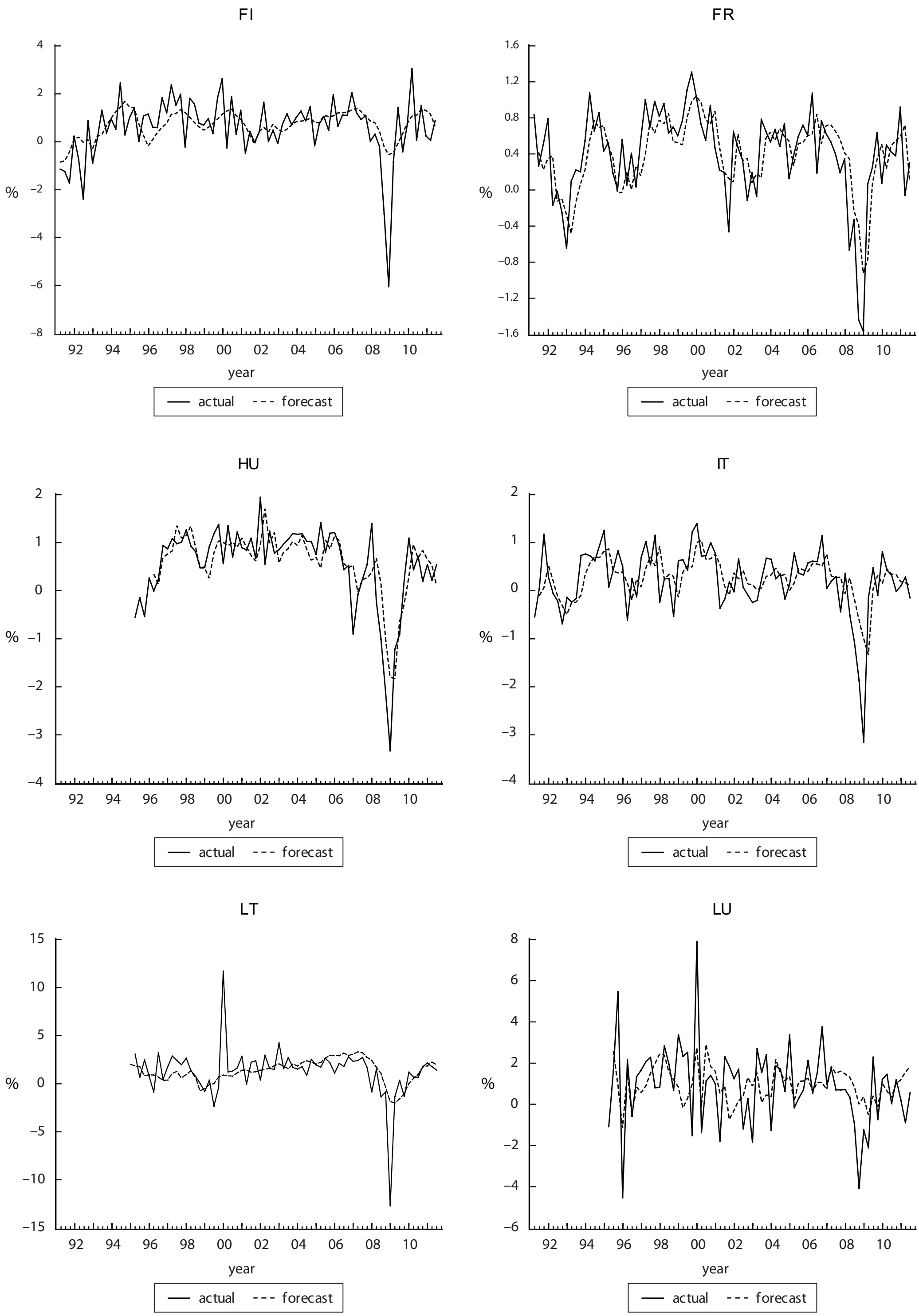

DK
4
3
2
1
%
0
−1
−2
−3
92 94 96 98 00 02 04 06 08 10
year
actual
forecast

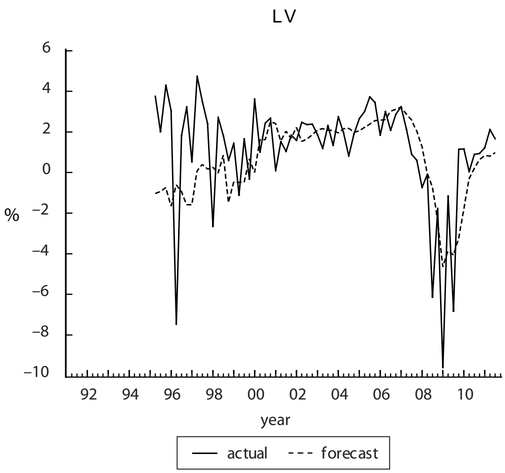

EE
10
5
% 0
-5
−10
92 94 96 98 00 02 04 06 08 10
year
actual
forecast

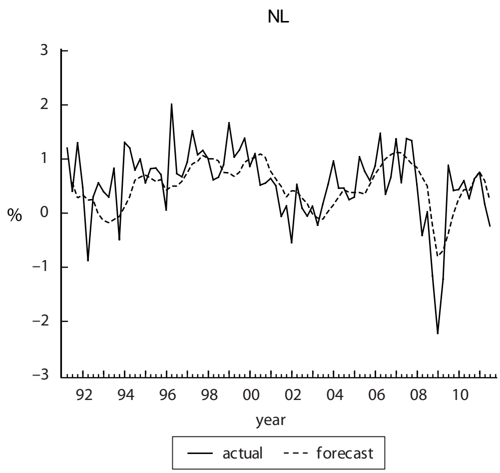

EL
4
3
2
% 1
0
−1
−2
92 94 96 98 00 02 04 06 08 10
year
actual
forecast

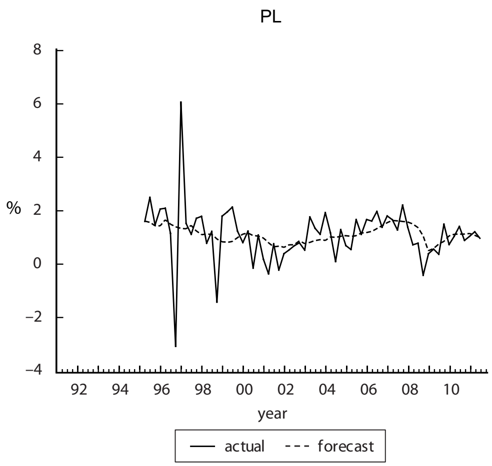

ES
1.6
1.2
0.8
0.4
0.0
%
−0.4
−0.8
−1.2
−1.6
−2.0
92 94 96 98 00 02 04 06 08 10
year
actual
forecast

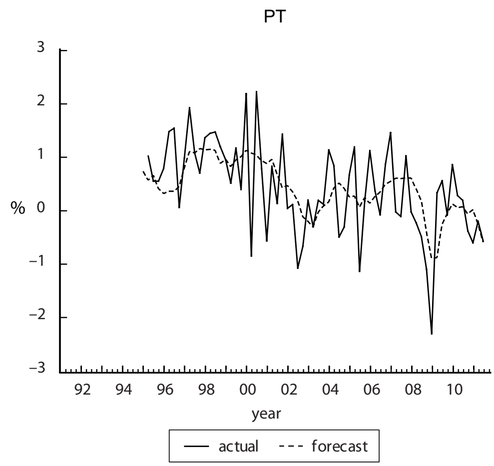

EU
2
1
0
%
−1
−2
−3
92 94 96 98 00 02 04 06 08 10
year
actual
forecast

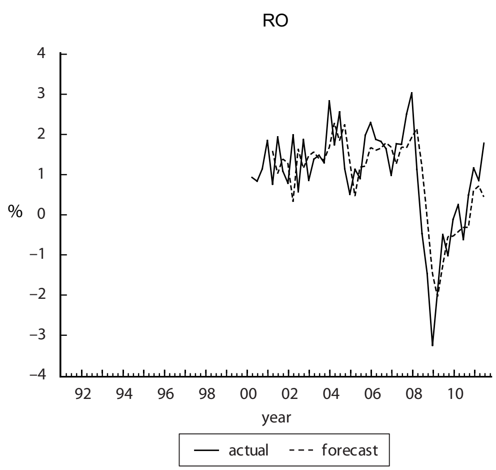

FI
%
year
actual
forecast
FR
%
year
actual
forecast
HU
%
year
actual
forecast
IT
%
year
actual
forecast
LT
%
year
actual
forecast
LU
%
year
actual
forecast

