Statistika

Vol. 48 (1) 2011

Journal of Official Statistics

ČSÚ published by the Czech Statistical Office

ISSN 1804-8765 (Online)
ISSN 0322-788X (Print)

Dear Readers,

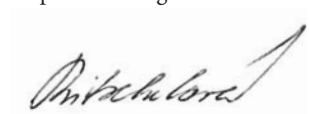
You are holding the new issue of the journal of Statistika published by the Czech Statistical Office. The journal of Statistika has been published since 1964. Over the years it has went through numerous changes following requirements of its readers, visions of the editorial board members, and professional interests of its authors. The journal editorial board has come to an agreement lately that the local character and format of the periodical are not sufficient to make it a suitable platform for discussions on the progress in the field of the modern statistics.

Our strategic aim is to create a platform enabling international and national statistical and research institutions to present the progress and results of complex analyses in the economic, environmental, and social spheres. Our mission is to promote the official statistics as a tool supporting the decision making at the level of international organisations, central and local government authorities, as well as businesses.

The uniqueness of the journal consists in the presentation of high quality analytical outcomes aimed at either supporting of decision-making processes or presenting of original and up-to-date methodologies. The value of any analytical result is directly determined by the quality of the data employed. In this respect, national statistical institutions producing these data can, undoubtedly, be taken as the most competent bodies for the comprehensive and relevant interpretation and use of available statistics. Concerning this, national statistical experts represent a significant group of the journal contributors.

On the other hand, the quality and information value of any analytical work is also conditioned by the maturity of the methodological apparatus applied. That is the reason other large group of our journal authors consists of representatives of the research and development sphere. The synergic connection of data producers and methodology creators, which one can see on the pages of our periodical, has a great potential to contribute to the debate and efforts in strengthening the bridge between theory and practice of the official statistics.

From now on the periodical will be published in English only. We have also invested a great effort into the development of its new layout, as well as a modern and user-friendly design. I wish the journal a lot of inspired readers and plenty of creative authors. I hope that papers presented on its pages will be valuable both for your everyday work and professional growth.



Iva Ritschelová
President of the Czech Statistical Office

CONTENTS

ANALYSES

METHODOLOGIES

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 organisations, 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.

Publisher

The Czech Statistical Office (CZSO) 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 CZSO 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 Praha 10 | Czech Republic
e-mail: statistika.journal@czso.cz | web: www.czso.cz/statistika_journal

The Role of the Official Statistics in the Context of the Global Crisis

Marie Bohatá[a] | *Eurostat, Luxembourg*

Abstract

Official statistics can be seen as a communication tool on variety of management levels and in different spheres of life of the society. The recent crisis has generated a number of challenges for official statisticians. Financial, economic, and political actors turned to statistics to describe the situation and to detect, assess, and even forecast these phenomena. Therefore, the main statistical consequence of the crisis is the recognition of the limits of the traditional approaches to statistical production and the importance to go beyond them. The first part of the paper shortly describes position and goals of the official statistics. In its second part the recent challenges are discussed with a particular stress on national accounts, the so-called Principle European Economic Indicators, as well as social statistics. The final chapter presents and analyzes general actions that are being taken at the European level in order to face these challenges and improve the overall quality of the official statistics.

Keywords

official statistics,
crisis,
challenge,
quality,
Europe

1 OFFICIAL STATISTICS

Official statistics can be seen as a communication tool that is indispensable for good public government, efficient business management, and also very helpful for assuring a democratic debate and facilitating societal life. Communicating about all sorts of phenomena requires a common language, which is universal in its ambition to transgress all sorts of borders, but has necessarily only a limited vocabulary and a limited grammar. It should be translatable into everyday speak as well as different expert speaks. It is a language that provides a societal perception framework which facilitates taking all sorts of decisions and enables to build a new or maintain an already established collective memory. To be understood, the language has to be based on conventions which are stable but simultaneously enable certain flexibility as the society develops.

Statistical service to society obliges us official statisticians to expose our work to a broad public debate and scrutiny. Our work is shaped by the political, legal, and administrative framework and our means (resources, rights, etc.) are ultimately determined by our political authority. Nevertheless, we should engage in a permanent dialogue with all parties concerned by official statistics (civil society, business community, scientific community, etc.) on the results of our work (the quality debate in the narrow sense), but also about concepts and methods (the quality debate in a wider sense). Only such a dialogue will allow us to stay in tune with the societies we are describing statistically. In this context it should be stressed, however, that a crucial precondition for the

[a] *Eurostat, Luxembourg, e-mail: marie.bohata@ec.europa.eu*

provision of relevant and objective statistics is professional independence and impartiality of official statistical authorities. At the EU level, these fundamental institutional features have been strengthened by the adoption of the European Statistics Code of Practice (2005)[1], the establishment of the European Statistical Governance Advisory Board (2008)[2] and the revision of the umbrella regulation for the production of European statistics (2009)[3].

2 CHALLENGES AND THE ROLE OF OFFICIAL STATISTICS

The recent crisis has generated a number of challenges for official statisticians. Financial, economic, and political actors turned to statistics to describe the situation and to detect, assess, and even forecast these phenomena. The role of official statisticians who have been confronted with an increased number of requests for relevant statistical data is, therefore, three-fold:

- to provide sufficient advanced warning;
- to monitor the impacts on economy, society and environment;
- to monitor political responses and their impacts.

National Accounts

Due to the importance of national accounting for all kinds of macroeconomic analyses, this tool has been considered crucial. Since its inception, national accounting – representing a conceptual frame of reference that enables economists from all over the world to engage in dialogue on common ground – has played an essential role in economic analysis.

Because of their wide range of uses, national accounts need to be solid and properly tailored to their purposes. In the context of the crisis, government debt and deficit are at the central focus point, as the recession forced many States to intervene directly in the economy. It, therefore, gave new legitimacy to national accounting, which had in fact gained prominence during the crisis of 1929 as a means of helping states to intervene effectively in economic affairs. Renewed interest in all of the macroeconomic analyses made possible by national accounting could, therefore, be expected. Nevertheless, we should thoroughly examine to which extent national accounts make it possible to explain and to monitor both the financial aspects and the real aspects of the crisis.

The crisis first arose in the form of a financial crisis and the question must be asked whether, given its very nature, the information provided in national accounts enables economists to really predict a recession of this type. Obviously, national accounting was primarily devised to measure activity in the real economy and is not well suited to monitor the financial sphere. While national accounting does touch on the financial sphere, this is essentially from the perspective of financing for the real economy. However, we can observe that the financial sphere has become quite autonomous from the real economy. It is, therefore, not surprising that national accounting did not make available specific information to foresee the financial crisis.

One of the essential questions is whether national accounting can better incorporate the financial aspects of the economy. The fact that financial activity has specific characteristics, such as e.g. the extreme diversity of financial instruments and their constant evolution and the extremely rapid pace of operations, makes it difficult. Thus, it would be unrealistic to attempt to use national accounting as the main analytical tool for the financial sphere. There is certainly room to improve the integration of financial issues in national accounting and this is being done, but it seems necessary that a series of relevant alert indicators be implemented outside national accounts to enable the authorities to anticipate more effectively any major problems looming on the financial horizon.[4]

---

[1] http://epp.eurostat.ec.europa.eu/portal/page/portal/quality/documents/code_practice.pdf

[2] http://epp.eurostat.ec.europa.eu/portal/page/portal/esgab/introduction

[3] http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2009:087:0164:0173:EN:PDF

[4] In this context a set of indicators developed at the EU level for the European monetary policy, so called principle European economic indicators (PEEIs) can be inspiring (see later).

National accounts provide a very useful framework but we should be also aware of their weaknesses. One of the major weaknesses is certainly the narrowness of their scope. When it comes to understanding a globalised economy, the main drawback of national accounting is the very fact that it is national. Worldwide accounts, with breakdowns by main regional areas, would certainly be extremely useful to economists in that they would highlight any potential imbalances or tensions between the different regions of the world. The conceptual framework to create this already exists: it would not be any different from the current framework used for the System of National Accounts (SNA) as recently revised. Actually creating these accounts would, of course, require the different international organisations to work together under the umbrella of the United Nations. Work to create worldwide accounts will have to be undertaken in cooperation between Eurostat, the IMF, the World Bank, and the OECD.

A key aspect of the statistical reaction has been to ensure the appropriate and proper consideration of the statistical consequences of the financial crisis on key statistics used in the European Union for administrative purposes and for the assessment of public finance.

As the financial crisis escalated from late summer 2008, governments and central banks in European countries have intervened through various operations in an effort to restore confidence in the financial system, at first to rescue single financial institutions in distress, and then through co-ordinated interventions broadly targeting all financial institutions, recognising the systemic aspect of the situation.

All these operations required an appropriate recording and treatment in statistical terms, notably in the framework of public finance statistics. A key requirement for the ESS in this area was to ensure the consistency across time and across countries of the statistical treatment of public interventions in full respect of the European System of Accounts (ESA95) rules. In this field, Eurostat, in co-operation with ESS partners, has closely monitored the public interventions and their implications for national accounts data, notably for the government deficit and debt statistics used for the excessive deficit procedure (EDP). The outcome of this methodological analysis provided the background information for defining the methodological treatment in national accounts, of these types of operations (see Eurostat Decision on "The statistical recording of public interventions to support financial institutions and financial markets during the financial crisis"[5] published on 15 July 2009). A new element to the existing approach has been introduced in the form of a supplementary table which reflects certain operations exposing governments to risk, but where the measurement is uncertain, as these operations have not crystallised yet. This approach provides transparency and shows a potential size of impact of operations addressing the financial crisis, for example if all guarantees provided by governments are called in the future.

Principle European Economic Indicators (PEEIs)

The crises has required official statisticians to provide not only a coherent view of the economy but also to deliver promptly key short-term economic indicators for monitoring the impact of the crisis and the impact of the measures taken by governments to remedy it. To meet these requirements, the European Statistical System (ESS)[6] offers through selected Principle European Economic Indicators (PEEIs) a continuously updated overview of these effects of the crisis at the EU level and in Member States, notably from the macroeconomic point of view. The PEEIs are 19 key short-term macroeconomic indicators available in a harmonised way for EU Member States,

[5] http://epp.eurostat.ec.europa.eu/portal/page/portal/government_finance_statistics/documents/FT%20-%20Eurostat%20Decision%20-%209%20July%202009%20_3_%20_final_.pdf

[6] The ESS is a partnership between national statistical authorities in Member states and Eurostat.

euro area, and EU (and when available for major economic partners) broken in 6 sections: consumer prices, quarterly national accounts, business, labour market, external trade and housing. They are disseminated via the PEEIs website and progress in their EU-wide compilation is regularly reported to the Economic and Financial Committee (EFC Status reports[7]). This project started in 2003 as a dialogue between users and producers to identify the best set of indicators needed for economic and monetary policy purposes at EU level, complemented by quality requirements, especially timeliness, and a methodological background. The PEEIs successfully evolved over time and to a large extent anticipated several requirements that became relevant during the crisis. Among those housing statistics and integrated quarterly financial and non-financial accounts for institutional sectors should be mentioned in particular. It is worth noting that currently this set of key short-term indicators is serving as a model for a global initiative (global principal economic indicators) at the UN level.

The crisis has also stimulated reflections towards a real-time monitoring and the construction of an early warning system. Currently these reflections focus on:

- further improvements of timeliness without a significant decrease of accuracy: more flash estimates and nowcasts;
- construction of new monthly indicators, filling the gaps of those official statistics only available on quarterly basis;
- ensuring the availability of long time series to fit purposes of analysts, as official statistics are often too short;
- compiling indicators to extract signals and to fill the specific gaps in official statistics: cyclical estimates, turning points dating and detection, coincident and leading indicators.

It should be stressed that the above-mentioned ideas require working with statistical and econometric techniques, as well as compiling new and composite indicators using existing statistics. Traditionally, such approaches have not been considered part of official statistics. We may observe, however, that in the recent years more and more statistical agencies have been involved in such kind of activities. Obviously, we – official statisticians – due to our deep knowledge of data and production systems, are in a privileged position. We can provide estimates and indicators based on statistically sound methodologies, transparent, replicable, and well documented procedures, and a high degree of objectivity.

Taking into account the nature of these outputs, we should communicate clearly to our users their specificities compared to traditional official statistics. Nevertheless, the information derived from such estimations (experimental statistics) could be very useful. It could complement official statistics and provide users with indicators enabling to get a real-time picture of the economic situation and with reliable early warning signals.

Social Statistics

It stands to reason that social impacts of the crisis have to be tackled with a high priority and this also needs an appropriate statistical support. For social statistics the monitoring dimension has been at the centre of interest. This reflects the lagging nature of social phenomena during economic downturns. While a lot of relevant information is available, there are some gaps and challenges linked especially to timeliness and flexibility. A comprehensive review of social statistics and their shortcomings at the EU level has been conducted and a new strategy for modernisation of social statistics elaborated together with Member States (see next chapter). The starting point is the expectation that the landscape for social statistics in the next decade will be characterised by increased use of registers and administrative sources alongside sample surveys, multi-mode data collections with a strong component of web-interviewing and enhanced data linking/matching approaches.

[7] http://epp.eurostat.ec.europa.eu/portal/page/portal/euroindicators/peeis/efc_status_report

3 GENERAL ACTIONS AT EUROPEAN LEVEL

In addition to domain-specific actions sketched above, two general initiatives have to be stressed:

- a critical analysis of methodological and practical aspects related to the statistical production process (re-engineering of the business architecture of the ESS);
- introducing a robust quality management for European Statistics.

The financial and economic crisis has highlighted the need to transform the production system of official statistics into a modern and efficient tool, flexible enough to cope with increasing or unexpected new requirements. The ESS has started to speed-up changes already under way in some Member States, and to rethink the production of official statistics through the modernisation of its business architecture. The approach aiming at vertical (the production chain) and horizontal (across statistical domains) integration is described in COM(2009) 404 final[8]. It was translated into an ESS strategy which was adopted by the ESS Committee in May 2010.

The challenge ahead of us is that official statistics will have to be produced as integrated parts of comprehensive production systems based on common technical infrastructure and a network of databases. For this we have to develop and establish joint structures, tools, and processes through collaborative networks. We have to think about network production at all levels and network communication in all directions, not just among producers of official statistics and between producers of official statistics and their political masters, but with all concerned and able to contribute. This should include not only the public administration or the world of business associations or NGOs, but also the scientific community.

The impact of the economic and financial crisis has also led to a more general reflection on the economic governance structure for the Euro area and the European Union as a whole. As a result of this reflection, the Commission adopted on 29 September 2010 a package of legislative proposals[9]. Broader and enhanced surveillance of fiscal policies, but also macroeconomic policies and structural reforms is sought in the light of the shortcomings of the existing legislation. New enforcement mechanisms are foreseen in case of non-compliance by Member States. Therefore, it is crucial to ensure that the decisions are based on statistical information which is produced under robust quality management.

Statistical information is a product resulting from statistical production processes operating across the entire ESS. Users of European statistics should be able to confidently use this information as an input to their own decision-making. These products should be fit for purpose, with users central in determining what constitutes quality.

The overall quality of statistical information on the European level is highly dependent on the appropriateness of the entire production process for statistics. In case data provided by Member States were of insufficient quality, this would have a negative impact on the quality of European statistics[10].