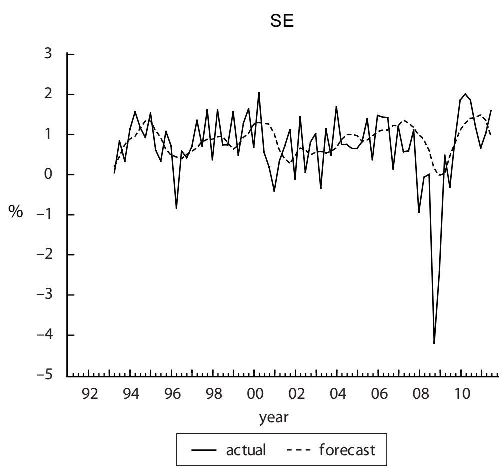

LV
6
4
2
0
%
−2
−4
−6
−8
−10
92 94 96 98 00 02 04 06 08 10
year
actual
forecast

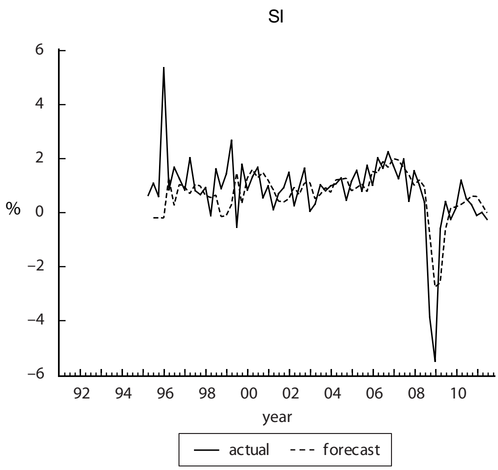

NL
3
2
1
%
0
−1
−2
−3
92 94 96 98 00 02 04 06 08 10
year
actual
forecast

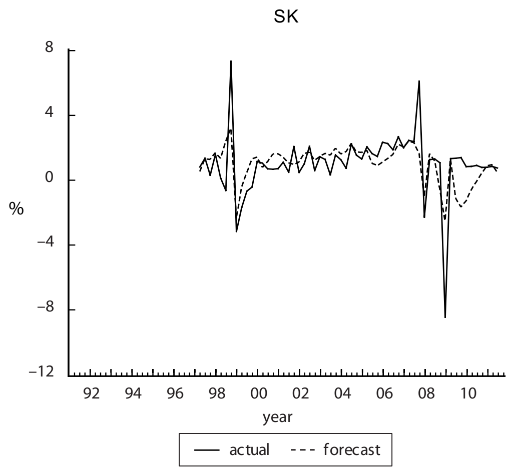

PL
8
6
4
%
2
0
−2
−4
92 94 96 98 00 02 04 06 08 10
year
actual
forecast

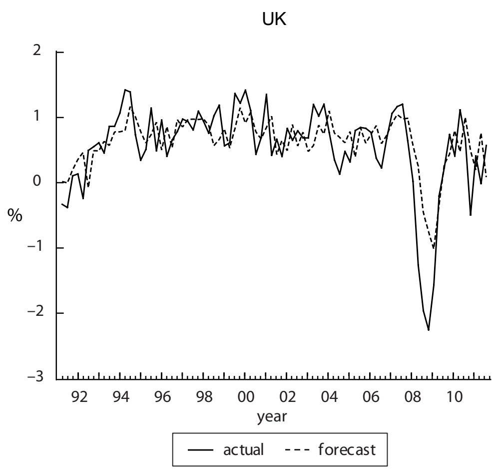

PT
3
2
1
%
0
−1
−2
−3
92 94 96 98 00 02 04 06 08 10
year
actual
forecast


RO
4
3
2
1
%
0
−1
−2
−3
−4
92 94 96 98 00 02 04 06 08 10
year
actual
forecast


SE
3
2
1
0
%
−1
−2
−3
−4
−5
92 94 96 98 00 02 04 06 08 10
year
actual
forecast







Source: Own construction

Using the Superpopulation Model for Imputations and Variance Computation in Survey Sampling

Petr Novák[1] | *Czech Statistical Office, Prague, Czech Republic*
Václav Kosina[2] | *Czech Statistical Office, Prague, Czech Republic*

Abstract

This study is aimed at variance computation techniques for estimates of population characteristics based on survey sampling and imputation. We use the superpopulation regression model, which means that the target variable values for each statistical unit are treated as random realizations of a linear regression model with weighted variance. We focus on regression models with one auxiliary variable and no intercept, which have many applications and straightforward interpretation in business statistics. Furthermore, we deal with cases where the estimates are not independent and thus the covariance must be computed. We also consider chained regression models with auxiliary variables as random variables instead of constants.

Keywords	***JEL code***
Survey sampling, variance estimation, imputation	*C13*

INTRODUCTION

For estimation of population characteristics (mainly totals, means, counts) in business statistics surveys, the Czech Statistical Office (CZSO) has been recently exploring a new approach, in which all data for units that are out of the sample are imputed based on predictions by regression, instead of estimating the population characteristics through weighting. The all-data imputation is based on the superpopulation model (i.e. Cassel et al., 1977, chapter 4). Compared to classical survey methodology (i.e. Hájek, 1960, 1981 or Cochran, 1977), the data are treated as realizations of an infinite population, some of which we know through the survey and some we want to estimate.

Traditional methods, on the other hand, work with the population at hand. All data are treated as fixed constants and the randomness of estimates then comes in form of sample inclusion indicators. The

[1] Department of Probability and Mathematical Statistics, Faculty of Mathematics and Physics, Charles University, Sokolovská 83, 186 75 Prague 8, Czech Republic. Corresponding author: e-mail: petr.novak@czso.cz, phone: (+420)274052141. Czech Statistical Office, Na padesátém 81, 100 82 Prague 10, Czech Republic.

[2] Czech Statistical Office, Na padesátém 81, 100 82 Prague 10, Czech Republic.

population totals are then estimated by weighting methods, such as the Horwitz-Thompson estimator or the ratio estimator. We show that some of the estimates coincide or are very similar.

The drawback of the superpopulation approach subsists in the fact that it relies heavily on the choice of the regression model and appropriate auxiliary variables. However, the all-data imputation allows to group the data and report the results in any desirable way, because we have a predicted value available for each unit in the population.

It is desirable to assess the quality of the obtained estimators by computing their variance, mean square error or the coefficient of variation. Because of the differences between classic and superpopulation modeling, new techniques for survey error computation had to be explored. At first, we derive the estimator of the standard error computation in simple cases with one auxiliary variable in the regression model. Then, we present extensions of the methods for cases where the population is divided in more strata and where the auxiliary variables used for the regression are themselves imputed and form a chain structure, as explored in Raghunathan et al (2001). We illustrate the methods on simplified examples from business statistics.

1 THE SUPERPOPULATION REGRESSION MODEL

In the superpopulation approach we treat the data as random realizations of an infinite population with some model distribution. Suppose that we have sampled n observations and $N - n$ more values must be estimated in order to cover the population of interest. To find appropriate estimates, we have to choose a suitable regression model, study the dependence between the variable of interest and the covariates on the observed data and use the results to predict the unknown part. First, we consider a simple superpopulation model with one regression variable and following assumptions:

- the data y_i are non-negative random variables with $y_i = x_i\beta + e_i$,
- the error terms e_i are independent with distribution $e_i \sim (0, c_i\sigma^2)$,
- x_i and c_i are known positive constants for all i = 1,..., N,
- β and σ^2 are unknown parameters.

By the notation $e_i \sim (0, c_i\sigma^2)$ we mean that the error terms have zero mean and that their variance is equal to $c_i\sigma^2$. Note that we do not assume normality of e_i.

The following methods rely heavily on these assumptions and therefore deviations from the model can make the results inaccurate. The variance scaling constants c_i must be chosen to fit the data well, often it is used $c_i = x_i$ or $c_i \equiv 1$. Methods of assessing the model fit are out of the scope of this paper (see Anscombe, 1961 or Cook and Weisberg, 1983 among others).

We observe n realizations of the variable, which we call the *sample* and denote as *sam*. There are N - n more realized variables, which values we wish to estimate with the knowledge of x_i and c_i. Let us call this unknown part of the population the *imputed* part and denote as *imp*. More accurately we want to estimate the sum:

$$Y = \sum_{i \in sam} y_i + \sum_{i \in imp} y_i , \tag{1}$$

by imputing an estimate for each y_i from the unknown part:

$$\hat{Y} = \sum_{i \in sam} y_i + \sum_{i \in imp} \hat{y}_i . \tag{2}$$

For space saving reasons we will mark the totals with just $\sum_{sam} y_i$ instead of $\sum_{i \in sam} y_i$ etc. We will further use the notation $Y_{sam} = \sum_{sam} y_i$, $Y_{imp} = \sum_{imp} y_i$ and $\hat{Y}_{imp} = \sum_{imp} \hat{y}_i$, similarly for sums of x_i and c_i.

We use classical linear regression model with one covariate and no intercept (the regression line passing through the origin). The estimator of β is obtained using weighted least squares and we use it to impute the data in the following way:

$$\hat{y}_i = x_i \hat{\beta} = x_i \times \frac{\sum_{sam} w_i x_i y_i / c_i}{\sum_{sam} w_i x_i^2 / c_i}, \tag{3}$$

where w_i are appropriately chosen weights (discussed later). Note that for $c_i := x_i$ we get the most commonly used weighted ratio:

$$\hat{\beta} = \frac{\sum_{sam} w_i y_i}{\sum_{sam} w_i x_i}. \tag{4}$$

For constant weights and $c_i :\equiv 1$, we have the classical least-squares estimator:

$\hat{\beta} = \frac{\sum_{sam} x_i y_i}{\sum_{sam} x_i^2}$ and $c_i := x_i^2$ gives the mean ratio $\hat{\beta} = \frac{1}{n} \sum_{sam} \frac{y_i}{x_i}$. It depends on each case, which c_i fits the data best.

We can easily verify regardless of the choice of c_i and w_i, that:

$$E\hat{\beta} = \frac{\sum_{sam} w_i x_i E y_i / c_i}{\sum_{sam} w_i x_i^2 / c_i} = \beta, \tag{5}$$

$$\operatorname{var} \hat{\beta} = \frac{\sum_{sam} w_i^2 x_i^2 / c_i^2 \operatorname{var} y_i}{(\sum_{sam} w_i x_i^2 / c_i)^2} = \frac{\sum_{sam} w_i^2 x_i^2 / c_i}{(\sum_{sam} w_i x_i^2 / c_i)^2} \sigma^2 =: \sigma_{\beta}^2. \tag{6}$$

Example

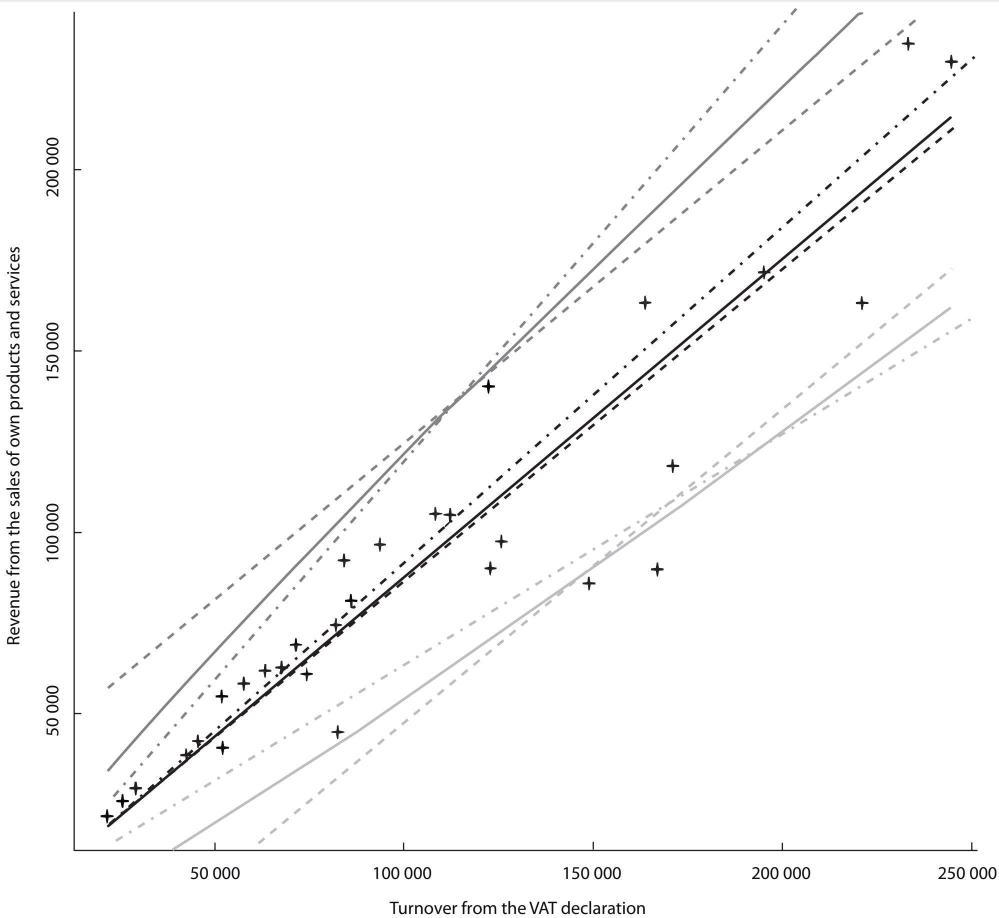
In Figure 1 we see sample data (n = 30) from one particular stratum of the annual structural business survey. We model the dependency of the revenue from the sales of own products and services (y_i) on the turnover given in the VAT declaration (x_i), both given in CZK 1 000. We fitted regression line using $c_i :\equiv 1$ (dashed) $c_i := x_i$ (full) and $c_i := x_i^2$ (dash dot). If the distribution of e_i was Gaussian, we could roughly approximate 95% – confidence bands for the predicted data as $\left(x_i \hat{\beta} - 2\sqrt{c_i}\hat{\sigma}, x_i \hat{\beta} + 2\sqrt{c_i}\hat{\sigma}\right)$, these are marked in gray. The estimated coefficients $\hat{\beta}$, their standard deviations $\hat{\sigma}_{\beta}$ and the constants $\hat{\sigma}$ are shown in Table 1.

Table 1 Estimated regression parameters

	$c_i := 1$	$c_i := x_i$	$c_i := x_i^2$
$\hat{\beta}$	0.864	0.879	0.923
$\hat{\sigma}$	19 295	53.11	0.143
$\hat{\sigma}_{\beta}$	0.029	0.03	0.026

Note: $\hat{\beta}$, $\hat{\sigma}$ – estimates of the regression slope β and the standard deviation σ, $\hat{\sigma}_{\beta}$ – estimated variance of $\hat{\beta}$, c_i – variance scaling.
Source: Simulation – own construction, Czech Statistical Office

Figure 1 Modeling the dependency of the revenue from the sales of own products and services on the turnover given in the VAT declaration, using different variance scaling constants c_i – estimated regression lines with approximate 95% – confidence bands for the data



Source: Czech Statistical Office, data modified to maintain confidentiality

Note that the estimated parameters and therefore also the regression lines are quite similar. Estimators with $c_i := x_i$ and $c_i := x_i^2$ are less sensitive to observations with higher covariate values. The standard deviation parameters $\hat{\sigma}$ differ, because in each case they have a different meaning. The standard deviation of the parameter estimates is again similar in each case. The observations seem to have an increasing deviation from the regression line with higher x_i, which suggests that $c_i := x_i$ or $c_i := x_i^2$ are better choices for the variance scaling than $c_i := 1$.