---

[8] http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2009:0404:FIN:EN:PDF

[9] COM (2010) 522 to 527:
http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2010:0522:FIN:EN:PDF
http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2010:0523:FIN:EN:PDF
http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2010:0524:FIN:EN:PDF
http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2010:0525:FIN:EN:PDF
http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2010:0526:FIN:EN:PDF
http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2010:0527:FIN:EN:PDF

[10] Following the weaknesses identified in the Greek case, extended audit-like powers in the EDP area were granted to Eurostat (EC Regulation 679/2010).
http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2010:198:0001:0004:EN:PDF

To help preventing such situations from happening, comprehensive and robust ESS quality management is needed. This systemic approach to quality will go hand in hand with the implementation of the above-mentioned vision for reforming the production method of European statistics, as it is expected to streamline the entire production chain.

We have to admit that, in the context of the crisis, the weaknesses in the quality of public accounting data and their statistical integration within the EDP reporting process were compounded by some weaknesses in the statistical governance arrangements in place. For the future it is, thus, crucial to move from a mainly corrective to a preventive approach regarding quality management for European statistics.

We can build on the existing framework comprised of the European Statistics Code of Practice, the European Statistical Governance Advisory Board (ESGAB) and the Regulation (EC) 223/2009 on European statistics, which provides solid foundations for an effective governance for the production of EU statistics, and address just those weaknesses which have become apparent on the basis of the recently gained experience. The new element in view of strengthening the quality of European statistics should be a risk-based approach taking into account also statistical implications of the legislative proposals, adopted by the Commission on 29 September 2010, on:

- strengthening the Stability and Growth Pact with prudent fiscal policy making;
- preventing and correcting macroeconomic imbalances;
- establishing national fiscal frameworks of quality, and in particular the need to have in place public accounting systems, subject to appropriate internal control and audit mechanisms, comprehensively and consistently covering all sub-sectors of general government;
- stronger enforcement[11].

This enhanced quality management will also take on board the conclusions of the Economic and Financial Affairs Council of 17 November 2010[12].

CONCLUSION

The worldwide nature of the crisis has underlined the global dimension of economic and financial phenomena, the integration of financial markets, and the rapidity of circulation of the information. All these aspects call for a global statistical view of the economic and financial reality, adequately supported by a statistical vision for the coming years. Therefore, the main statistical consequence of the crisis is the recognition of the limits of the traditional approaches to statistical production and the importance to go beyond them.

Even if the statistics concentrate on the recent past, this information is essential to enable economists and analysts to anticipate future scenarios. The crisis, thus, represents a challenge of strengthening official statistics by adapting relevant analytical tools at international level and always aiming at objectivity and credibility stemming from professional independence and impartiality of statistical services.

The ESS has acknowledged these challenges and is stepping up efforts to expedite the changes already under way, including the modernisation of the business architecture for the production of official statistics and further strengthening its governance including enhancing quality management.

[11] http://ec.europa.eu/economy_finance/articles/eu_economic_situation/2010-09-eu_economic_governance_proposals_en.htm

[12] http://www.consilium.europa.eu/uedocs/cms_data/docs/pressdata/en/ecofin/117762.pdf

References

EC. *European Statistics Code of Practice.* 2005.
<http://epp.eurostat.ec.europa.eu/portal/page/portal/quality/documents/code_practice.pdf>.

EC. *European Statistical Governance Advisory Board.* 2010.
<http://epp.eurostat.ec.europa.eu/portal/page/portal/esgab/introduction>.

EC. *Regulation (EC) No 223/2009 of the European Parliament and of the Council of 11 March 2009 on European statistics*, 2009. <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2009:087:0164:0173:EN:PDF>.

EC. Decision of Eurostat on deficit and debt: The statistical recording of public interventions to support financial institutions and financial markets during the financial crisis. 2009.
<http://epp.eurostat.ec.europa.eu/portal/page/portal/government_finance_statistics/documents/FT%20-%20Eurostat%20Decision%20-%209%20July%202009%20_3_%20_final_.pdf>.

EC. *EFC Status Reports.* 2010.
<http://epp.eurostat.ec.europa.eu/portal/page/portal/euroindicators/peeis/efc_status_report>.

EC. *Communication from the Commission to the European Parliament and the Council on the production method of EU statistics: a vision for the next decade (Com(2009) 404 final).* 2009.
<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2009:0404:FIN:EN:PDF>.

EC. P*ackage of legislative proposals adopted by the Commission on 29 September 2010 – COM(2010)522-527.* 2010.
<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2010:0522:FIN:EN:PDF>.
<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2010:0523:FIN:EN:PDF>.
<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2010:0524:FIN:EN:PDF>.
<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2010:0525:FIN:EN:PDF>.
<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2010:0526:FIN:EN:PDF>.
<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2010:0527:FIN:EN:PDF>.

EC. *Council Regulation (EU) No 679/2010 of 26 July 2010 amending Regulation (EC) No 479/2009 as regards the quality of statistical data in the context of the excessive deficit.* 2010.
<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2010:198:0001:0004:EN:PDF>.

EC. *A new EU economic governance – a comprehensive Commission package of proposals.* 2010.
<http://ec.europa.eu/economy_finance/articles/eu_economic_situation/2010-09-eu_economic_governance_proposals_en.htm>.

EC. *Conclusions on EU Statistics – 3045th Economic and Financial Affairs Council meeting on 17 November 2010.* 2010.
<http://www.consilium.europa.eu/uedocs/cms_data/docs/pressdata/en/ecofin/117762.pdf>.

Composite Leading Indicators Based on Business Surveys: Case of the Czech Economy

Vít Pošta[a] | *Ministry of Finance of the Czech Republic, Prague*
Vilém Valenta | *European Central Bank, Frankfurt*

Abstract

Consistent and countercyclical economic policies are conditioned on an accurate assessment of the position of an economy within the economic cycle and also on its prediction. The indicators of economic cycles represent a frequently used tool for monitoring and predicting economic cycles. In the paper we present the current practice in this area as it is has evolved over the years at the Ministry of Finance of the CR. At the same time, we point out the critical issues connected with the construction and interpretation of the indicators both in general sense but especially in the environment of the Czech economy. Finally, we present two composite leading indicators for the Czech economy and assess their prediction capacity in relation to the economic cycle.

Keywords

business surveys, economic cycle, leading indicators, short-term prediction

INTRODUCTION

Market economies typically experience fluctuations in economic activity. From the point of view of economic policy it is the mid-term fluctuations which are in the centre of interest and which are referred to as economic (business) cycles. Consistent and countercyclical economic policies are conditioned on a precise assessment of the position of the economy within the economic cycle and also on its prediction. The indicators of economic cycles represent a frequently used tool for monitoring and predicting economic cycles.

The paper presents the methodology of leading indicators as it has been applied at the Ministry of Finance of the Czech Republic (MF CR) for several years with only minor changes. At the same time we follow up on the recent discussion given in papers by Czesany and Jeřábková [1] and Czesaný and Jeřábková [2]. As far as older domestic sources are concerned we refer to Filáček [3]. The paper represents an update and an extension of Valenta [4]. First we summarize the basic principles on which the use of the indicators of economic cycles for the purpose of analysis and prediction rests. In the second part we focus on the methodology of construction of composite leading indicators, and, finally, we present the results of our analysis for the case of the Czech economy.

1 FUNDAMENTALS OF THE INDICATORS OF ECONOMIC CYCLES

An economic cycle may be defined as mid-term repetitive fluctuations in the deviation of economic

[a] *Letenská 15, Praha 1, phone: +420 257 042 725, e-mail: Vit.Posta@mfcr.cz*

activity from its long-term trend. It is characterized by alternation of expansions and contractions. To carry out a quantitative analysis, it is necessary to find a proxy series which captures the fluctuations of the economy – a reference series.

In principle, economic cycles can be predicted if the sources of the fluctuations lie inside the economic system or are represented by exogenous variables which themselves are predictible and also whose effects are predictable. The approach followed in this paper can be classified as a qualitative prediction of economic cycles. It is focused on the estimation of the position of the economy within the cycle, prediction of the trend of economic activity and, in particular, prediction of turning points of economic cycles.

Potential indicators of economic cycles must meet several criteria. They must have economic interpretation, they must exhibit cyclical behaviour themselves, they must show a statistically significant relationship to the economic cycle (reference series) and they must be regularly and timely available in sufficient quality. From the point of view of economic forecasting the most interesting group of indicators meeting these criteria are indicators that move ahead of the economic cycle. Such indicators can be used to forecast future trends and turning points in the level of economic activity. The predictive capacity of the indicators is, however, usually limited to a relatively short horizon.

For the purspose of predictions, it is useful to aggregate individual leading indicators into composite indicators. The aggregation may help to achieve a higher correlation of the composite indicator with the reference series and, thus, to limit both the risk of misinterpretation of the information given by the indicators and the number of false signals of turning points.

Czesany and Jeřábková [1] both give a standard classification of composite leading indicators as leading, coincident and lagging indicators and discuss the principles of their use. They also present the general approach to the construction of composite indicators. Hence, we do not discuss those general aspects of the issue in this paper and rather focus on the description of the practice with leading composite indicators at the MF CR.

Reliability of leading indicators is critically dependent on the quality and length of the input time series. The data for the Czech economy do not fully meet these requirements; this will be further pointed out below. Compared to developed market economies, the Czech time series are significantly shorter and suffer from structural breaks on account of rapidly changing economic environment. It is important to take account of these limitations when interpreting the composite leading indicators in the Czech Republic.

2 CONSTRUCTION OF A COMPOSITE LEADING INDICATOR

The very principle of the use of composite leading indicators is the existence of correlation between these indicators and the reference series. The procedure of constructing a composite leading indicator given in this paper makes use of statistical methods and is fully formalized. The assessment of the results is, however, qualitative and to some extent rests on expert's judgement. The method presented below does not allow a quantitative prediction of the reference series.

The procedure which is applied at the MF CR is based on the methodology developed by the OECD [6] and [7] while the method of composition follows the approach of the U.S. Department of Commerce. This method is also used by, for example, The Conference Board, Inc. [5].

The procedure of construction of a leading composite indicator can be divided into several steps. First, it is necessary to choose a suitable reference series, which will serve as a proxy of the economic cycle. At the same time, potential leading indicators needs to be identified. Second, it is necessary to adjust the input series for seasonality and calender effects and then to extract their trends. Third, the input series must be synchronized and standardized. When this is completed, in the fourth step, the individual indicators which will enter into the composite indicator are selected according to several criteria. In the next step the weights of the individual indicators are computed. Finally, the composite indicator itself is calculated.

As regards requirements for the reference series, it should capture the economic activity as widely as

possible. This condition is sufficiently met by the series of gross domestic product (GDP). At the same time the reference series should reflect the cyclical nature of economic development. In this regard the statistics of production, e.g. index of industrial production, may seem preferable. In the practice at the MF CR the series of GDP at constant prices has been selected. We discuss this choise later.

Candidates for individual leading indicators have been selected from the set of business survey indicators as published by the Czech Statistical Office (CSO). Business survey indicators are presented in the form of a balance of answers „it will improve" or „it will deteriorate" to questions concerning the economic situation of respondents. The third possible answer „it will not change" can also be given, but it does not affect the value of the indicator. Moreover, it may be also useful to monitor also correlation between the economic cycle and the answers „it will improve" or „it will deteriorate", not just the balance. We did not consider the composite sentiment indicators, which generally meet the given criteria but are themselves aggregates of individual indicators of which some are only weakly correlated to the economic cycle. During the process of construction we consider indicators from four sectors: industry, construction, trade and services.

The fact that the composite leading indicators used at the MF CR are based solely on business survey indicator, so-called soft data, makes it different when compared to the indicators based on both soft and hard data, which is a common practice. Recently such indicators have been presented in the paper by Czesaný and Jeřábková [2].

While the quarterly series of real GDP runs from the first quarter of 1996, as published by the CSO, the series of business survey indicators start in January 2003. This is due to the switch to the new classification of economic activity CZ-NACE (national version of NACE Rev. 2) from the previous classification OKEC (national version of NACE Rev. 1.1) in May 2010. The new data cannot be linked to the original series, thus limiting reliability of the composite leading indicator. However, it should also be noted that first years of business surveys, that have been lost, may have been of a lower quality due to the inexperience of the respondents. What is even more, one should take account of the fact that the nature of the economic cyle of the Czech economy changed during the transition from the transformation to post-transformation period. Therefore, the prediction capacity of surveys from 1990s for the curent situation would likely be questionable.

The series of real GDP is officially published as seasonally adjusted by the CSO and we use it in this form in the process of construction of the composite leading indicators. The other time series (almost 200) are seasonally adjusted using Census X12. In the cases where we find the estimated model within the framework of Census X12 unsatisfactory, we use Tramo-Seats instead. The analysis is carried out within the EViews package.

Estimation of the trend output of the economy, the output gap, has been a controversial issue ever since this question was raised. Two kinds of methods are at hand: purely statistical model or econometrical model which at least partially draws from the economic theory. For the purpose of this analysis we use a purely statistical method of a decomposition of the time series of GDP into trend and cycle: Hodrick-Prescott filter. The main advantage of this method is the fact that it is easy and fast. On the other hand the problem of „end points" is well documented. At this time when the economy has been recovering from a recent recession and we can see significant revisions of the time series of GDP done by the CSO, the estimate of the output gap is all the more a demanding task.

The set of potential indicators based on business survey comes in a monthly frequency while the series of GDP is published quarterly. To avoid losing the monthly information available, we intrapolate the quarterly series of GDP into monthly series. Due to the fact that we did not find a suitable baseline series to accomplish the time desaggregation of GDP series we chose a purely statistical method. The method rests on using quadratic polynomial. Based on every three consecutives elements of the original quarterly series a quadratic polynomial is estimated which is further used to estimate the monthly elements associated with the particular quarter so that the average of the monthly estimates of a given quarter equals the actual value

of the quarter in the original series. We apply the method within the EViews package.

The choice of the GDP as a reference series and the method of time desaggregation to monthly series is not straightforward and may require some defense. First, we assert that a typical candidate series for the reference series in the form of index of industrial production is not satisfactory for mainly two reasons: (i) index of industrial production does not capture the development of the whole economy while we later compose an indicator which reflects virtually all the sectors of the economy (industry, construction, trade and services) and also (ii) index of industrial production is, in the CR, a much less prominent indicator than GDP is. Second, the series of the corresponding relative cyclical component of the index of industrial production is much more volatile than the series of relative cyclical component of GDP (the volatility of relative cyclical component of industrial production index is more than double than that of GDP as measured by standard deviation on quarterly data from 2000Q1 to 2009Q3), which increases the risk of misinterpretation of the model and poses risks during the composition of the model in the first place. Third, the fact that monthly series of GDP does not exist is not a problem for interpreting the results of the model because (i) the monthly GDP series is used only in the phase of a selection of the individual leading indicators and (ii) regarding the intervals when data are published by the CSO and Macroeconomic Forecasts of MF CR are made, we are able to make qualitative forecasts based on CLI one or two quarters ahead depending on the actual lead of the model (three or five months, respectively).

Determination of the actual leads and the following synchronization are often based on the comparison of the average lead of turning points of the individual time series. Given the fact that in the CR the individual indicator series run from 2003, and therefore their turnng points can be assessed only against two turning points in the economic activity, it is not possible to use this approach as the primary method. The alternative method is cross-correlation. The visual analysis of the time series with the emphasis on their behavior near the turning points is, thus, due to the insufficient length of the series, just a supporting method.