2 VARIANCE ESTIMATION WITH SIMPLE REGRESSION IMPUTATIONS

Let us derive the formula for the error of $\hat{Y}$. Because of the superpopulation model, the variables y_i which we estimate are random variables instead of constants. Therefore we cannot use the common formula:

$$\operatorname{var}\hat{Y} = E(\hat{Y} - E\hat{Y})^2. \tag{7}$$

In fact, we are interested in the mean square error of the difference of the real and estimated (predicted) values of the random variables:

$$mse\hat{Y} = E(\hat{Y} - Y)^2, \tag{8}$$

given the realization of the sample data. We should write $E(\hat{Y} - Y \mid sam)^2$, but we leave the condition out for space saving reasons. This is the main difference from the usual theoretical methods in survey sampling, where all data are taken as constants and the randomness is included in the models in form of inclusion indicators. If we take y_i as realizations of random variables from the superpopulation model, we can derive the formulas for the variance also in more complex situations.

For the imputed data we have:

$$E\hat{y}_i = Ex_i\hat{\beta} = x_i\beta = Ey_i, \tag{9}$$

therefore $E\hat{Y}_{imp} = EY_{imp}$. For the mse we then get:

$$\begin{aligned} E(\hat{Y} - Y)^2 &= E(\hat{Y}_{imp} - Y_{imp})^2 = E(\hat{Y}_{imp} - E\hat{Y}_{imp} - (Y_{imp} - EY_{imp}))^2 \\ &= E(\hat{Y}_{imp} - E\hat{Y}_{imp})^2 + E(Y_{imp} - EY_{imp})^2 - 2E[(\hat{Y}_{imp} - E\hat{Y}_{imp})(Y_{imp} - EY_{imp})]. \end{aligned} \tag{10}$$

The third (covariance) term will be zero, because it consists of two independent terms, both with a zero mean ($\hat{Y}_{imp}$ is computed form the sample, Y_{imp} is the rest). Denote $c_{imp} = \sum_{imp} c_i$. Then:

$$\begin{aligned} mse\hat{Y} &= \operatorname{var}\hat{Y}_{imp} + \operatorname{var}Y_{imp} = \operatorname{var}X_{imp}\hat{\beta} + c_{imp}\sigma^2 \\ &= X_{imp}^2\sigma_\beta^2 + c_{imp}\sigma^2. \end{aligned} \tag{11}$$

The constants x_i and c_i are known and $\sigma_\beta^2 = \dfrac{\sum_{sam} w_i^2 x_i^2 / c_i}{(\sum_{sam} w_i x_i^2 / c_i)^2}\sigma^2$. For establishing the estimate $m\hat{s}e\hat{Y}$ we only need to use an appropriate estimate of σ^2, i.e.

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{sam}\frac{(y_i - \hat{\beta}x_i)^2}{c_i}, \tag{12}$$

or

$$\hat{\sigma}^2 = \frac{1}{\sum_{sam} w_i - \overline{w}}\sum_{sam}\frac{w_i(y_i - \hat{\beta}x_i)^2}{c_i}, \tag{13}$$

where $\overline{w} = \frac{1}{n}\sum_{sam} w_i$.

We see, that the estimate of *mse* consists of the model parameter estimates on the sample part and of the sums of auxiliary variables on the imputed part of the data.

It is clear that the more data is in the imputed part, the higher is the mean square error. On the other hand, the more sampled data we have, the more accurately we can estimate $\hat{\beta}$ and therefore σ_β^2 is smaller in the most cases. For example if the weights are constant, then $\sigma_\beta^2 = \dfrac{1}{\sum_{sam} x_i^2 / c_i}\sigma^2$ is a non-increasing function of *n*.

Example (continued)
In the stratum from the example given in the last section, the revenue from the sales of own products and services was $Y_{sam} = 3\,693\,886$. Suppose we have 50 non-sampled units in the observed stratum. We want to impute the data with the help of known turnover from VAT declaration, for which $X_{imp} = 6\,317\,817$.

We use the same c_i and the estimated regression parameters from above. In Table 2 we see the auxiliary totals c_{imp}, estimated totals $\hat{Y} = Y_{sam} + X_{imp}\hat{\beta}$, the mean square error $m\hat{s}e\hat{Y} = X_{imp}^2\hat{\sigma}_{\beta}^2 + c_{imp}\hat{\sigma}^2$ and the modified coefficient of variation $CV(\hat{Y}) = \frac{\sqrt{m\hat{s}e\hat{Y}}}{\hat{Y}}$ for each choice of c_i.

Table 2 Estimated characteristics

	$c_i := 1$	$c_i := x_i$	$c_i := x_i^2$
$\hat{Y}$	9 154 548	9 247 872	9 527 206
c_{imp}	50	6.31×10^9	1.79×10^{12}
MSE	5.26×10^9	5.42×10^9	6.39×10^9
CV	2.51%	2.52%	2.65%

Note: $\hat{Y}$ – estimated total, MSE – mean square error, CV – coefficient of variation, c_i – variance scaling, c_{imp} – total of c_i – over the imputed part.
Source: Simulation – own construction, primary data: Czech Statistical Office

3 VARIANCE COMPUTATION FOR MORE COMPLEX CASES

By using the superpopulation model, we get closer to linear regression theory and therefore we can derive the variance of the population estimators in various situations where using the classic survey sampling methodology can be overly complicated.

3.1 Variance of chain imputations

Suppose we deal with data y_i estimated with the help of random auxiliary variables x_i, which are known only for the units in the sample, elsewhere it is imputed with the help of known constants z_i. For each step, we assume the same model as above:

$$y_i \mid x_i \sim (\beta_y x_i, c_i \sigma_y^2), \qquad x_i \sim (\beta_x z_i, d_i \sigma_x^2), \tag{14}$$

with $y_i \mid x_i$ meaning the conditional distribution of y_i given x_i and d_i being the variance-scaling factors of y_i. The regression parameters are estimated in following way:

$$\hat{\beta}_y = \frac{\sum_{sam} w_i x_i y_i / c_i}{\sum_{sam} w_i x_i^2 / c_i}, \qquad \hat{\beta}_x = \frac{\sum_{sam} v_i z_i x_i / d_i}{\sum_{sam} v_i z_i^2 / d_i}. \tag{15}$$

The estimates have then similar properties:

$$\hat{\beta}_y \sim \left(\beta_y, \sigma_{\beta_y}^2 := \frac{\sum_{sam} w_i x_i^2 / c_i}{(\sum_{sam} w_i x_i^2 / c_i)^2} \sigma_y^2 \right), \qquad \hat{\beta}_x \sim (\beta_x, \sigma_{\beta_x}^2). \tag{16}$$

Note that the distribution of $\hat{\beta}_y$ is conditional given the values of x_i, $i = 1,\ldots,n$. At first, $\hat{x}_i$ are imputed, afterwards we impute $\hat{y}_i$ with their help:

$$\hat{x}_i = \hat{\beta}_x z_i, \qquad \hat{y}_i = \hat{\beta}_y \hat{x}_i. \tag{17}$$

Using the conditional expectation, for the imputed part we have:

$$E\hat{y}_i = E[E[\hat{y}_i \mid x_i]] = E[E[\hat{\beta}_y \hat{x}_i \mid x_i]] = E\beta_y \hat{x}_i = \beta_y E\hat{x}_i = \beta_y \beta_x z_i = E[E[y_i \mid x_i]] = Ey_i. \tag{18}$$

We want to compute the mean square error of the prediction of the random variables Y estimated by $\hat{Y}$. With the help of conditional variance decomposition we get:

$$\begin{aligned} mse\hat{Y} &= E(\hat{Y}_{imp} - Y_{imp})^2 = E(\hat{Y}_{imp} - E\hat{Y}_{imp})^2 + E(Y_{imp} - EY_{imp})^2 \\ &= \operatorname{var}\hat{Y}_{imp} + \operatorname{var} Y_{imp} \\ &= E\operatorname{var}[\hat{Y}_{imp} \mid X] + \operatorname{var} E[\hat{Y}_{imp} \mid X] + E\operatorname{var}[Y_{imp} \mid X] + \operatorname{var} E[Y_{imp} \mid X] \\ &= E\hat{X}^2_{imp}\sigma^2_{\beta_y} + \operatorname{var}[\hat{X}_{imp}\beta_y] + Ec_{imp}\sigma^2_y + \operatorname{var} X_{imp}\beta_y \\ &= E[\hat{X}^2_{imp}\sigma^2_{\beta_y} + c_{imp}\sigma^2_y] + \beta^2_y(\operatorname{var}\hat{X}_{imp} + \operatorname{var} X_{imp}) \\ &= EE[(\hat{Y}_{imp} - Y_{imp}) \mid X]^2 + \beta^2_y E(\hat{X}_{imp} - X_{imp})^2 \\ &= Emse(\hat{Y} \mid X) + \beta^2_y mse(\hat{X}). \end{aligned} \tag{19}$$

The second term may be estimated by plugging $\hat{\beta}_y$ and $m\hat{s}e\hat{X}$ into the formula. The estimation of the expectation with respect to the distribution of x_i in the first term would be relatively complex, because of the values x_i which are in both nominator and denominator of $\sigma^2_{\beta_y}$. We need to find an appropriate estimate, we can use instead of $Emse(\hat{Y} \mid X)$ the term:

$$m\hat{s}e(\hat{Y} \mid \hat{X}) = \hat{X}^2_{imp}\hat{\sigma}^2_{\beta_y} + \hat{c}_{imp}\hat{\sigma}^2_y. \tag{20}$$

We get $\hat{X}^2_{imp}\hat{\sigma}^2_{\beta_y} + \hat{c}_{imp}$ $\hat{X}_{imp}$, $\hat{\sigma}^2_y$ and $\hat{\sigma}^2_{\beta_y}$ through the estimates of $\hat{x}_i$. The estimate $\hat{c}_{imp}$ follows from the chosen model of the variance (i.e. $c_i := x_i$ or $c_i := x_i^2$). We get:

$$m\hat{s}e(\hat{Y}) = m\hat{s}e(\hat{Y} \mid \hat{X}) + \hat{\beta}^2_y m\hat{s}e(\hat{X}). \tag{21}$$

When we work with a chain structure having more levels, the first term $m\hat{s}e(\hat{Y} \mid \hat{X})$ and $\hat{\beta}_y$ remain the same, because they are conditional estimates given their auxiliary variable. The second term may be obtained through another chain estimation, so we are getting a recurrent formula, which leads so far until it reaches an auxiliary variable which is known for all units (i.e. administrative data sources).

3.2 Stratification level shifts – covariance computation

The CZSO works with the stratification approach, where the surveyed enterprises are divided into strata depending on the number of employees, type of economic activity, region etc. The stratification has more levels, going from relatively small groups to larger ones. In each stratum, the regression parameters are estimated separately. When it is not possible to obtain the estimates in given stratum, mainly because of a low number of responding units, we use the estimates in the corresponding superior stratum at a higher stratification level.

Let us consider the non-chained regression from section 2. Let m be a small stratum where the estimates for β_m and σ^2_m could not be obtained. Let S be its superior stratum (one or more levels higher), with enough units to compute the estimates:

$$\hat{\beta}_S = \frac{\sum_{S_{sam}} w_i x_i y_i / c_i}{\sum_{S_{sam}} w_i x_i^2 / c_i}, \tag{22}$$

for the variance of the estimate of the sum Y_m we impute $\hat{y}_i = \hat{\beta}_S x_i$ and we get:

$$\begin{aligned} mse\hat{Y}_m &= \operatorname{var}\hat{Y}^m_{imp} + \operatorname{var} Y^m_{imp} \\ &= \operatorname{var}\hat{X}^m_{imp}\hat{\beta}_S + \operatorname{var} Y^m_{imp} = (X^m_{imp})^2\sigma^2_{\beta_S} + c_{imp}\sigma^2_m. \end{aligned} \tag{23}$$

The estimate for $\sigma^2_{\beta_S}$ is obtained from the superior stratum S, σ^2_m is completely unknown and cannot be estimated from m, therefore we use the estimate for σ^2_S instead.

Suppose we now have one stratum *S* in a higher level, which consists of two substrata: one too small (m) and one good (d), where it is possible to estimate β_d and σ_d^2. We want to obtain the variance for the sum *Y* for the whole *S*. Using the above given formulas and the independence assumption for e_i, we get:

$$\begin{aligned} mse\hat{Y} &= \operatorname{var}\hat{Y} + \operatorname{var}Y = \operatorname{var}(\hat{Y}_m + \hat{Y}_d) + \operatorname{var}(Y_m + Y_d) \\ &= \operatorname{var}\hat{Y}_m + \operatorname{var}\hat{Y}_d + 2\operatorname{cov}(\hat{Y}_m, \hat{Y}_d) + \operatorname{var}Y_m + \operatorname{var}Y_d \\ &= mse(\hat{Y}_m) + mse(\hat{Y}_d) + 2\operatorname{cov}(\hat{Y}_m, \hat{Y}_d). \end{aligned} \tag{24}$$

The covariance is computed in the following way:

$$\begin{aligned} \operatorname{cov}(\hat{Y}_m, \hat{Y}_d) &= \operatorname{cov}(X_{imp}^m \hat{\beta}_S, X_{imp}^d \hat{\beta}_d) = X_{imp}^m X_{imp}^d \operatorname{cov}(\hat{\beta}_S, \hat{\beta}_d) \\ &= X_{imp}^m X_{imp}^d \operatorname{cov}\left(\frac{\sum_{S_{sam}} w_i x_i y_i / c_i}{\sum_{S_{sam}} w_i x_i^2 / c_i}, \frac{\sum_{d_{sam}} w_i x_i y_i / c_i}{\sum_{d_{sam}} w_i x_i^2 / c_i} \right) \\ &= \frac{X_{imp}^m X_{imp}^d}{\sum_{S_{sam}} w_i x_i^2 / c_i \sum_{d_{sam}} w_i x_i^2 / c_i} \operatorname{cov}\left(\sum_{S_{sam}} w_i x_i y_i / c_i, \sum_{d_{sam}} w_i x_i y_i / c_i \right). \end{aligned} \tag{25}$$

The variables y_i belonging to m and d are mutually independent, therefore it is enough to take the sum only through d in the first term of the covariance. Denote as B_S and B_d the sums we have taken out of the parentheses in the denominator:

$$\begin{aligned} &= \frac{X_{imp}^m X_{imp}^d}{B_S B_d} \operatorname{cov}\left(\sum_{d_{sam}} w_i x_i y_i / c_i, \sum_{d_{sam}} w_i x_i y_i / c_i \right) \\ &= \frac{X_{imp}^m X_{imp}^d}{B_S B_d} \operatorname{var} \sum_{d_{sam}} w_i x_i y_i / c_i = \frac{X_{imp}^m X_{imp}^d}{B_S B_d} \sum_{d_{sam}} w_i^2 x_i^2 / c_i^2 \operatorname{var} y_i \\ &= \frac{X_{imp}^m X_{imp}^d}{B_S B_d} \sum_{d_{sam}} w_i^2 x_i^2 / c_i \sigma_d^2 = X_{imp}^m X_{imp}^d \frac{B_d}{B_S} \sigma_{\beta_d}^2 . \end{aligned} \tag{26}$$