The primary method of cross-correlation was used in almost all the 200 cases of potential leading indicators and it served as the main screen to select the candidates. In the next step we applied visual analysis with the focus on turning point prediction and we also excluded potential duplicities, i.e. the composite leading indicator should not include both three-month outlook of total demand in industry and three-month outlook of foreign demand in industry as it is highly probable than there are the same economic factors behind both indicators. The leads were not necessarily set at the highest correlation between the indicator and relative cyclical component of GDP. The difference in correlation at neighboring points of the series are very often negligible regarding the length of the series and relatively high instability of the relative cyclical component of GDP. Hence, we also paid attention to the results of visual analysis.

To make the time series comparable, it is necessary to transform them so that their levels of variability are the same. This is achieved by assigning weights to the individual indicators which are equal to inverse values of their variabilities. Hence, the individual components contribute the composite indicator equally. We do not apply weights in the sense of assigning some of the individual components a higher economic importance within the composite indicator.

As we have already pointed out, the selection of the individual components is based on the key criteria in the first place: sufficient correlation between the indicator and the relative cyclical component of GDP and timely availability on a regular basis. The other criteria are the stability of the lead and a capacity to predict the turning points.

It is a common practice to finally choose the individual indicators so that they cover all the sectors of the economy. On the other hand, there is another line of reasoning asserting that if the economy is governed by a particular sector, perhaps with a strong link to the world economy, the other sectors of the economy are under significant influence of this dominant sector. The composite indicator may then reflect especially this domi-

nant sector. In practice, we can see that there is a relatively stronger link between the individual indicators of industrial sector and the relative cyclical component of GDP and those links are at longer leads than in the cases of the other sectors. This, in turn, enables to construct a composite indicator with a longer lead.

The composition procedure itself is based on the calculation of symmetric month-on-month changes in the individual components. The monthly changes in the composite indicator are then computed as weighted average of the changes in the individual components where the weights are defined as inverse standard deviations of the respective components. The composite leading indicator is finally calculated by chaining its monthly changes.

3 COMPOSITE LEADING INDICATORS FOR THE CZECH ECONOMY

In this section, we will present two composite leading indicators: one with the lead of three months which is based on surveys from all sectors of the economy and one with the lead of five months which is based on the sector of industry only.

Both composite indicators draw from the new set of business survey indicators, i.e. after the switch to the new classification of economic activity CZ-NACE (national version NACE Rev. 2) in May 2010.

The first composite leading indicator covers all sectors covered by the business surveys, i.e. industry, construction, trade and services. If we want to cover all sectors of the economy, we are able to construct a composite indicator with the lead of three months. However, indicators from industrial sector show a prediction capacity for a longer horizen, thus alowing for a longer lead if only industrial surveys are taken into account. In such a case, the composite indicator has a lead of five months. Tables 1 and 2 show the composition of both indicators.

Both the all-sector and the industrial-sector composite indicators show correlation with the reference series at the level of around 88%. The key turning points in the GDP cycle were reflected by both of them. It is important to bear in mind that the actual shape of relative GDP cyclical component is still subject to possible changes due to data revisions. It is also important to note that the recent economic recession was set off by such a huge shock that it is captured by several dozens of individual indicators.

In the cases of such abrupt changes the interpretation of composite indicators is typicallly easy. The tough part is the assessment of the indicators within periods of normal fluctuations. The OECD recommends that some signal statistics are accepted, e.g. an annualized six-month change in the level of

Table 1 Composition of the indicator with the lead of 3 months

name of component			lead (months)	correlation (%)
sector	**individual indicator**	**type**		
industry	economic situation of firm	balance	3	76,0
industry	foreign demand level	balance	3	68,7
industry	employment: 3-month outlook	balance	3	81,1
industry	production capacitiy relative to orders: 3-month outlook	decrease	3	73,7
construction	total demand: 3-month outlook	balance	3	48,2
trade	economic situation of firm	balance	3	45,6
trade	number of employees: 3-month outlook	balance	3	55,3
trade	economic situation of firm: 6-month outlook	balance	3	57,2
services	number of employees: 3-month outlook	balance	3	72,3
services	economic situation of firm: 6-month outlook	increase	3	69,9

Source: Own calculations
Note: Type of the series may include a balance (difference between answers "it will improve" and "it will deteriorate", an increase (only answers "it will improve") and a decrease (only answers "it will deteriorate").

Table 2 Composition of the indicator with the lead of 5 months

name of component			lead (months)	correlation (%)
sector	**individual indicator**	**type**		
industry	economic situation of firm: 6-month outlook	balance	5	67,7
industry	solvency: 3-month outlook	balance	5	67,5
industry	total demand	increase	5	66,9
industry	economic situation of firm: 3-month outlook	balance	6	53,9
industry	number of employees: 3-month outlook	balance	5	76,5
industry	total demand: 3-month outlook	balance	6	54,8
industry	production capacitiy relative to orders: 3-month outlook	decrease	5	63,2

Source: Own calculations
Note: We report lags at which the individual indicator enters into the composite indicator. We also report the correlation coefficient between the individual series and relative cyclical component of GDP at the particular lag.

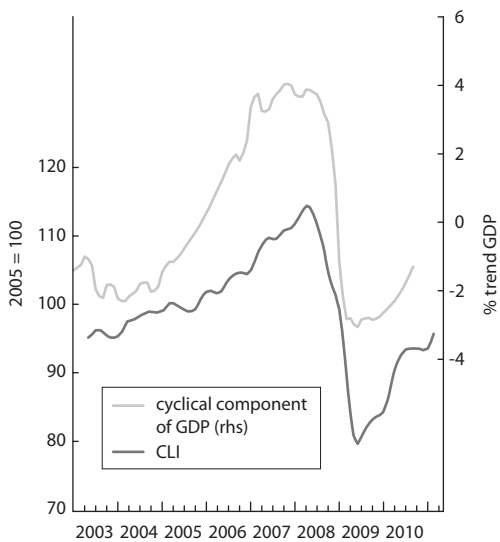
the indicator. A common rule is to take as a signal when there are continual changes in the level of the indicator in the opposite direction of the last trend for a given period of time. To some extent the interpretation of the indicators is, however, arbitrary and based on experience.

Both indicators signal a further increase in the economic activity in the near future. This may be interpreted as signal of quarter-on-quarter GDP growth leading to a gradual closure of the negative output gap.

CONCLUSION

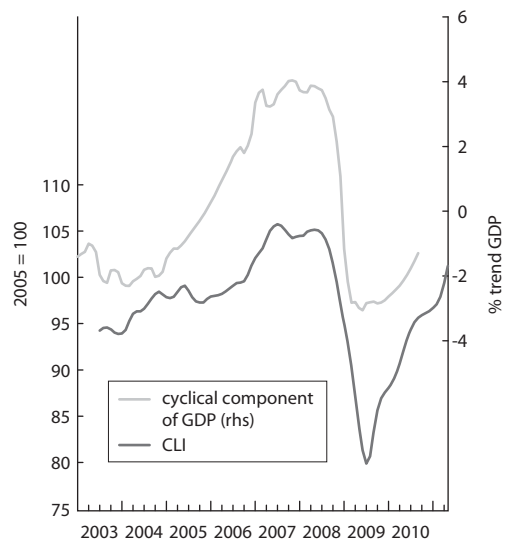
The database of the Czech economy provides indicators which possess the basic properties of leading indicators. Among others, some indicators based on business survey run by the CSO meet those criteria. The composite leading indicators we presented in

Figure 1 Composite leading indicator with the lead of 3 months



Source: Own calculations
Note: Synchronized with the GDP cycle.

Figure 2 Composite leading indicator with the lead of 5 months



Source: Own calculations
Note: Synchronized with GDP cycle.

this paper show a capacity to predict the key turning points in the economic cycle and show a high correlation with the relative cyclical component of GDP. Both composite indicators are based solely on the results of the CSO's business surveys. This fact makes our approach different from the usual practice which is inclusion of both the „soft" and „hard" data.

The available time series are of an insufficient length due to the switch to the new classification of economic activity CZ-NACE in May 2010. They are also less stable as a result of a changing economic environment as compared with more developed market economies. Therefore, it is necessary to interpret the composite leading indicators with caution.

Indicators of the economic cycle are widely used for a short-term prediction of the turning points in the economic cycle or underlying trends in the economic activity. They represent a relatively easy-to-implement forecasting method and it is reasonable to expect they will become even more reliable and useful in the future. The presented approach allows only a qualitative assessment. The possibility of the use of business survey indicators as a tool for a quantitative prediction of the reference series is a subject to the current research.

References

[1] CZESANÝ, S., JEŘÁBKOVÁ, Z.: Metoda konstrukce kompozitních indikátorů hospodářského cyklu pro českou ekonomiku. *Statistika*, 2009, č.1. s. 21–31, ISSN 0322–788x.

[2] CZESANÝ, S., JEŘÁBKOVÁ, Z.: Kompozitní indikátory hospodářského cyklu české ekonomiky. *Statistika*, 2009, č. 3., s. 256–274. ISSN 0322–788x.

[3] FILÁČEK, J.: *Odhad bodu obratu české ekonomiky za použití metody leading a coincident indikátorů.* Praha: ČSÚ, 2000.

[4] VALENTA, V.: Využití předstihových indikátorů pro krátkodobou predikci HDP. *Sborník prací účastníků vědeckého semináře doktorského studia Fakulty informatiky a statistiky VŠE v Praze*, 2007.

[5] *Business Cycle Indicators Handbook.* The conference board, Inc.: New York, 2001. <www.conference-board.org>.

[6] *Cyclical Indicators and Business Tendency Surveys.* OECD: Paříž, 1997. <www.oecd.org/dataoecd/20/18/1844842.pdf>.

[7] *Composite Leading Indicators: a tool for short-term analysis.* OECD: Paříž, 1997. <www.oecd.org/dataoecd/4/33/15994428.pdf>.

Selected Views on Fixed Assets in the Czech Economy

Drahomíra Dubská[a] | *Czech Statistical Office, Prague*

Abstract

Gross fixed capital formation (investment) in the Czech economy has been remaining the significant component of gross domestic product (GDP) from expenditure side during last roughly decade – in spite of the modest decrease in its share in GDP in nominal terms (from less than one third in 1995 to just about one quarter in 2008). In real terms its share in the GDP shows more likely stagnation (from 29% in 1995 to 28% in 2008). But correlation between GDP and investment in the Czech Republic is weaker compared to EU27. The increases of the various types of fixed assets in the Czech economy in constant prices of the year 2000 show clearly more dynamic growth in machinery and equipment in comparison with the dwellings and other buildings and structures. NACE K Real estate, renting and business activities participated most in the total state quantity of the gross capital formation. Although the state of the gross fixed capital formation was growing on average by about 2%, y-o-y, the investment in dwellings in constant prices of the year 2000 only stagnated. The article describes the reasons for that. The tangible assets in the households sector are mentioned, as well. The last part analyses current problems of investment in the Czech economy during economic crisis of 2009. Nevertheless, the fall of gross fixed capital formation in that year should not influence negatively the convergence of the Czech economy towards EU27 level.

Keywords

gross fixed capital formation, GDP, EU27, industry, households, dwellings

INTRODUCTION

Gross fixed capital formation (investment) in the Czech economy has been remaining the significant component of gross domestic product (GDP) from expenditure side during last roughly decade – in spite of the modest decrease in its share in GDP in nominal terms (from less than one third in 1995 to a quarter in 2008). In real terms its share in the GDP for the mentioned period shows more likely stagnation on the level between 28 to 30%[1]. The article describes development related to gross fixed capital formation and its states in the Czech economy and its institutional sectors. In the context of briefly mentioned development of investment in the EU27 countries, the main attention is paid to the area of investment and states of gross fixed capital in the economy of the CR in general (primarily tangible assets) and further in main institutional sectors and industries. From the point of view of material breakdown of assets, bigger attention is paid to the development of investment in dwellings (acquisition of multi-dwelling buildings, family houses and flats) and changes in states of gross fixed capital with the character of a dwelling

[a] *Czech Statistical Office, Na padesátém 81, 100 82 Prague 10, e-mail: drahomira.dubska@czso.cz*

as they are defined in the national accounts statistics. With regard to the strength of the share of the households sector in the material item of fixed assets, we analyse possible reasons for some surprising findings and there is also a view on profits and losses from holding of non-financial assets in the households sector and the entire economy. The last part deals with the necessity to finance investment from external sources at lower formation of gross national saving and there is also an attempt to deduce from the development of gross fixed capital formation in the year of crisis (2009) what impacts it will have on the future growth of the Czech economy.

The article has not been written with the intention to capture the issue of fixed assets and gross fixed capital formation in the economy of the CR in a whole extent. It focuses on some areas related to the topic.

Methodological bases:

- The term "Investment" in this text always means the item from the ESA95 classification: Gross fixed capital formation. Unlike this flow quantity, state quantities are mentioned as States of gross fixed capital.
- Industries in the article are observed according to the Industrial Classification of Economic Activities *(i.e. the national version of NACE Rev. 1)* valid until the end of the year 2008. Conversion of time series of national accounts items to the new Classification of Economic Activities (CZ-NACE, *i.e. the national version of NACE Rev. 2*) is assumed in the year 2011.
- Households sector consists of the segments: Households-individuals and Households-entrepreneurs. Major part of volumes of non-financial assets (fixed assets and inventories) belongs to the segment of Households-entrepreneurs.
- Data are primarily mentioned in prices of the year 2000; the nominal view results from current prices. Wherever growth rates data are given in prices of the preceding year it is mentioned in the text or graphs. Reference periods are years 1997 to 2007 if not stated otherwise (state quantities are available with some delay). In the last part observing the crisis year of 2009, data of the CZSO and Eurostat are updated as at 22 June 2010.

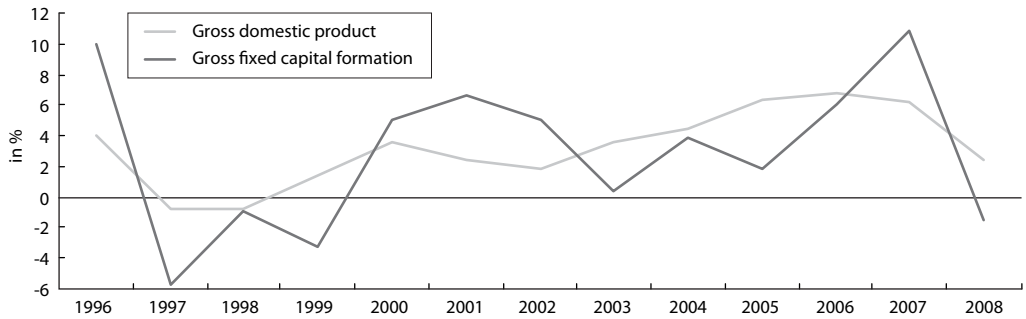
1 DEVELOPMENT OF GROSS FIXED CAPITAL FORMATION IN THE CR AND EU27

Gross fixed capital formation as a flow quantity reported in real terms (starting in the year 1995, from when CZSO time series are available, until the year 2008) a more volatile development than the gross domestic product (Graph 1). It is generally characteristic for growth rates of both features in most of the countries. The growth rate of investment as a growth multiplier reported for the period of 1996 to 2008 in real terms in the CR a lower correlation with the growth rate of the gross domestic product (correlation coefficient 0.66) than the correlation for the same period in the EU27 (0.85). A rather lower closeness of the dependence between GDP dynamics and dynamics of gross fixed capital formation was obvious in the mentioned period also in Slovakia (0.61), while it was higher in Germany (0.78) or Ireland (0.83).