If we estimate the parameter $\sigma_{\beta_d}^2$ from the good stratum d, we get the whole variance. In a similar way, the covariance of estimates for any two strata can be obtained. Take m_1 and m_2, for which the estimates are taken from the strata S_{m_1} and S_{m_2}. Denote m_1^{sam} the sampled part of the stratum m_1 etc. If m_1 is a good stratum, then $m_1^{sam} = S_{m_1}^{sam}$, otherwise $m_1^{sam} \subset S_{m_1}^{sam}$. The same for m_2. Suppose that the stratification structure is well ordered, in the way that each substratum is contained in exactly one superior stratum. Denote $S_d^{sam} = S_{m_1}^{sam} \cap S_{m_2}^{sam}$ and $S^{sam} = S_{m_1}^{sam} \cup S_{m_2}^{sam}$. Because of the well-ordered stratification, S_d^{sam} is necessarily either the smaller of the sets $S_{m_1}^{sam}$ and $S_{m_2}^{sam}$ or an empty set if the strata do not overlap. For the covariance we get:

$$\operatorname{cov}(\hat{Y}_{m_1}, \hat{Y}_{m_2}) = \frac{X_{imp}^{m_1} X_{imp}^{m_2}}{B_{S_{m_1}^{sam}} B_{S_{m_2}^{sam}}} \sum_{i \in S_d^{sam}} w_i^2 x_i^2 / c_i \sigma_{S_d^{sam}}^2 = X_{imp}^{m_1} X_{imp}^{m_2} \frac{B_{S_d^{sam}}}{B_{S^{sam}}} \sigma_{\beta_{S_d^{sam}}}^2 . \tag{27}$$

It cat be further shown, that for a larger stratum S consisting of $d = 1,\ldots,D$ good and $m = 1,\ldots,M$ small strata we get:

$$mse(\hat{Y}_S) = \sum_{d=1}^{D} mse\hat{Y}_d + \sum_{m=1}^{M} mse\hat{Y}_m + 2\sum_{m=1}^{M} X_{imp}^m \sum_{d=1}^{D} X_{imp}^d \frac{B_d}{B_S} \sigma_{\beta_d}^2 + \sum_{m_i \neq m_j} X_{imp}^{m_i} X_{imp}^{m_j} \sigma_{\beta_S}^2 . \tag{28}$$

3.3 Stratification level shifts – chained imputations

We generalize now the methods used for stratification level shifts for the cases, when the data y_i are imputed with help of estimated auxiliary variables x_i, which are obtained through regression with respect to known constants z_i. In terms of model parameters we have $y_i \mid x_i \sim (\beta_y x_i, c_i\sigma_y^2)$, and $x_i \sim (\beta_x z_i, d_i\sigma_x^2)$. Let S be a large stratum consisting of substrata m (small) and d (good). Then the mean square error can be decomposed as:

$$\begin{aligned} mse\hat{Y}_S &= \operatorname{var}\hat{Y}_S + \operatorname{var}Y_S = \operatorname{var}\hat{Y}_m + \operatorname{var}\hat{Y}_m + 2\operatorname{cov}(\hat{Y}_m,\hat{Y}_d) + \operatorname{var}Y_m + \operatorname{var}Y_d \\ &= mse(\hat{Y}_m) + mse(\hat{Y}_d) + 2\operatorname{cov}(\hat{Y}_m,\hat{Y}_d). \end{aligned} \tag{29}$$

Both *mse* of sums just in strata d and m can be estimated through methods given in section (3.1):

$$m\hat{s}e(\hat{Y}_d) = m\hat{s}e(\hat{Y}_d \mid \hat{X}) + \hat{\beta}_{yd}^2 m\hat{s}e(\hat{X}_d), \tag{30}$$

$$m\hat{s}e(\hat{Y}_m) = m\hat{s}e(\hat{Y}_m \mid \hat{X}) + \hat{\beta}_{yS}^2 m\hat{s}e(\hat{X}_m). \tag{31}$$

The covariances are derived with help of conditional covariance decomposition:

$$\begin{aligned} \operatorname{cov}(\hat{Y}_d,\hat{Y}_m) &= E\operatorname{cov}[\hat{Y}_d,\hat{Y}_m \mid X] + \operatorname{cov}(E[\hat{Y}_d \mid X], E[\hat{Y}_m \mid X]) \\ &= E\operatorname{cov}[\hat{Y}_d,\hat{Y}_m \mid X] + \beta_{yd}\beta_{yS}\operatorname{cov}(\hat{X}_d,\hat{X}_m). \end{aligned} \tag{32}$$

The estimation of the mean of the first term with respect to X would be rather difficult, we substitute it with the estimate with the help of $\hat{X}$:

$$\hat{\operatorname{cov}}(\hat{Y}_d,\hat{Y}_m) = \hat{\operatorname{cov}}[\hat{Y}_d,\hat{Y}_m \mid \hat{X}] + \hat{\beta}_{yd}\hat{\beta}_{yS}\hat{\operatorname{cov}}(\hat{X}_d,\hat{X}_m). \tag{33}$$

The coefficients and $\hat{\beta}_{yd}$ and the first term of the sum can be computed given the estimates $\hat{x}_i$:

$$\hat{\operatorname{cov}}[\hat{Y}_d,\hat{Y}_m \mid \hat{X}] = \hat{X}_{imp}^m \hat{X}_{imp}^d \frac{\hat{B}_d^x}{\hat{B}_S^x}\hat{\sigma}_{\beta_y d}^2, \tag{34}$$

the second covariance term may be estimated as:

$$\hat{\operatorname{cov}}(\hat{X}_d,\hat{X}_m) = Z_{imp}^m Z_{imp}^d \frac{B_d^z}{B_S^z}\hat{\sigma}_{\beta_x d}^2. \tag{35}$$

Similarly as for the mean square errors, we now also have a recurrent formula for the covariances. If z_i would have an auxiliary variable which must be estimated, the estimate of the second term will be chained until it leads to constant covariates.

It can be also shown, that the formula will work also when in the strata m or d are some values y_i imputed, but corresponding values x_i are observed in the sample.

The covariance estimation for more than two strata can be generalized in a similar way as in the case with no chain structure.

4 REMARKS

4.1 Special cases

The above described techniques are quite general. Often we work simply with $c_i := x_i$. The population estimate is then:

$$\hat{Y} = Y_{sam} + X_{imp}\frac{\sum_{sam} w_i y_i}{\sum_{sam} w_i x_i}, \tag{36}$$

which is an analogy to the ratio estimator from the classic survey methodology (i.e. Levy and Lemeshow, 1999),

$$\hat{Y}_R = X_{all}\frac{\sum_{sam} w_i y_i}{\sum_{sam} w_i x_i}. \tag{37}$$

The mean square error then reduces to:

$$mse\hat{Y} = X_{imp}^2\sigma_\beta^2 + c_{imp}\sigma^2 = X_{imp}^2\frac{\sum_{sam} w_i^2 x_i}{\left(\sum_{sam} w_i x_i\right)^2}\sigma^2 + X_{imp}\sigma^2. \tag{38}$$

When the weights are constant, we get:

$$\hat{Y} = Y_{sam} + X_{imp} \frac{Y_{sam}}{X_{sam}} = X_{all} \frac{Y_{sam}}{X_{sam}}, \tag{39}$$

which is equal to the ratio estimator. For the error we get:

$$mse\hat{Y} = X_{imp} \left(X_{imp} \frac{X_{sam}}{X_{sam}^2} + 1 \right) \sigma^2 = X_{imp} \frac{X_{all}}{X_{sam}} \sigma^2. \tag{40}$$

If no auxiliary information is available, we may use $x_i \equiv 1$, which means that we impute just the sample mean for each unit. We obtain:

$$mse\hat{Y} = (N-n)\frac{N}{n}\sigma^2 = \frac{N^2}{n}\left(1 - \frac{n}{N}\right)\sigma^2, \tag{41}$$

which is the commonly used formula for simple random sampling variance.

4.2 Choosing the weights

For getting the population estimates, we use imputations with help of the superpopulation model, rather than the commonly used weighting techniques. The weights are used in the estimates $\hat{\beta}$ and, therefore, they have a different meaning.