In the Czech Republic during 1996–2008 a double-digit year-on-year increment of investment was reached only in 2007 (+10.8%). In 2008, gross fixed capital formation decreased by 1.5% compared to 2007, which was more than the actual stagnation for the EU27 on average (–0.3%). On the contrary, for the year 2009, in which the world financial crisis struck by transforming to an economic crisis with full power, Eurostat estimated for the EU27 economy a very marked drop of investment in the amount of –11.4%. The Czech Republic, according to this estimate, would have a lower fall of gross fixed capital formation than that average (–7.2%); however, in any case, it would be the deepest fall of year-on-year investment since 1995.

[1] However, this result is stated in the knowledge that structural shares in chained volumes do not have to be absolutely correct.

Graph 1 Gross domestic product and gross fixed capital formation in the economy of the CR
(year-on-year changes in %, from data in prices of the year 2000)



Source: CZSO

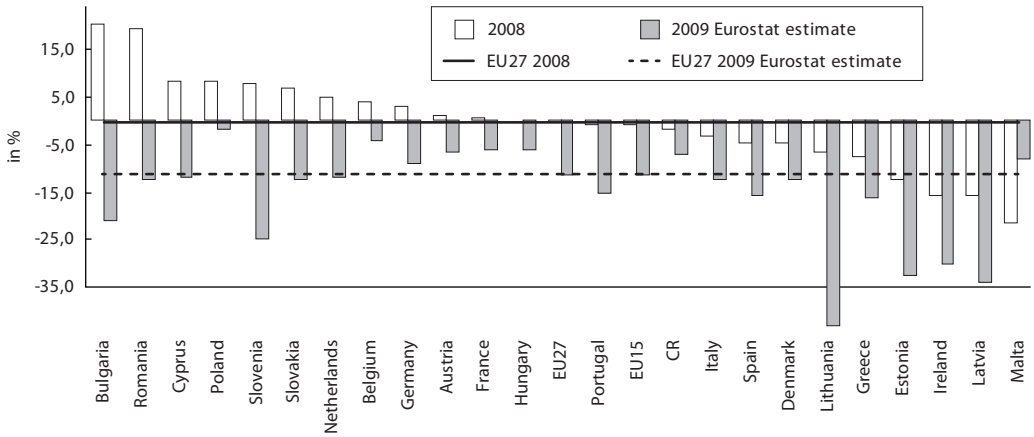
As it is depicted by Graph 2, marked decreases of investment for the year 2009 were expected primarily in the Baltic States, but also in Slovakia, Bulgaria, and, most of all, in Ireland. Since investment is somewhat under-dimensioned in the aforementioned countries, strong increase in gross fixed capital formation especially in the second half of the 1990's to the first years of the new millennium can be attributed namely to that circumstance.[2] For the year 2009 Eurostat's estimate for the fall of investment in each of the three Baltic States was more than 30%, y-o-y (Estonia –32.8%, Latvia –34%, and Lithuania even –43%). However, very surprising investment cycle is in Ireland (Graph 3), in which after a strong investment wave in the second half of the 1990's there was another one – much weaker already – in the first years of the new decade, and then in the time of European economic boom in 2004–2007 the pace of investment in Ireland already slowed down markedly and for the year 2009 Eurostat expected their fall also by more than 30% (–30.4%), which was a year-on-year decrease comparable to the drops in the Baltic States. Therefore, in some sense, we can speak about a "tiger's disease" affecting economies with extreme GDP growth, in which volatility of investment is proportional – and conditioning – to the economic cycle development.

In the estimate of a year-on-year change of gross fixed capital formation for the year 2009 Eurostat did not mention any increment for any single country of the EU27; the lowest drop was forecasted for Poland (–1.9%), which is the single country of the EU27, for which GDP growth was expected, and Belgium (–4%). The deepest fall of investment was predicted, as already mentioned, for Lithuania (–43%), at estimated fall of the economy by 18.1%.

As for the dynamics of gross fixed capital formation in the CR and EU27, synchronicity of curves for the years 1996–2008 is not clear (Graph 3). A very volatile development was reached by Slovakia; development of investment in Germany to great extent anticipates the curve of development of gross fixed capital formation in the entire EU27. In 2008, "old" countries of the EU (EU15) recorded a deeper investment fall than in the EU27 (–0.8% compared to –0.3%), for 2009 the estimate of the fall elaborated by Eurostat was roughly the same for both EU27 and EU15 (–11.5% in the EU15 compared to –11.4% in the EU27). What influences the deepness of the estimated fall is the double-digit loss of the pace of gross fixed capital formation in big countries, such as Italy (–12.2%) and Spain (–15.6%), but also in

[2] For example, in Latvia investment increased in 1998 by 61.4%, y-o-y, and grew with a double-digit growth rate until 2004 and 2005 (+23.8% and +23.6%, respectively).

Graph 2 Gross fixed capital formation in the EU27 countries
(in real terms, year-on-year changes in %, 2008 reality, 2009 estimate of Eurostat)



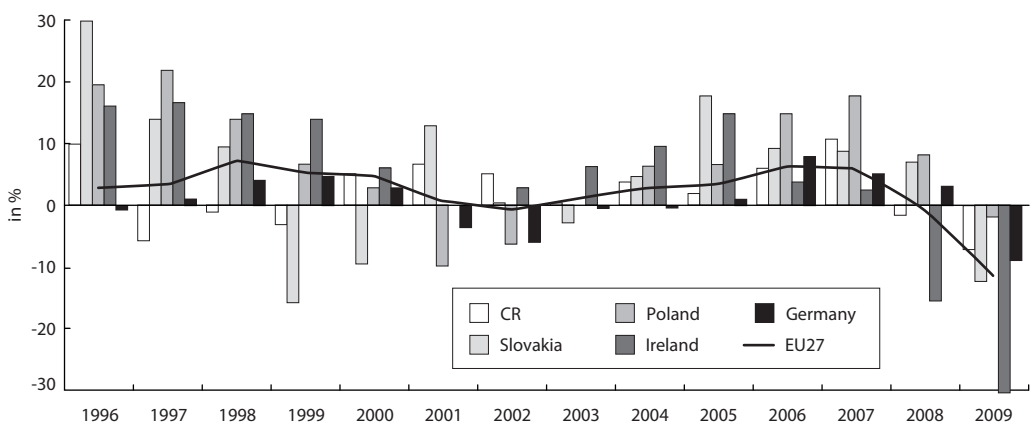
Source: CZSO

Germany (−8.9%). For Slovenia with the highest GDP per capita in the purchasing power parity from the ten countries that joined the European Union in May 2004, a very steep fall of investment was estimated (−24.8%), while in 2008 they still grew there by 7.7%, y-o-y. This fact would thus in 2009 significantly influence the development of Slovenian economy (−7.4%).

2 STATES OF GROSS FIXED CAPITAL IN THE ECONOMY OF THE CR WITH REGARD TO INVESTMENT TO TANGIBLE ASSETS

State of gross fixed capital in the economy of the CR in the end of the year 2007 compared to 1997 (not taking into account the influence of prices) increased for the mentioned period by about a fifth (+20.9%). In comparison to the year 1995 it was

Graph 3 Dynamics of gross fixed capital formation in selected countries
(year-on-year changes in %, 1996-2008 reality, 2009 estimate of Eurostat)



Source: Eurostat

even by more than a quarter (+26.2%). In 2007 it thus reached CZK 20.610 trillion compared to CZK 17.045 trillion in 1997 or CZK 16.327 trillion in 1995 (in prices of the year 2000).

From those volumes, the vast majority are gross fixed assets[3], the share of which moves in the long-term on the level of 99% of total volumes of the gross fixed capital in the economy of the CR. Their slight gradual decrease in state quantities in favour of intangible fixed assets (i.e. software and other intangible fixed assets), clear when expressed in percentage since 2001, however, takes place on the second place behind the decimal point: in 2001 tangible fixed assets made up 99.24% of the state of gross fixed capital in the economy of the CR and in 2007 it was 99.11%. This chapter analyses changes, which occurred in the period 1997–2007 in the structure of invested volumes of fixed capital, mainly at the item of dwellings and other buildings and structures, which are directly related to real estates development. Comparison with the year 1995 in some cases is to indicate that in the years 1995 and 1996 preceding to the monetary crisis there were rather massive investment increases with a strong differentiation by industry.

2.1 Differentiation of gross fixed capital by industry

CZ-NACE K "Real estate, renting and business activities" participated in 2007 in the total state of fixed capital in the CR by more than a quarter (27.9%); however, in 1995 it was still almost a third. For the period 1997–2007 states of gross fixed capital increased there with relatively low dynamics – the increment by 7.1% to CZK 5.743 trillion is the third lowest of all the industries of the CZ-NACE; however, there was an influence of high comparative basis. In absolute expression, during the years 1997 to 2007 the state of fixed capital increased by CZK 382 bn, which was more than a tenth of its total increment for the entire economy (CZK +3 565 bn) in the mentioned period.

The biggest increments of gross fixed capital were reached in the mentioned period in manufacturing (CZK +1 111 bn) and also in "Transport, storage and communication" (CZK +626 bn), which is logical, because this industry includes also transport infrastructure investment that is highly financially demanding.

Thus, transport, storage and communication contributed with 16.4% to the total state of gross fixed capital in the Czech economy in 2007, manufacturing with 14.2%, and public administration and defence with 10%. Compared to the year 1995, real estate industry decreased its share (−4.8 p.p.), while the share of manufacturing increased by about the same amount (+4 p.p.). The share of transport and public administration and defence in the total state of gross fixed capital in the economy of the CR remained basically unchanged in the period of 1995 to 2007 (+0.3 p.p. and −1.6 p.p., respectively).

Besides that, representation of individual sectors in states of fixed capital in the economy of the CR was also changing.

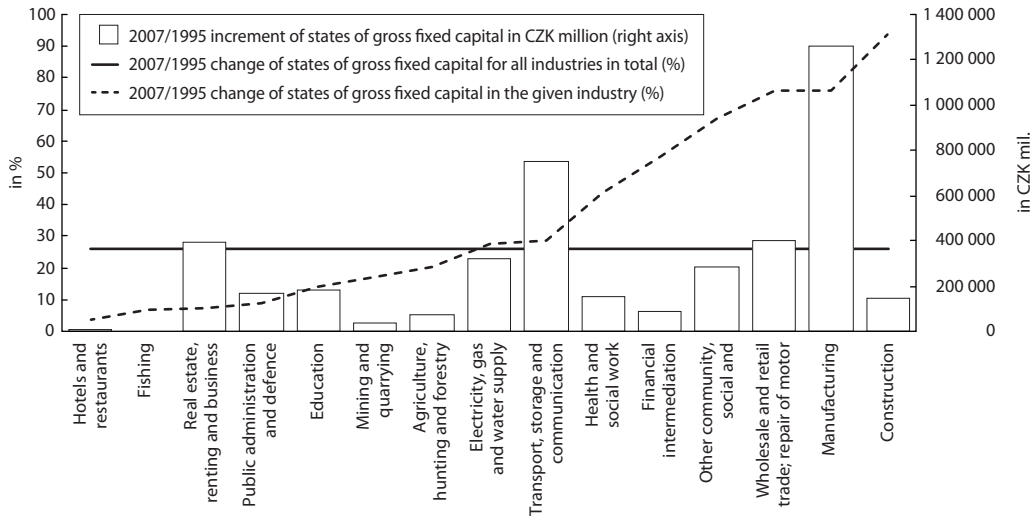
In addition to the already mentioned two industries, share of fixed assets in their total state in the economy decreased in industries of primary production, i.e. agriculture and forestry (−0.1 p.p.), although for the years 1995 to 2007 there was rather marked dynamics of fixed assets with the growth by a fifth). Also the share of mining and quarrying dropped (−0.1 p.p. as well, at the growth of states by more than 17%). Among other industries, rather significant losses in shares were recorded in education (−0.8 p.p.) and hotels and restaurants (−0.2 p.p.); in the latter case it was primarily due to slight dynamics in growth of fixed assets, because in hotels and restaurants the state of gross fixed capital increased in 2007 compared to 1995 only by 3.6%, which is the least of all industries of the Czech economy. Compared to the end of the year 1997, there is even an obvious fall by 0.5%.

2.2 Structure of tangible fixed assets

Industrial view in the part 2.1 does not capture a comparison how individual types of fixed assets

[3] I.e. in the methodology of national accounts dwellings, other buildings and structures, transport equipment, other machinery and equipment, and also cultivated assets.

Graph 4 Change of states of gross fixed capital in industries for the years 1995 to 2007 (change of states in CZK mil. in prices of the year 2000, or changes of states in % of data in prices of the year 2000)



Source: CZSO

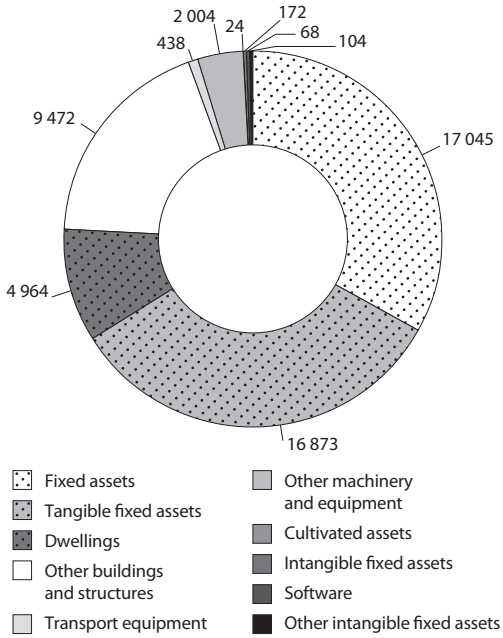
– or rather more narrowly specified types of tangible and intangible assets (it means how much from the total states belongs to machinery or construction investment or cultivated assets) share in the total states of fixed capital in the economy of the CR. This view is provided by the total summary of tangible fixed assets and their structure including the share of intangible assets.

For the economy of the CR it is typical that there is almost a 100% excess of tangible fixed assets – software and other intangible fixed assets represented in the end of 2007 only 0.89% of the total volume of fixed capital (CZK 183.7 bn). This share was slightly growing during the time (0.76% in 2001), however, for example, in 1995 to 1997 it was over 1%.

Intangible fixed assets grew on average according to the states in the end of the years 1997–2007 or 1995–2007 only by 0.9% and 0.8%, year-on-year, respectively. Compared to other types of fixed capital, this growth rate is relatively low. Therefore, it can be concluded that the low share of intangible assets (as logics suggests) results from very high dynamics of investment to machinery – and lower in construction investment – from the period from the year 2000. It is related to a wave of investment imports of companies belonging to foreign owners in the CR, who were equipping by them acquired production capacities.