If we observe just one stratum alone with no relation to others, it would be appropriate to use constant weights (which may simply be equal to one for that case, because the constants in the numerator and denominator of $\hat{\beta}$ cancel out).

If we apply some outlier-detection methods to identify observations that may not fit the model (see e.g. Grubbs, 1969 or Barnett and Lewis, 1994), we can simply put $w_k := 0$ for that units, meaning that they will not influence the parameter estimates in any way.

In the case when we need to use higher level stratification to obtain the estimates, the weights can be chosen in a way that they reflect the proportion of sampled units in each sub-strata, i.e. $w_k := N_k / n_k$ for sub-stratum k with n_k from N_k units sampled. Therefore the data from the greater strata influence the estimates more than the data from the smaller strata. However, this approach is rather simplified. The proportion of sampled units can be much lower in the studied small stratum than in the neighbouring strata, resulting in overly high weights. Also the dependency of the studied and auxiliary variables may differ between the strata. These considerations open an entire field of Small Area Estimation, which has been extensively studied for example by Rao (2003).

4.3 Multivariate regression

The methods as described in sections 1–3 can be easily generalized to accommodate more regression variables with modeling the data as $y_i = \vec{x}_i^T \vec{\beta} + e_i$, with $e_i \sim (0, c_i\sigma^2)$, covariate vector $\vec{x}_i = (x_{1i}, ..., x_{pi})^T$ and the vector of parameters $\vec{\beta} = (\beta_1, ..., \beta_p)^T$. We would then have to work with matrix calculus, for the sampled part using $\vec{Y} = (y_1, ..., y_n)^T$, a $n \times p$ matrix $X = (\vec{x}_1, ..., \vec{x}_n)^T$, vector and an $n \times n$ diagonal variance scaling matrix C with $C_{ii} = \frac{w_i}{c_i}$. Because c_i are one-dimensional, they have to be chosen as a function of one or more of the covariates.

The regression parameters can be than estimated as $\hat{\vec{\beta}} = (X^T CX)^{-1} X^T C\vec{Y}$ with $p \times p$ variance matrix $V = \text{var}\,\hat{\vec{\beta}} = (X^T CX)^{-1}(X^T C^2 X)(X^T CX)^{-1}\sigma^2$. We can take:

$$\hat{\sigma}^2 = \frac{1}{n-p} \sum_{sam} \frac{(y_i - x_i^T \hat{\beta})^2}{c_i}. \tag{42}$$

We estimate the target variable as $\hat{y}_i = \vec{X}_i^T \hat{\hat{\beta}}$ for $i \in imp$. Denote $\vec{X}_{imp} = \left(\sum_{imp} X_{1i}, \ldots, \sum_{imp} X_{pi} \right)^T$.

Then the mean square error of the estimate is:

$$mse\hat{Y} = \vec{X}_{imp}^T V \vec{X}_{imp} + c_{imp}\sigma^2, \tag{43}$$

and the results from section 3 can be generalized similarly. Note that in this way we could include also the intercept term.

Difficulties can arise when the matrix $X^T CX$ is singular or under-determined, which can be the case when there is a linear dependency between the regression variables. It is then impossible to compute the inverse $(X^T CX)^-$, therefore for estimating $\hat{\hat{\beta}}$ one must either omit one or more of the covariates or use some pseudo-inverse matrix $(X^T CX)^-$, such as the Moore-Penrose pseudo-inverse matrix (Penrose, 1955).

5 EXAMPLES

Mean square error estimation by the means of the superpopulation model as shown here has been adapted by the CZSO for business statistics. Larger surveys often have a very detailed stratification structure, with many small strata consisting of only a few units. Also a sequential approach is used, when the most important variables are estimated first and with their help the other ones are imputed, building a chain structure. We show here examples of mean square error and coefficient of variation estimation.

5.1 Revenue from sales of own products and services

First, suppose we want to estimate the aggregate revenue from sales of own products and services in one particular two-digit NACE stratum using the annual structural business statistics survey data from year 2010. The population of enterprises was divided into sampling substrata by size class (1–9, 10–19, 20–49 employees according to the business register) and by three-digit NACE (in this case there are three subgroups, say 1–3). We estimate the regression coefficients for each of the groups separately. If there are less than 15 responding enterprises in one group, we use there the coefficient $\hat{\beta}$ computed over the whole corresponding size class group. As the auxiliary variable x_i, the total turnover from tax declaration was taken. We take again the variance scaling as $c_i \equiv 1$, $c_i = x_i$ and $c_i = x_i^2$ and compare the results. An outlier detection technique based on assessing the influence of each observation on the estimate $\hat{\beta}$ was used.

In Table 3, we see the number of enterprises sampled (*sam*) and non-sampled or non-responding (*imp*) in respective groups. The sample was designed to pay more attention to larger companies. In the higher size classes, all units were sampled and some of them did not respond. There are some strata with relatively few sampled units (enterprises of higher size in 3-digit NACE groups 1 and 3, marked in italics).

Table 3 The number of enterprises in the sampling strata

		NACE3					
		1		2		3	
		Sam	Imp	Sam	Imp	Sam	Imp
Size class	0–9	20	38	86	110	42	82
	10–19	*4*	1	35	4	*14*	0
	20–49	*10*	0	25	1	*12*	1

Note: Sam – sampled part, Imp – imputed part.
Source: Czech Statistical Office

Table 4 The number of enterprises in the imputation groups

		NACE3							
		1		2		3		Total	
		Sam	Imp	Sam	Imp	Sam	Imp	Sam	Imp
Size class	**0–9**	(2,11,7)	(6,22,10)	(6,61,19)	(11,52,47)	(2,28,12)	(20,35,27)	(10,100,38)	(37,109,84)
	10–19	(1,2,1)	(0,0,1)	(2,21,12)	(1,0,3)	(2,9,3)	0	(5,32,16)	(1,0,4)
	20–49	(0,7,3)	0	(1,16,8)	(0,1,0)	(1,9,2)	(1,0,0)	(2,32,13)	(1,1,0)
	Total	(3,20,11)	(6,22,11)	(9,98,39)	(12,53,50)	*(5,46,17)*	(21,35,27)	(17,164,67)	(39,110,88)

Note: Sam – sampled part, Imp – imputed part.
Source: Czech Statistical Office

The regression coefficient estimates would not be reliable, if taken in these strata separately. Therefore we compute estimates for each whole size class so that the coefficients in smaller NACE groups 1 and 3 are obtained using information also from the group 2. Fortunately, there are no units to estimate in two of the small strata and the other two small strata have both just one non-responding unit.

We estimated $\hat{Y}$, corresponding $m\hat{s}e$ and coefficients of variation first for the whole population and then for regional division in which enterprises were divided into three groups by place of residence: i) those residing in the capital city of Prague, ii) in the rest of Bohemia and iii) in Moravia. The number of sampled and non-sampled enterprises in each region can be seen in Table 4 in parentheses (Prague, Bohemia, Moravia).

Table 5 Revenue from sales of own products and services – the whole population

c_i	$\hat{Y}$	MSE	CV
1	11 578 276	5 632 297 044	0.65%
x_i	11 699 438	3 255 484 884	0.49%
x^2_i	11 739 074	7 428 077 251	0.73%

Note: $\hat{Y}$ – estimated total, MSE – mean square error, CV – coefficient of variation, c_i – variance scaling.
Source: Simulation – own construction, primary data: Czech Statistical Office

Table 6 Revenue from sales of own products and services – regions

Region	c_i	$\hat{Y}$	MSE	CV
Prague	1	1 133 291	1 102 787 426	2.93%
	x_i	1 158 584	350 602 637	1.62%
	x^2_i	1 159 533	1 501 735 362	3.34%
Bohemia	1	7 118 493	1 970 034 661	0.62%
	x_i	7 179 980	1 124 227 221	0.47%
	x^2_i	7 202 045	2 714 996 570	0.72%
Moravia	1	3 326 493	1 375 424 879	1.11%
	x_i	3 360 874	644 562 108	0.76%
	x^2_i	3 377 496	1 546 626 660	1.16%

Note: $\hat{Y}$ – estimated total, MSE – mean square error, CV – coefficient of variation, c_i – variance scaling.
Source: Simulation – own construction, primary data: Czech Statistical Office

The mean square error is computed in each of the regions separately, using the coefficients estimated over the sampling strata and the totals of auxiliary data in the region. Note that because the coefficients for small strata are taken from the size-class groups, covariance between estimates has to be computed as shown in section 3.2. We can see the results for each type of variance scaling c_i in Tables 5 and 6.