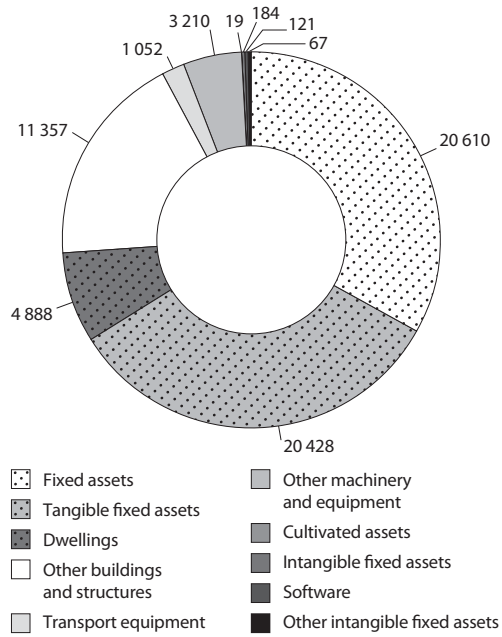
In the period of 1997–2007 (or 1995–2007), the fastest growing were the fixed assets observed in the system of national accounts as Transport equipment and Other machinery and equipment (besides them, tangible fixed assets comprise also Dwellings, Other buildings and structures, and Cultivated assets). Investment in transport equipment increased with the average annual growth rate of +8.3% (for the period of 1995 to 2007 even +9.2%). Investment to Other machinery and equipment were growing on average by 4.9% or 4.8% each year. Cultivated assets are the most volatile item of tangible fixed assets, which results from the character of commodities that are reported in this item. Their big increments in 1995 and 1996 resulted for the period 1995 to 2007 in an average annual increment by 2.5%; however, without that influence, i.e. for the period of 1997 to 2007, cultivated assets recorded a 2% y-o-y decrease on average. Graph 5 compares states of individual types of fixed capital in the CR in 1997 and 2007, which are the results of investment activity for that period.

Graph 5a States of gross fixed capital by type of fixed capital in the year 1997
(in CZK bn, constant prices of the year 2000)



Source: CZSO

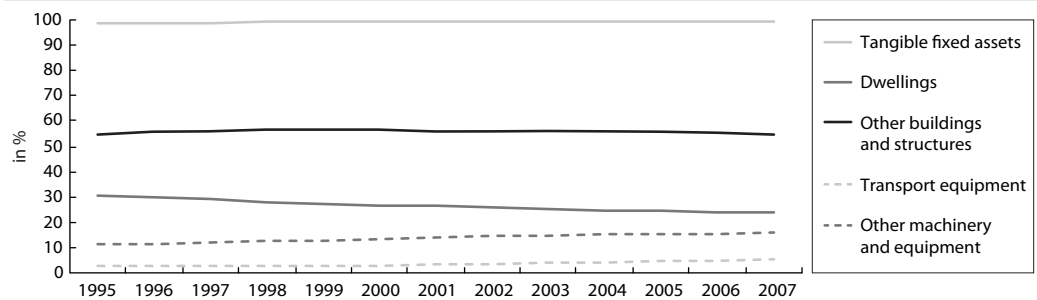
Graph 5b States of gross fixed capital by type of fixed capital in the year 2007
(in CZK bn, constant prices of the year 2000)



Source: CZSO

Despite strong growth of tangible fixed assets of the type of machinery and transport equipment, however, tangible fixed assets of the type of dwellings and other buildings and structures are still big as for their volume. Although there is primarily in the years 2005 to 2007 a strong construction-investment activity in the CR, which is underlined by co-financing of the projects from the structural funds of the European Union, the share of tangible fixed assets of the type of construction investment in the total states of fixed capital in the case of dwellings is decreasing, among other buildings and structures it is stagnating (Graph 6).

Graph 6 Share of individual types of tangible fixed capital in fixed assets in total
(in %, states in CZK bn, expressed in prices of the year 2000)



Source: CZSO

As for industries, strong dynamics of machinery investment and investment in transport equipment – i.e. much higher than average – is characteristic for manufacturing industry. There, thanks to the dynamics, the overall state of gross fixed capital increased in real terms in the year 2007 compared to 1997 by 62% (compared to 1995 even by 76%). This growth in manufacturing industry was during 1997–2007 three times faster than the growth for the entire economy (+21%).

Investment to machinery and equipment were, mainly in the first years after the year 2000, related to the inflow of investment from abroad. Their biggest recipient, manufacturing industry, participated in the year 2007 in the total volume of these external sources with 42.2%. The share of foreign investment directed to manufacturing industry moved during the period of 1997 to 2007 between 35 to 45%. A marked fluctuation in the year 2003 can be explained by the fact that by the influence of a steep decrease of the total volume of direct investment to the CR in the given year their inflow to the manufacturing industry was higher than the overall level for the CR (the volume of direct investment was influenced primarily by a buy-back of the stake in the joint stock company from the industry of telecommunications).

The manufacturing industry is a typical recipient of foreign investment primarily in "material" form, which can be with some simplification understood as fixed capital, while in some other industries it can have in major part the form of inputs to the registered capital of companies and thus strengthening of own sources of companies, which as a result does not need to have the form of tangible fixed assets.

2.2.1 Tangible fixed assets of the dwellings- and other buildings and structures type

During the years 1997–2007, the share of fixed assets of dwellings type in the state of total fixed assets in the CR was reduced according to prices of the year 2000. In 1997, fixed assets related to dwellings participated in the overall state of fixed assets in the CR with 30.6%, while in 2007 it was already only 23.7%. In volumes (in prices of the year 2000) states of fixed assets of dwellings type decreased even to CZK 4 888.3 bn in the year 2007, compared to CZK 4 963.9 bn from the year 1997. On the contrary, investment to other buildings and structures caused that their states increased for the mentioned period from CZK 9 471.7 bn to CZK 11 357 bn. How this development – paradoxical on the first sight – can be explained, when construction of houses and flats was experiencing mainly in the second half of the reference period an obvious boom?

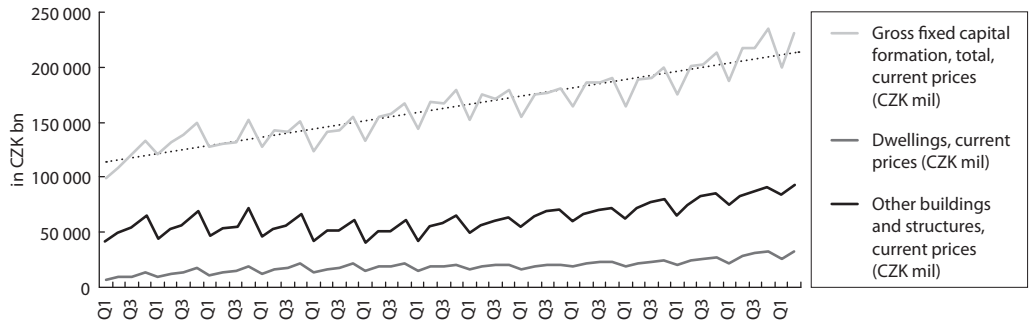
Rather surprising development of states of gross fixed capital of dwellings type can be partly explainable by movements in the dwelling stock of the Czech Republic and its evaluation. As for the breakdown by institutional sector, this period is characteristic by a relatively big increment of states of fixed assets in the households sector (privatisation); however, on the contrary, in the remaining two institutional sectors, i.e. non-financial corporations and government sector they were decreasing. In total, tangible fixed assets of dwelling character report virtually stagnation for the period of 1997–2007 (expressed in prices of the year 2000). Life span of buildings of dwellings type is calculated for 80 years and after such structure exceeds the limit it is eliminated from the statistics, which is a methodological influence, related to gradual depreciation.

Another result follows from the data in current prices. States of fixed assets of dwellings type in nominal expression report on average for years 1995 to 2007 an average annual change of +5.6% with stronger dynamics in the period from the second quarter of 2005 to half-year of 2008 (+6.7%). Tangible fixed assets of dwellings type reached stronger nominal increments for the period from the beginning of the year 2000 until the end of the first half-year of 2002 and the highest average increments were recorded in the period with high inflation (nominally +17.5% in the year 1996).

Total fixed assets in the CR grew on average by 2%, y-o-y, for the years 1996 to 2007; states of fixed assets of dwellings type, on the contrary, decreased on average by 0.2%, year-on-year (calculation from constant prices of the year 2000).

Increments of individual types of fixed assets in conditions of the Czech economy expressed in prices of the year 2000 clearly point at more dynam-

Graph 7 Development of tangible fixed assets (by quarter, 1995 to 2008, CZK bn in current prices)



Source: CZSO

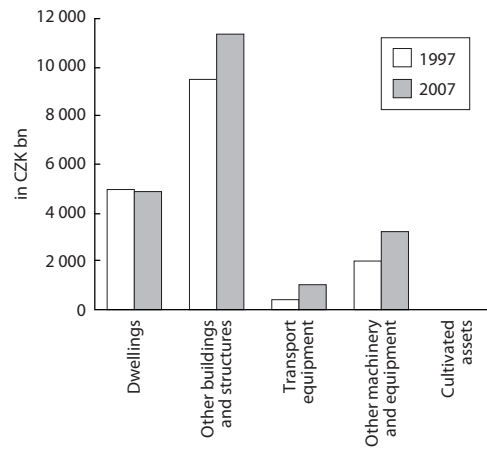
ic growth of investment in machinery and equipment compared to construction investment.[4] In total, states of total fixed assets in the CR increased for the years 1996 to 2007 on average by 2%, year-on-year; at transport equipment this growth was 8.3% and at other machinery and equipment on average by 4.9%.

On the contrary, states of fixed assets in the case of other buildings and structures grew on average by 2%, i.e. roughly the same as states of fixed assets for the economy as a whole. As for fixed assets of dwellings type, however, there was reported a decrease by 0.2% in the year-on-year growth rate; for the period of 1998 to 2007 it was by 0.1%. Year-on-year decrease of states of investment to dwellings occurred (again in prices of the year 2000) in 1997 to 1999 and further in the years 2002 to 2004.

2.3 Tangible fixed assets in the households sector

Czech households accelerated their investment in new dwellings (according to calculations from current prices) starting from the year 2001. Nominally, development of investment of households to dwelling (housing)[5] in the CR during the years 1995 to 2006 was not closely related to the development of investment to dwellings for the economy as a whole (according to net acquisition of tangible fixed assets) – households pushed the growth of investment to dwellings for all institutional sectors approximately since 2003, when the development curve in both cases is almost the same. However, a closer mutual dependence can be found in investment to new dwellings (Graph 9). It is clear from the graph, that

Graph 8 States of fixed assets in the end of 1997 and 2007 (in CZK bn, in prices of the year 2000)

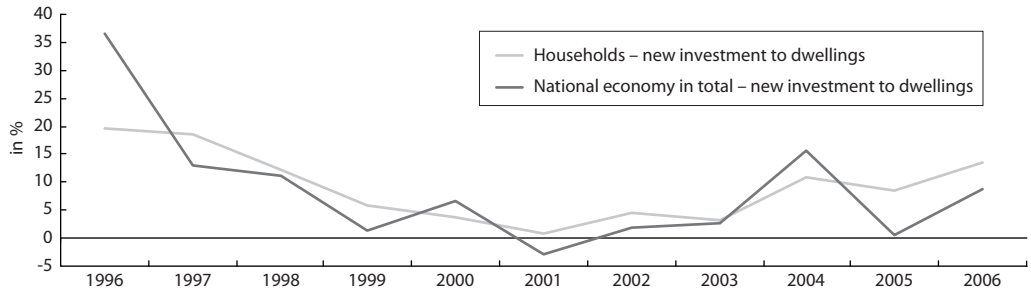


Source: CZSO

[4] Nevertheless, depreciations could relativize this deduction.

[5] In this part of the article, in relation to acquisition of tangible fixed assets – focused by type of fixed capital on the Dwellings item – we use a simplified term "investment to dwellings".

Graph 9 Investment to new dwellings
(acquisition of new tangible fixed assets, year-on-year changes in % from the data in current prices)

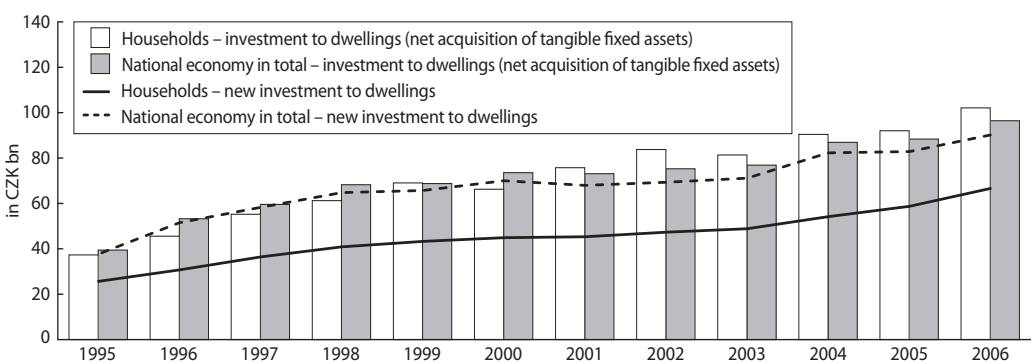


Source: CZSO

acquisition of new tangible fixed assets was slowing down the growth in the period from 1995 to 2001, year-on-year; however, since that time an increase can be recorded mainly in investment of households to dwellings, i.e. new fixed assets of the type of houses and flats. Nevertheless, the decreasing dynamics during individual years in the mentioned period can be partially explained by a price development of real estates (year-on-year changes are calculated from the data in current prices).

During the years 1995–2006 about two thirds of the value of acquisition of new flats and houses for the CR as a whole[6] belonged to households – the lowest share in acquisition of new houses and flats belonged to households in 1996 (59%), the highest in the period of a boom on the market with flats in 2006 (73.9%). In total, for new and used flats and houses, in nominal expression, investment of this sector to dwellings increased in 2006 three times compared to 1995 according to data in current prices on net acquisition of tangible fixed assets of dwellings type. While in 1995 the value of houses and flats acquired by households was CZK 37.3 bn in current prices, in 2006

Graph 10 Investment to dwellings and from that to new dwellings
(net acquisition of tangible fixed assets, from that of new fixed assets, in CZK bn, current prices)



Source: CZSO

[6] The rest of the acquired value belonged, first of all, to developer companies and real estate agencies.

it was CZK 102.1 bn. Starting in 2001, the value of net acquisition of tangible fixed assets of dwellings type for the households sector began to even exceed the value for all sectors, i.e. for the entire Czech economy. The reason for that was the reported negative value of net acquisition of tangible fixed assets related to dwellings at non-financial corporations, in which the value of acquisition of the new and used was lower than the value of sales (privatisation).

3 REVENUES FROM NON-FINANCIAL ASSETS

It can be said in general that even without a transaction taking place the value can change also thanks to a change in evaluation of an asset or liability. Thus, a profit or loss of a business or sector from holding of them can be quantified. For example, holding of assets in the form of real estates itself (provided that the real estate does not serve to own usage, i.e. housing or doing business) can generate profits in the form of a rental, but in every case also profits (or losses) resulting from simple holding of a given real estate (non-financial asset). Generally, in the national accounts system it has to be distinguished whether the changes in the volume of assets and liabilities take place due to transactions or because of other changes. The account of other changes is further broken down to the account of other changes in the volume of assets and the account of revaluations. On the account of revaluations the change of the volume of a given asset caused by the change of its price is captured; thus, a nominal profit (or nominal loss), which results from holding of the given asset by a business can be quantified.