The estimated totals $\hat{Y}$ using different c_i are similar. The coefficient of variation differs, we can see that $c_i = x_i$ yields more accurate results than $c_i \equiv 1$ or $c_i = x_i^2$ in each case. Generally the estimated coefficients of variations are quite low, which is partly because the sampling ratio was high and the sample focused on larger and more important enterprises and partly also due to good regression fit.

5.2 Revenue from the lease of land

Suppose we want to estimate the total revenue from the lease of land in the same population and the corresponding prediction error. As auxiliary variables x_i, for each enterprise we take the predicted values of the revenue from the sales of own products and services from above. Thus we have a chain structure and therefore it is necessary to use the method described in section 3.2. Because there are some small strata, the covariance has to be computed via the chain structure as shown in section 3.3.

Again, we take the variance scaling as $c_i \equiv 1$, $c_i = x_i$ and $c_i = x_i^2$ and compare the results.

In Tables 7 and 8 we see that the estimated totals are again similar to each choice of c_i. The coefficient of variation of $\hat{Y}$ for the whole population is the lowest with $c_i \equiv 1$. Among the regions it is not so clear, the mean square error is lowest in two cases with $c_i \equiv 1$ and in one case with $c_i = x_i$.

Table 7 Revenue from the lease of land – the whole population

c_i	$\hat{Y}$	MSE	CV
1	31 492	31 291	0.56%
x_i	31 629	53 565	0.73%
x^2_i	31 751	138 821	1.17%

Note: $\hat{Y}$ – estimated total, MSE – mean square error, CV – coefficient of variation, c_i – variance scaling.
Source: Simulation – own construction, primary data: Czech Statistical Office

Table 8 Revenue from the lease of land – regions

Region	c_i	$\hat{Y}$	MSE	CV
Prague	1	15 119	9 898	0.66%
	x_i	15 139	4 542	0.45%
	x^2_i	15 153	13 312	0.76%
Bohemia	1	14 981	16 999	0.87%
	x_i	15 059	38 704	1.31%
	x^2_i	15 123	68 859	1.74%
Moravia	1	1 393	3 307	4.13%
	x_i	1 431	4 909	4.89%
	x^2_i	1 475	33 480	12.41%

Note: $\hat{Y}$ – estimated total, MSE – mean square error, CV – coefficient of variation, c_i – variance scaling.
Source: Simulation – own construction, primary data: Czech Statistical Office

CONCLUSION

The superpopulation regression model and all-data imputation presents an alternative approach to estimate the population totals in survey sampling. It is then easier to provide estimates with respect to various groupings. We have shown how to compute the mean square error in order to assess the accuracy of the estimators. In simple cases, this approach leads to similar estimators as the commonly used formulas for classic simple random sampling. However, using the superpopulation model it is easier to derive error estimates in more complex cases with sophisticated stratification and chain structure, as we have shown.

Because the superpopulation approach is model-based, the results can be inaccurate if the model assumptions are not met. Further research can concern sensitivity analysis on departures from the assumed model, presence of outliers and goodness-of-fit tests.

References

ANSCOMBE, F. J. Examination of Residuals. *Proc. 4th Berkeley Symp.* 1961, 1, pp. 1–36.

BARNETT, V., LEWIS, T. *Outliers in Statistical Data, 3rd edition.* New York: John Wiley & Sons, 1994.

CASSEL, C. M., SÄRNDAL, C. E., WRETMAN, J. H. *Foundations of Inference in Survey Sampling.* New York: John Wiley & Sons, 1977.

COCHRAN, W. G. *Sampling Techniques, 3rd edition.* New York: John Wiley & Sons, 1977.

COOK, R. D., WEISBERG, S. Diagnostics for Heteroscedasticity in Regression. *Biometrika*, 1983, 70, pp. 1–10.

GRUBBS, F. E. Procedures for Detecting Outlying Observations in Samples. *Technometrics,* 1969, 11, pp. 1–21.

HÁJEK, J. *Teorie pravděpodobnostního výběru s aplikacemi na výběrová šetření* (Probability Sampling Theory with Applications to Sample Surveys). Prague: ČSAV (Czechoslovak Academy of Sciences), 1960.

HÁJEK, J. *Sampling from a Finite Population*. New York: Marcel Dekker, 1981

LEVY, P. S., LEMESHOW, S. *Sampling of Populations: Methods and Applications, 3rd edition.* New York: John Wiley & Sons, 1999.

PENROSE, R. A Generalized Inverse For Matrices. *Proceedings of the Cambridge Philosophical Society,* 1955, 51, pp. 406–413.

RAGHUNATHAN, T., LEPKOWSKI, J., HOEWYK, J. A Multivariate Technique for Multiply Imputing Missing Values Using a Sequence of Regression Models. *Survey Methodology.* 2001, 27 (1), pp. 85–95.

RAO, J. N. K. *Small Area Estimation.* New York: John Wiley & Sons, 2003.

Recent Publications and Events

New Publications of the Czech Statistical Office

Czech Republic in International Comparison. Prague: Czech Statistical Office, 2011.

Other Selected Publications

How´s Life? Measuring Well-Being. Paris: OECD, 2011.

Conferences

In order to celebrate the 100th anniversary of the Polish Statistical Association (timed to the celebration of the Polish Statistics Day) the ***Congress of Polish Statistics*** in Poznan, Poland, from 18th to 20th April 2012 will take place. Organizing Institutes: Polish Statistical Association, Central Statistical Office of Poland, Poznan University of Economics and Statistical Office in Poznan. More information is available at: *http://www.stat.gov.pl/pts/kongres2012/english/index.htm*.

The ***27th International Workshop on Statistical Modelling (IWSM)*** will be held in Prague, Czech Republic, from 16th to 20th July 2012. Organizing Institutes: Charles University in Prague (Faculty of Mathematics and Physics) and University of Economics, Prague (Faculty of Informatics and Statistics). More information is available at: *http://iwsm2012.karlin.mff.cuni.cz.*

Statistical Week at the Vienna University of Technology, Austria, will be held from 18th to 21st September 2012. Organizing Institutes: the German and Austrian Statistical Societies with the Association of German Municipal Statisticians. More information is available at: *http://www.statistische-woche.de/en.*

The Czech Statistical Office is pleased to invite to the ***International Marketing and Output Database Conference (IMAODBC 2012)*** which will be held at the Conference Centre of Průhonice Chateau, Prague, Czech Republic, from 8th to 12th October 2012. The deadline for the submission of abstracts is August 6, 2012. Potential participants should contact the organizing committee (by email at: *imaodb@czso.cz*) for access to the website which provides all the details they will need.

Papers
We publish articles focused at theoretical and applied statistics, mathematics and statistical methods, econometrics, applied economics, 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.

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.

The *Book Review* section brings reviews of recent books in the field of the official statistics. Reviews shall have up to 600 words or one (1) 1.5-spaced page.

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 (except headings).

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*. 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., 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, Economy and Statistics 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.

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. You will be informed by our managing editor about all necessary details and terms.

Contacts
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

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 369 (incl. postage) for the Czech Republic,
EUR 110 or USD 165 (incl. postage) 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: Jana Chocholoušová
Print: Czech Statistical Office

All views expressed in the journal of Statistika are those of the authors only and do not necessarily represent the views of the Czech Statistical Office, the Editorial Board, the staff, or any associates of the journal of Statistika.

© **2012** by the Czech Statistical Office. All rights reserved.

published by the Czech Statistical Office
ISSN 1804-8765 (Online)
ISSN 0322-788X (Print)