Changes of prices of assets and liabilities at all businesses influence their wealth. From the aforementioned it results that this change occurs also on condition that businesses or households (or government or financial sector, but also non-residents) do not make any transactions with their assets or liabilities, i.e. they do not purchase or sell them. Thus, by a mere change of prices, "value" of these assets and liabilities changes, which has an impact on the level of wealth of the mentioned businesses.

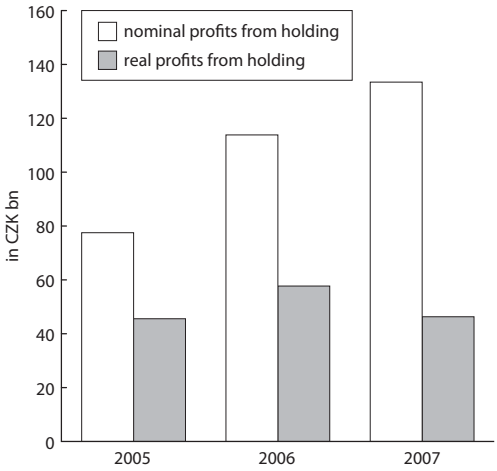
If a household or a business or any other entity (in the national accounts terminology: institutional sector or sub-sector) acquires some type of an asset (i.e. it acquires a real estate, deposits money in a bank, saves in a pension fund or on an account of a life assurance, purchases allotment certificates or invests in shares, buys someone's receivable, and the like) we speak about transactions. If, for example, they borrow money, conclude a leasing contract, and others – they thus undertake an obligation, it is also a transaction. The value of assets, both financial and non-financial, the same as the value of liabilities, can be quantified. Nominal profit or loss from their holding is generated also by the price development.

As it is stated in the results processed by the Annual National Accounts Department of the CZSO[7] published in October 2008 and capturing the period of 2005–2007, real profits or losses from holding (assets and liabilities in general) can be calculated. When there is faster growth of prices of a given asset (as for the subject dealt with in our article it is mainly a non-financial asset) in comparison to the growth rate of the price level[8] – the holder of the asset reaches in such case a real profit (i.e. a profit from holding). On the contrary, if a price of a given asset is growing slower than the overall price level in the economy (or in the relevant price range), then the holder of the asset is losing or records a real loss (i.e. a loss from holding). The Graph 11 shows nominal and real profits from holding of non-financial assets for households sector in 2005–2007. It is clear

[7] See http://czso.cz/csu/redakce.nsf/i/zisky_ztraty_z_drzby_dopady_pohybu_cen published on 14 October 2008 by the Annual National Accounts Department of the CZSO (Czech only).

[8] Expressed by an implicit deflator of final national uses excluding changes in inventories, which contains price change of expenditure for final uses and gross fixed capital formation. This deflator is recommended by the ESA95 methodology (paragraph 6.45).

Graph 11 Nominal and real profits of the households sector from holding of non-financial assets (in CZK bn)



Source: CZSO

from the graph that while nominal profits from holding of non-financial assets in this sector for the mentioned three years were growing, the real profits from their holding in 2007 decreased, year-on-year. It can be expected that data for the year 2008 and, primarily, for 2009 will brought with regard to the decrease of prices of houses and flats as fixed assets of dwellings type another fall of profits from holding or even losses from holding.

An important feature of profits from holding is that they strengthen the net worth, i.e. they increase wealth of businesses (analogically, losses from holding are reducing that wealth). Real profits from holding of non-financial assets consist of the difference between nominal and neutral profits from their holding. Results for the years 2005–2007 show that non-financial assets in the Czech economy strengthened the total net worth in the national economy in a dynamic trend – in 2005 thanks to the profits from non-financial assets the wealth was strengthened (as expressed by growth of the net worth in the economy) by CZK 315 bn; in 2006 it was by CZK 457.8 bn.

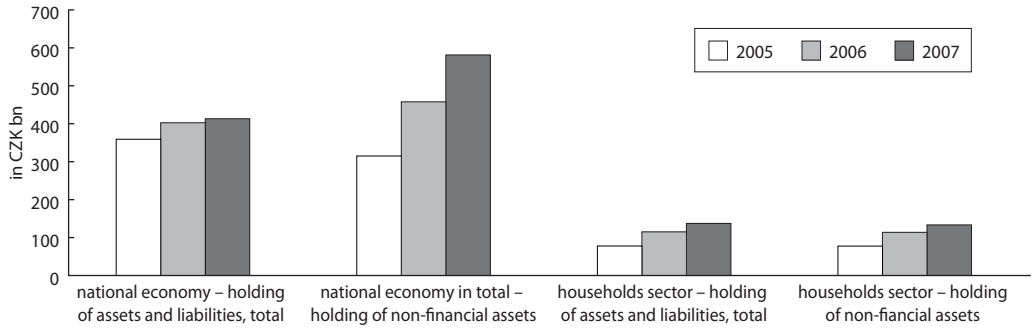
Individual institutional sectors differ not only as for the volume of nominal profits from holding, which is logical, but also in their dynamics. While in the year 2006, for example, the net worth of the households sector increased faster (thanks to nominal profits from holding of assets and liabilities) compared to 2005 than for the entire national economy, namely for non-financial assets and all their assets and liabilities (Graph 12), in 2007 it was not valid any more. For example, non-financial assets of corporations recorded in the year-on-year comparison by a third higher nominal profits from holding and the Czech economy as a whole an increase by 27%. Nevertheless, in the households sector the nominal profits from holding of non-financial assets increased only by 17% in 2007 compared to 2006.

It is different when looking at real profits (or losses) from holding of assets and liabilities, i.e. with depreciation by changes in the price level in the economy, namely via the already mentioned implicit deflator of final uses[9]. According to the calculations of the Annual National Accounts Department of the CZSO, real profits from holding of assets and liabilities in 2007 compared to 2006 in the economy as a whole significantly decreased – from CZK 148.3 bn to CZK 58.6 bn. There was a drop already in the preceding period (however, not that big) because in 2005 real profits from holding of assets and liabilities for the entire economy were CZK 214.8 bn.

At non-financial assets, the decrease of real profits from holding for the entire national economy in 2007 compared to 2006 was only small (from CZK 186.8 bn to CZK 172 bn) – in the year 2006 compared to 2005 these profits even increased (CZK 160.6 bn). With the exception of the financial sector and the government sector, other insti-

[9] However, it may not always reflect the actual amount of the profits or losses for individual institutional sectors and a better result could be reached by calculation for changes of price levels in price ranges pertaining to those institutional sectors – i.e., for example, at non-financial corporations via changes of production prices, at households by change of the consumer price index. Those calculations, however, are not available now.

Graph 12 Changes of net worth thanks to nominal profits and losses from holding (in CZK bn)



Source: CZSO

tutional sectors recorded a decrease of the profits from holding.

4 SOME TOPICAL PROBLEMS OF THE INVESTMENT DEVELOPMENT IN THE CZECH REPUBLIC AND EU27

The crisis development of the Czech economy in 2009 entailed – similarly as in all European countries except for Switzerland – a slow-down of investment activity. Decisive for that was an anxiety of companies to focus on development at a decrease in the demand, in the time of existential problems, credit tightness and decrease of production capacities utilization. At the same time, in reaction to the crisis, during individual quarters of 2009, gross saving was falling, year-on-year, in absolute majority of European economies.

4.1 Gross fixed capital formation and gross national saving

In 2009, gross fixed capital formation in the Czech economy markedly decreased for the first time since 1995; however, already in 2008 there was a year-on-year decrease. It has been up to now the third year-on-year decrease of investment since 1995 (in 1999 the gross fixed capital formation also dropped, however, with an absolutely low volume of CZK 148 million, which is characteristic rather for a y-o-y stagnation). Compared to the year 2008 with the volume of investment of CZK 883.2 bn in current prices,[10] gross fixed capital formation in 2009 dropped to CZK 822.1 bn, i.e. by 6.9%. The y-o-y drop of investment in 2008 was 0.8%.

An international comparison shows that the pace of investment according to year-on-year growth dynamics is the highest among countries on the lower level of economic development. According to data of Eurostat, the average growth rate of gross fixed capital formation for 2001–2009 in the Czech Republic (+7.1%) was roughly three times higher than in the EU27 (+2.2%) or EU15 (+1.8%). Even higher dynamics was recorded, for example, in the Baltic States (Estonia +9.6%, Latvia +10.6%, Lithuania +9.8%), Romania (+18.9%), Bulgaria (+17.9%) or Slovakia (+11.8%) – data are provided in nominal expression with regard to the need of other comparisons in this chapter, real growth rates are analysed above in Chapter 1. The need of investment in those countries corresponds to the fact that they are generally under-dimensioned for a long-term. On the contrary, negative average annual dynamics of investment in current prices was reported for the years 2001–2009 by United Kingdom (–1.1%), Portugal (–0.5%) but also, for example, Germany, in which, however, the average result should be interpreted rather as a stagnation

[10] Values at current prices are used in this sub-chapter to compare investment and gross national saving, which is stated only in nominal expression (difficult possibility to deflate).

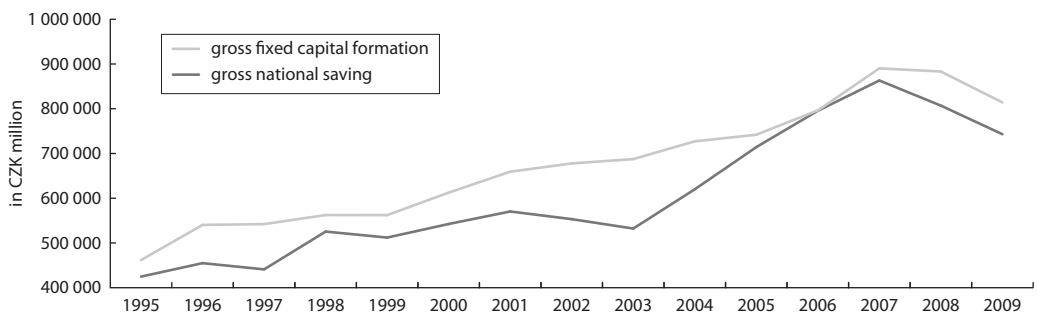
(−0.1%). The average decrease was in general significantly influenced by y-o-y falls in the year 2009 (in the case of United Kingdom even by 23.1%). Even in a much higher extent, the average drop of investment was recorded for the years 2001–2009 by big world economies (the United States −3%, Japan −5.5%), in which, however, during much of that period, year-on-year decreases occurred. Especially in Japan, with its deflation conditions, there was y-o-y growth of investment only in two years of the mentioned period.

The start of a distinctive slowdown of the growth rate of the Czech economy resulting in a technical recession (the economy of the CR was decreasing quarter-on-quarter from the fourth quarter of 2008 to the second quarter of 2009) was preceded by a decrease of investment. Both in real and nominal expression, there were year-on-year decreases of gross fixed capital formation (also by the influence of the comparison basis after big investment actions) starting already in the second quarter of 2008 (−1.2% or more precisely −0.9%) with continuing marked falls until the third quarter of 2009 (−11.7% or more precisely −10.8%), which then started to mitigate. The GDP dropped both in real and nominal terms from the fourth quarter of 2008 or rather the first quarter of 2009. However, also the comparison basis had an influence on the "sooner" y-o-y fall of investment according to the annual data in comparison to the GDP dynamics, because the year 2007 was, on the contrary, very "strong" as for investment – the gross fixed capital formation increased in it by 10.8% in real terms, by 11.8% in nominal terms, i.e. with the highest rate for the reference period of 2001–2009.

The ability of an economy to finance gross fixed capital formation from gross national saving points at the extent of dependency on external financing of investment, which is then reflected in the balance of national current transactions with the rest of the world. The years 2008 and 2009, as it is clear from Graph 13, brought a decrease of gross fixed capital formation as well as fall of gross national saving in the economy of the Czech Republic.

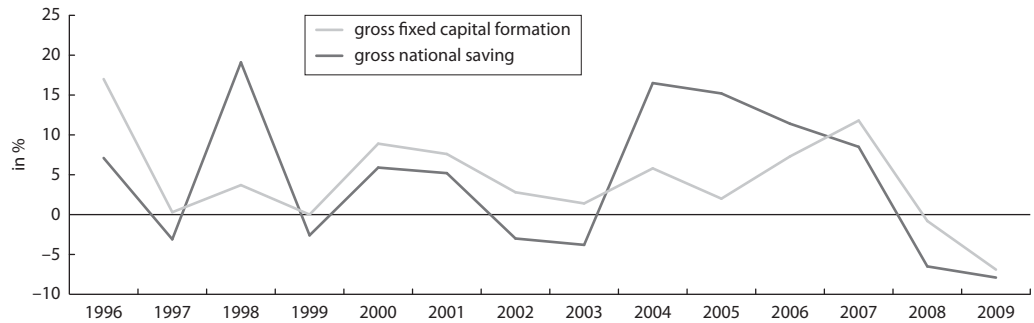
The necessity to finance investment from external sources in those years – with the beginning and further continuation of the economic crisis – thus again soared. A decrease of gross national saving was logical in relation to the year-on-year drop of gross disposable income in the year 2009; at the same time, final consumption expenditure continued to grow, although with very reduced dynamics compared to the rates of the previous years. In 2008, still, final consumption expenditure increased with the highest growth rate of the new decade (+7.6% in nominal terms), while the gross disposable income in the Czech economy lost already much of its dynamics in that year and, compared to the increments from the time of boom, it grew only with a half pace (+3.9%). The decrease of gross national saving in the Czech economy in 2008 and 2009 can be explained namely by different dynamics of both quantities. Thus, in combination with the dynamics of investment (Graph 14), the need to finance investment again in a larger extent

Graph 13 Gross fixed capital formation and gross national saving (in CZK million)



Source: CZSO

Graph 14 Dynamics of gross fixed capital formation and gross national saving
(year-on-year changes in %, from data in current prices)



Source: CZSO

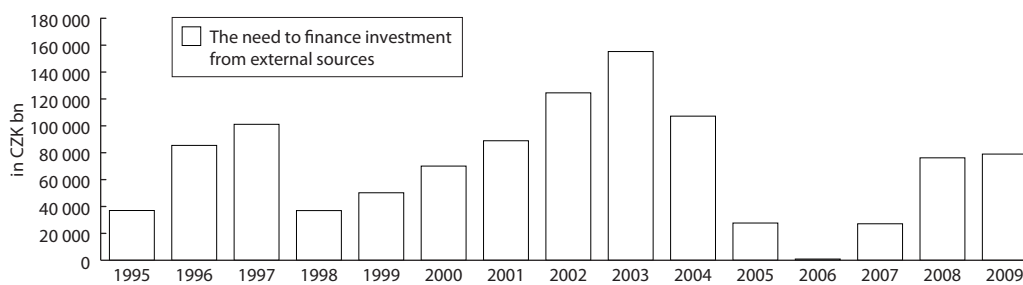
from external sources arose (Graph 15), namely already in the year 2008, when the crisis has not been fully developed, yet.

Although in the years 2008 and 2009 both the gross fixed capital formation and gross national saving were decreasing in their dynamics (the latter even faster than the investment) – the need to finance investment from external sources in the crisis year 2009 remained almost the same as in 2008.

Therefore, it seems that the Czech economy is able to reach balance in the sphere of investment and gross national saving exclusively in the periods of extreme growth – the top of the boom in the year 2006 and virtually balanced financing of investment thanks to gross national saving in the mentioned year, as well as decreasing of the need of external sources in 2004 and 2005 clear from Graph 15 are confirming that fact.

Thus, it depends on the willingness and possibilities of external investors to finance investment in the Czech Republic, which depends on the intensity of flows based on the cyclical phase of the world or rather main European economies. In general, investment depends on income, which will be generated by the state of overall economic activity[11] – besides them also by the expense-revenue ratio (i.e. interest rates and tax policy)

Graph 15 The need to finance investment from external sources (in CZK bn)



Source: CZSO

[11] Samuelson, Nordhaus: "Investment bring income to a company only provided that it can sell more." Ekonomie. 1992, page 136.

and future expectations. It can be expected that the Czech economy small in its size, into which direct foreign investment was coming especially during the first five years after the year 2000 (besides others also in sales of the state's stakes in important companies) and the main recipient was manufacturing industry, can be saturated in the future by external sources of financing rather in the sphere of services.

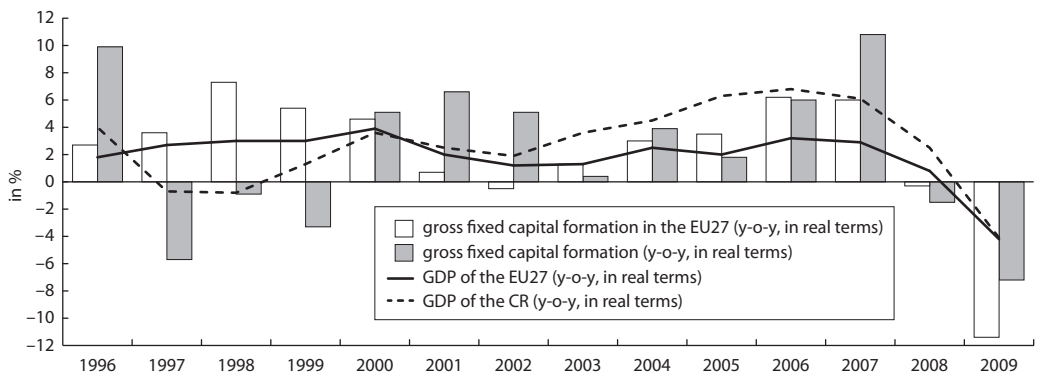
4.2 Development of investment in the time of crisis and influence on the potential GDP growth

Can drops of investment in the years 2008 and 2009 have markedly negative influence on the potential GDP growth of the Czech Republic in the following periods? Will it slow down the real convergence of the Czech economy towards the European Economic and Monetary Union? Mutual relationship in real terms between the dynamics of economic growth and the rate of gross fixed capital formation in the Czech Republic compared to the EU27 (or rather countries of the Euro-zone) during the period, which can be labelled as an economic crisis – including quarters that were preceding to it – shows differences between both entities being compared for both the start of the crisis and depths of falls caused by that crisis.

From year-on-year changes according to data seasonally and working days adjusted (as well as at data not seasonally adjusted) it results that in the CR compared to the EU27 drop of investment was clear already in the second quarter of 2008, i.e. by a quarter sooner than in the EU27 and according to the data available when this article was being elaborated, it lasted the same as in the EU27 until the fourth quarter of 2009. However, a huge difference was in the depth of the fall – during each quarter of that period investment in the EU fell deeper than their fall in the CR. While in the CR the gross fixed capital formation dropped by more than 10% only in the third quarter of 2009, in the EU27 as a whole falls deep like that were recorded in each quarter of 2009; they were pushed – besides small countries that were hit the most by the crisis – also by the development of some big economies (Italy, Germany).

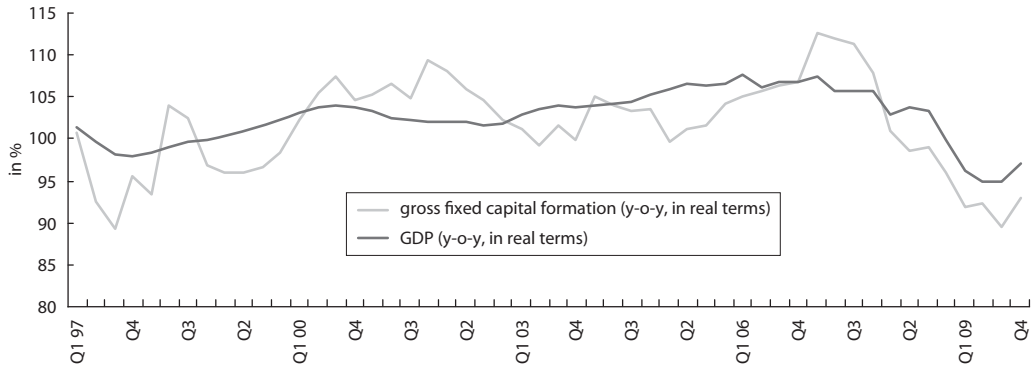
The up-to-now strongest boom of the Czech economy in 2005–2007, however, was not clearly determined in GDP growth rates by increments of investment, but it was "spread" more largely also to other components of the expenditure side of GDP, primarily to external trade (net export). For example, expenditure for final uses of households including non-profit institutions serving households (NPISHs) has in the GDP volume a higher weight in the EU27 countries as a whole than in the CR (in the period of 2000–2009 the difference was from 6.2 p.p. in 2000 and 2004 up to 8.5 p.p. in the year 2007; the boom of the Czech economy was

Graph 16 Dynamics of GDP and gross fixed capital formation in the CR and EU27
(year-on-year changes in %, in real terms)



Source: Eurostat

Graph 17 Dynamics of gross fixed capital formation and GDP by quarter
(year-on-year index, in real terms, seasonally not adjusted)



Source: CZSO

decreasing the mentioned share namely in favour of the higher share of net export).

The share of gross fixed capital formation itself in the economic performance of the country has distinctively decreased since 1996 according to calculations from average prices of the previous year – while in 1996 it exceeded a third (33.3%), in 2008 it made up almost a quarter (24.2%) and in the crisis year 2009 it further decreased to 22.6%. Also in price conditions of the year 2000 there was a clear drop, although not that significant as in the previous comparison (from 30.4% to 27.6% in the year 2008, and 26.7% in 2009). By this tendency, the CR was close to the proportions of most of the developed countries.

Despite gradual decreasing of the share of investment in GDP[12,13] the Czech economy in the context of Europe belongs to countries, where this proportion is higher than the average for the EU27 (19.7% in 2009). Nevertheless, it is interesting that the EU27 as a whole was affected by the crisis as for investment much more than the Czech Republic when comparing decrease of shares of gross fixed capital formation in GDP between 2008 and 2009 – while in the EU it dropped by 1.7 p.p. (at lower proportion than in the CR), in the Czech economy it was by 1.5 p.p.

From this point of view, the slowdown and drop of investment activity is a less endangering factor for the economic growth of the CR than in the countries, in which the investment dynamics is crucial for their economic growth.

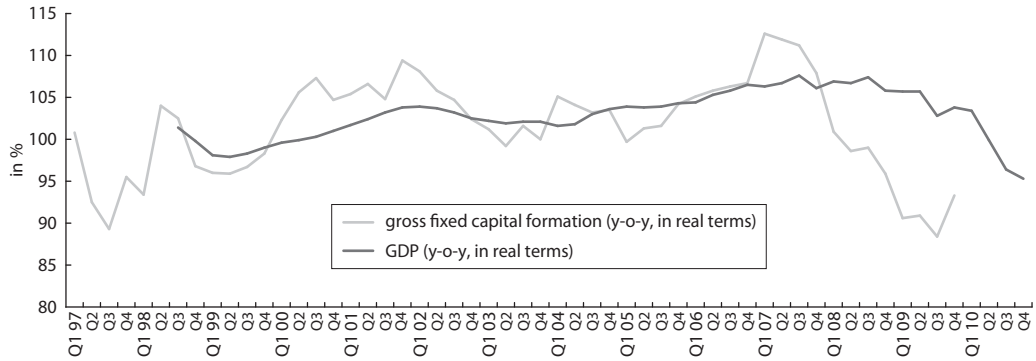
The answer to the question to what extent the GDP development of the Czech Republic is determined by the development of investment lies, among others, also in the weight, which the gross fixed capital formation has in GDP in individual years and on the multiplication effect of investment (2006, 3). The Czech economy has already seen an investment fall similar to that of the year 2009 in the second half of the 1990's.

A steep fall of investment during the year 1997, which was similar to the situation in 2009 as for its depth (Graph 17), however, was relatively soon replaced by a marked mitigation of the fall and following growth of investment (during mere four quarters by 18.9 p.p.). Together with that, also the real GDP got out of the decrease. However, at that time, there was an economic recession due to strong restrictions reacting on the monetary crisis, which had its origin in the Czech economy. The situation from the year 2009 was a consequence of a crisis "implanted" from

[12] According to data of Eurostat, shares calculated from national currencies in constant prices of the year 2000.

[13] See footnote 1.

Graph 18 Dynamics of gross fixed capital formation and GDP in the shift by six quarters
(year-on-year changes in %, in real terms, seasonally not adjusted)



Source: CZSO

the outside via a steep drop of external demand, which decimated the Czech economy that is oriented on exports.

Time horizon of investment actions (clear mainly at construction investment) causes that the effect of investment does not have to be captured in statistics directly in an increment of the product in the given year but only (which is usual) in the future periods. An example, when investment was falling and GDP was further growing, monitored according to curves with GDP shift by six quarters (Graph 18) can be seen basically only in the beginning of 2004, 2005 and then in a massive way also in the beginning of the year 2008. Non-shifted curves according to data in real terms then show the contradictory development in the end of the year 1998, beginning of 2003, end of 2004, and in the first quarter of the year 2008.

The question from the introduction of this sub-chapter regarding a negative influence of the development of investment in 2009 on expected GDP can be answered from the following points of view:

- In the Czech economy the share of investment in GDP is higher than in the EU27, but during the entire decade of 2000–2009 it markedly decreased – the difference was two times higher in percentage points than in the EU27 (while in the EU the share of gross fixed capital formation in GDP decreased in 2009 compared to the year 2000 by 0.9 p.p., in the CR it was by 1.8 p.p.). At the same time, however, the CR belonged to few European countries, in which this share decreased also in "non-crisis" years 2000–2008 (similarly as in Germany or Ireland). Nevertheless, here it has to be born in mind that the decrease of investment share was markedly influenced by an export boom that was started by the accession of the CR to the EU, which increased the share of net export in the economic performance of the country and thus suppressed the influence of investment. Should this factor survive, the 2009 investment fall should not have a significant influence on the future growth rates of the Czech economy in that sense that the real convergence of the Czech Republic to the level of the EU would be markedly decreased.

- Correlation between gross fixed capital formation and GDP was lower than in the EU27 during the period of 1995–2009, which is again a factor in favour of less important influence on the pace of convergence as well as

- year-on-year falls of investment during the 2009 economic crisis – and first of all at observing year-on-year changes by individual quarters of that year – that were again much weaker compared to the EU27.

From the mentioned development during the crisis it can be deduced that the dynamics of gross fixed capital formation in the CR in the future does not justify an assumption that the convergence of the Czech Republic to an average level in the EU27 should be markedly slowed down due to the development of investment.

CONCLUSION

State of gross fixed capital in the Czech economy in the end of 2007 expressed in prices of the year 2000 increased in comparison to the year 1997 by more than a fifth, compared to 1995 even by more than a quarter. It reached CZK 20.6 trillion compared to 17 trillion in 1997. States of gross fixed capital in the CR consist in an overwhelming majority of tangible fixed assets the share of which moves in the long-term on the level of 99% of the gross fixed capital of the Czech economy.

From the point of view of industries, what is dominating to the states of gross fixed capital in the economy of the CR is the national economy industry, which includes besides services to businesses and research and development also activity in the area of real estate, renting and business activities (CZ-NACE K). In the end of the year 2007, it contributed to the states of gross fixed capital in the economy of the CR by almost 28% (CZK 5.7 trillion). In the mentioned year, real estate activities (CZ-NACE 70) had the share of almost three quarters of tangible assets of dwellings type in its gross fixed capital formation. The correlation between investment and gross value added formation in that industry (into which also stock of dwellings belongs) is higher than the closeness of dependency of both quantities for the economy as a whole.

For the years 1997 to 2007, states of total fixed assets in the CR expressed in prices of the year 2000 grew by 2%, y-o-y, on average. Investment in dwellings (multi-dwelling houses, family houses and flats), however, reported a real year-on-year decrease by 0.2%; year-on-year decrease of states occurred in 1997–1999 and also in 2002–2004. As a result, state of dwellings type fixed assets expressed in prices of the year 2000 in the end of 2007 remained almost unchanged in comparison to the end of 1997 (CZK 4.9 trillion against 5 trillion). States of other buildings and structures – into which also infrastructure and non-residential houses belong – on the contrary, increased during that period by CZK 2.4 trillion to CZK 11.4 trillion. States of gross fixed capital in the form of machinery and equipment including transport equipment also increased (by CZK 2.1 trillion to 4.3 trillion). As it is obvious from the ratios, the gross fixed capital of dwellings type and other structures type exceeds significantly the gross fixed capital in the form of machinery and transport equipment. However, compared to the states of the gross fixed capital of other structures type and by its state it is less than a half.

Rather surprising actual stagnation of states of tangible fixed assets of dwellings type in the CR between 1997 and 2007 can be explained partly factually: tangible fixed assets in the households sector grew also due to privatisation; on the contrary, in the sector of non-financial corporations and in government sector their states were decreasing. Another explanation can be made by the methodology of reporting: life of the assets of dwellings type is set for 80 years and every year assets over this age limit are eliminated from the national accounts statistics. When investment (gross fixed capital formation) in dwellings in a given year exceed the value of those eliminated assets, their total states are increasing, in the opposite case they are decreasing. Actual stagnation in the 1997 to 2007 decade thus means that the value of multi-dwelling houses, family houses and flats eliminated from the statistics was replaced by gross fixed capital formation of dwellings type roughly in the same scope during that period.

Thanks to these changes, the share of fixed assets of dwellings type (that actually stagnated in real terms in the mentioned period or decreased by 0.2%, respectively) in total volume of fixed assets in the CR (that, on the contrary, increased by more than a fifth) decreased in real terms according to the states from the end of 2007 compared to the end of 1997. It was a logical decrease from 31% to 24%. However, in nominal terms (in current prices) investment to houses and flats in the period of 1996–2007 increased by 5.6%, y-o-y,

on average, with acceleration during the years of the boom of the Czech economy in the end of the observed period (from the second quarter of 2005 to the second quarter of 2008 the year-on-year growth was +6.7% on average).

Households have tripled their investment in dwellings – according to fixed assets of dwellings type in that sector – for the years 1995–2006 in nominal terms (from CZK 37 billion to 102 billion according to net acquisition of this type of tangible fixed assets). The households sector thus participated in the mentioned period with about two thirds in the total value of newly acquired houses and flats in the economy and its share in time was increasing (from 59% in 1996 to 74% in 2006 as a reflection of the beginning of the boom on the market with flats). Strengthening of the volume of tangible assets of households in the form of houses and flats – as one of the types of non-financial assets, which share in their total volume in the households sector by more than three quarters[14] – was reflected also in nominal and real profits of households from holding of non-financial assets. Thus, while nominal profits of households increased (from CZK 78 billion in 2005 to 133 billion in 2007), their real profits in 2007 decreased, year-on-year (from CZK 58 bn to 46 bn). For the Czech economy in total, non-financial assets strengthened the total net worth in the national economy in a dynamic trend – during the years 2005 to 2007 the influence of profits from these assets strengthening the wealth (expressed by growth of the net worth in the economy) was higher by 85% (between years 2006 and 2005 it was by 45%). In the year 2005 profits from holding of non-financial assets increased the net worth in the economy of the CR by CZK 315 bn; in 2007 it was by CZK 581.3 bn.

During major part of the period of the boom of the Czech economy in 2004–2006 gross national saving grew faster than investment. It caused that in 2006 it even was not necessary to finance an increase of gross fixed capital formation in the Czech economy from other than national sources, as it was usual during the entire period from 1995. In 2008 and 2009, however, when gross national saving dropped faster than investment in 2008, the need of financing from external sources was again confirmed.

The crisis year of 2009 brought together with the drop of gross fixed capital formation also a question to what extent this circumstance will "harm" the future economic growth. Despite the up-to-now higher share of investment in economic performance of the country than the EU27 average, but at its gradual decreasing, at lower correlation of gross fixed capital formation and GDP in the Czech economy compared to the EU27 and also lower year-on-year falls of investment during the 2009 economic crisis compared to the EU – all that leads us to an assumption that those decreases in a medium-term horizon should not lead to significant losses in the GDP growth rate of the Czech Republic. Their decelerating influence – which is evident with regard to drop rates – however, together with slump of external demand that determined in 2009 in a decisive way the fall of the Czech economy to recession, became only one of the elements causing the length of duration of the recession. In relation to this, however, in the future development of especially the enterprise sector, it will be important that in the strong investment wave from the boom period enterprises probably set their forecasts of investment returns according to their income from the boom period, which might have been markedly corrected by the following development in the second half of 2008 and mainly the 2009 crisis environment. It might be a risk also for financial institutions giving credits for those investment projects.

[14] Since the Households sector consists of the segment of Households-individuals and Households-enterprises, an important part of the volumes of non-financial assets (fixed assets and inventories) belongs to the segment of Households-enterprises.

References

1. *Průmyslová odvětví zblízka 2000–2005 (definitivní údaje).* Prague: Czech Statistical Office, 2007. <http://czso.cz/csu/2006edicniplan.nsf/p/1146-06>.
2. DUBSKÁ, D. *Fixní aktiva v české ekonomice.* Prague: Czech Statistical Office, 2009. <http://czso.cz/csu/2009edicniplan.nsf/p/1157-09>.
3. DUBSKÁ, D. *Investice a ekonomický růst v České republice: kam se ztrácí vysoká míra investic?* Prague: Czech Statistical Office, 2006. <http://czso.cz/csu/2006edicniplan.nsf/p/1135-06>.
4. KUBÍČEK J. *Rovnovážná cena fixního aktiva v rostoucí ekonomice.* Working Papers No. 9/2003. Prague: The University of Economics, Prague, 2003.
5. DUBSKÁ, D. *Srovnání výkonnosti cizích zdrojů ve formě bankovních úvěrů v produkčních odvětvích české ekonomiky.* Prague: Czech Statistical Office, 2009. <http://czso.cz/csu/csu.nsf/informace/ckta061709.doc>.
6. VLKOVÁ, J. *Zdroje podnikání a účetní přidaná hodnota nefinančních podniků v období let 2000 až 2004.* Prague: Czech Statistical Office, 2005. <http://czso.cz/csu/2005edicniplan.nsf/p/1532-05>.
7. DUBSKÁ, D. *Role domácího kapitálu v transformaci české ekonomiky.* Prague: Czech Statistical Office, 2005. <http://czso.cz/csu/2005edicniplan.nsf/p/1530-05>.
8. SAMUELSON, P., NORDHAUS, W.D. *Ekonomie.* Victoria Publishing, 1992.
9. DUBSKÁ, D. I*nvestiční aktivita v odvětvích české ekonomiky během let 1995–2005.* Prague: Czech Statistical Office, 2009. <http://czso.cz/csu/csu.nsf/informace/ckta211106.doc>.
10. CZESANÝ, S., DUBSKÁ. D., JEŘÁBKOVÁ, Z. *Komponenty bohatství zemí.* Prague: Czech Statistical Office, 2009. <http://czso.cz/csu/2008edicniplan.nsf/p/1150-08>.

Analysis of the Development in Wage Distributions of Men and Women in the Czech Republic in Recent Years

Diana Bílková[a] | *University of Economics, Prague*

ABSTRACT

The paper presents the development of the distributions of gross monthly wages of men and women in the Czech Republic in the recent years and foreshadows future development of these distributions in next years providing the elimination of the effect of economic recession. For the characterization of the development of wage distributions of the recent years we used the following characteristics of location of frequency distribution – arithmetic mean, median and medial; within the frame of the characterization of the development of variability we used the moment characteristic of variability – standard deviation and in light of tracing of the development of shape of the wage distributions we used the moment characteristic – coefficient of skewness. Within the framework of modelling of these wage distributions we used the free-parametric lognormal curve, whereas the parameters of this curve were estimated using the moment method and the quantile method of the point estimation. Using the trend analysis we predicted the development of descriptive characteristics of wage distributions for the next two years, from which we obtained the lognormal model of wage distributions for the next two years, whereupon we estimated the shares of employees in bounds of the gross monthly wages namely for the total set of men and women together and separately for the set of men and for the set of women. These predictions present the development of wage distributions of men and women in the Czech Republic in the years 2009 and 2010 providing abstraction of the effects of economic recession, which broke out in the autumn of the year 2008. Comparing the predictions with the really observed wage distributions in the two years we can quantify the effect of economic recession on the development of the wage distributions of men and women in the Czech Republic.

Keywords

wage distributions, lognormal curve, probability density function, moment method of parameter estimation, quantile method of parameter estimation, prediction of wage distributions, shares of employees in the bounds of gross monthly wages

INTRODUCTION

Estimations of wage distributions development allow, among other things, to link the findings relating to wage differentiation with socio-political considerations for which the estimation of average wage development is often not sufficient and which also require estimations of the shares of workers with low, medium and high wages, or shares of workers in all wage groups. The estimation of wage distributions based on a certain idea

[a] *University of Economics, Prague, e-mail: bilkova@use.cz*

of wage differentiation also allows us to estimate, for example, the total volume of wages to be paid in the future, etc., see [22].

This study is based on the data presented by Czech Statistical Office – "Percentages of employees in bands of gross wage by gender" for the years 2002 to 2008 in the Czech Republic and the data on sampled worker numbers by gender. The investigated variable was the worker's gross monthly wage in CZK. The input data were provided in the form of tables with interval frequency distribution with open extreme intervals. Data relevant to a longer period of time could not be used due to their incomparability. The data were processed using Microsoft Excel and the SAS and Statgraphics statistic programs.

1 SETS OF DESCRIPTIVE CHARACTERISTICS

A description of a unimodal frequency distribution usually concerns basic characteristics, such as the position (level), variability, skew, or sharpness, see [15]. Sets of moments and moment characteristics or sets of quantiles and quantile characteristics, see [6], have been traditionally used for this purpose.

We will mark the centers of individual wage intervals as x_i, $i = 1, 2, \ldots, k$, and the absolute numbers in individual intervals as n_i, $i = 1, 2, \ldots, k$, where k is the total number of intervals. From the moment characteristics of the position, we can apply the arithmetic mean

$$\bar{x} = \frac{\sum_{i=1}^{k} x_i \cdot n_i}{\sum_{i=1}^{k} n_i}. \tag{1}$$

standard deviation

$$s_x = \sqrt{\frac{\sum_{i=1}^{k} (x_i - \bar{x})^2 \cdot n_i}{\sum_{i=1}^{k} n_i}}, \tag{2}$$

see [5]; or variation coefficient

$$v_x = \frac{s_x}{\bar{x}} \cdot 100 \tag{3}$$

to characterize the variability (absolute or relative). The third moment of the standard variable is used to characterize the skew of the observed distribution of frequencies.

$$\sqrt{b_1(x)} = \frac{\sum_{i=1}^{k} (x_i - \bar{x})^3 \cdot n_i}{s_x^3 \cdot \sum_{i=1}^{k} n_i}. \tag{4}$$

For right-side asymmetry it holds that $\sqrt{b_1(x)} > 0$, for left-side asymmetry it holds that $\sqrt{b_1(x)} < 0$, and for symmetric distribution of frequencies it holds that $\sqrt{b_1(x)} = 0$. Parameter $\left|\sqrt{b_1(x)}\right|$ grows with the growing asymmetry of the observed distribution of frequencies. Moment characteristics have an important advantage for the description of the distribution of frequencies – they react sensitively to a change of any characteristic feature of the distribution as a whole, see [17]. Except for the arithmetical mean, the disadvantage of the moment characteristics for description and comparison, is their difficult logical and in our case also economic interpretability, see [18].

The P% quantile of variable x is understood as value $\tilde{x}_P$, $0 < P < 100$, where P % of the observed values of variable x is less or equal to the value of P% quantile $\tilde{x}_P$ and the remaining (100 – P) % of the observed values of the x variable is larger or equal to the value of P% quantile $\tilde{x}_P$, see [12]. The most important quantile is the median (middle value), which divides the non-decreasing sequence of the observed values of variable x in two halves with equal count and which is marked as $\tilde{x}_{50}$, or only $\tilde{x}$. Quartiles represent three values (lower quartile = 25% quantile $\tilde{x}_{25}$, upper quartile = 75% quantile $\tilde{x}_{75}$ and the median), which divide the non-decreasing sequence of observed values of variable x in four parts with equal count. Apart from quartiles, there are also terciles (two values dividing the non-decreasing sequence of values of the observed variable x into three parts with equal counts), quintiles, sextiles, septiles, octiles, noniles. Deciles represent nine values dividing the non-decreasing sequence of the values of variable x in ten parts with equal count; the first decile is the 10% quantile $\tilde{x}_{10}$, the second decile is the 20% quantile $\tilde{x}_{20}$, etc., up to the ninth decile, which

is the 90% quantile $\tilde{x}_{90}$. And finally, percentiles are 99 values, which divide the non-decreasing sequence of values of the observed variable x in 100 parts with equal count. There is the first percentile, which is the 1% quantile $\tilde{x}_1$, the second percentile, which is the 2% quantile $\tilde{x}_2$, etc., up to the 99th percentile representing the 99% quantile $\tilde{x}_{99}$. Quantiles lower than the median are called lower quantiles, while quantiles higher than the median are called the upper quantiles. Quantile characteristics are determined according to the quantiles, see [15].

In general, we can say that if we want to describe the characteristics of the entire frequency distribution, it is more convenient to construe characteristics that are based on quantiles more distant from the median, i.e. percentiles are more suitable than deciles and deciles are more suitable than quartiles. However, when the calculations are based on interval frequency distribution with open extreme intervals (our case), the extreme percentiles are often situated in these open intervals, and therefore have to be estimated using linear interpolation. As these estimations are not too accurate, we prefer, in such a case, characteristics based, for example, on deciles to those based on percentiles. Quantiles are conventionally used to determine the interval of medium wages. The first decile is used to define lower wages, while the ninth decile is used to determine high wages. The first percentile is used as a limit for minimum wages and the last percentile as a limit of maximum wages. The median is used to characterize the medium level of wages, see [7] and [8].

P% tantile is a value of $\check{x}_P$, $0 < P < 100$, where the sum of values lower or equal to the value of the P% tantile $\check{x}_P$ represents P % of the total sum of values of variable x and the sum of values higher or equal to P% tantile $\check{x}_P$ represents the remaining $(100 - P)$% of the total sum of values of variable x. An interesting characteristic of position is the medial, which represents the 50% tantile (sum centre) $\check{x}$. Employees whose wage is no more than equal to the medial value receive one half of the total volume of wages and employees whose wages are at least equal to the medial value receive the other half of the total volume of wages.

2 LOGNORMAL CURVE

The most important of the models used to model wage and income distributions is the lognormal distribution, see [1] and [2]; with various modifications in the form of two-parameter, three-parameter and four-parameter lognormal curve, see [14]. The importance of the lognormal curve for the modeling of empiric distributions is doubtless, see [3]. Typical features of the processes modeled using this curve are: gradual actuation of interdependent factors, tendency towards development in a geometric progression and transition of random variability to systematic variability, see [16]. Wages and incomes belong to the group of economic phenomena that can be interpreted using the lognormal model, see [23]; three-parameter lognormal distribution is most commonly used in these models, see [3].

The probability density function for random variable X with three-parametric lognormal distribution $LN(\mu; \sigma^2; \xi)$ with parameters μ, σ^2 a ξ, $-\infty < \mu < \infty$, $\sigma^2 > 0$, $-\infty < \xi < \infty$, is

$$f(x) = \frac{1}{\sigma (x-\xi)\sqrt{2\pi}} \cdot \exp\left\{-\frac{[\ln(x-\xi)-\mu]^2}{2\sigma^2}\right\}, \quad x > \xi, \qquad (5)$$
$$= 0, \quad x \le \xi,$$

see [19]. If the random variable X has three-parametric lognormal distribution $LN(\mu; \sigma^2; \xi)$, then the random variable

$$Y = \ln(X - \xi) \qquad (6)$$

has normal distribution $N(\mu; \sigma^2)$ and the random variable

$$U = \frac{\ln(X - \xi) - \mu}{\sigma} \tag{7}$$

has a standardized normal distribution N(0; 1), see [9] and [10].

The basic moment characteristics of the location of the three-parametric lognormal distribution is the expected value of random magnitude *X*, having the form

$$E(X) = \xi + \exp\left(\mu + \frac{\sigma^2}{2}\right). \tag{8}$$

The quantile characteristics of the location of the three-parametric lognormal distribution is the 100 *P*% quantile x_P, which is determined as the value for which the value of distribution function *F*(x) of random variable *X* in point x_P is equal to *P*

$$F(x_P) = P, \quad 0 < P < 1, \tag{9}$$

i.e., for the three-parametric lognormal distribution the 100 *P*% quantile x_P has the form

$$x_P = \xi + \exp(\mu + \sigma u_p), \; 0 < P < 1. \tag{10}$$

If we use *P* = 0,5 in relation (10), we obtain the 50% quantile of the three-parametric lognormal distribution, also called the median, which is the basic quantile characteristics of location of this distribution

$$x_{0,50} = \xi + \exp(\mu). \tag{11}$$

The basic moment characteristics of the variability of the three-parametric lognormal distribution is the variance

$$D(X) = \exp(2\mu + \sigma^2) \cdot [\exp(\sigma^2) - 1], \tag{12}$$

the square root of the variance, i.e. the standard deviation

$$\sigma(X) = \sqrt{D(X)} = \exp\left(\mu + \frac{\sigma^2}{2}\right) \cdot \sqrt{\exp(\sigma^2) - 1}, \tag{13}$$

or the characteristics of the relative variability – variation coefficient

$$V(X) = \frac{\sigma(X)}{E(X)} = \frac{\exp\left(\mu + \frac{\sigma^2}{2}\right) \cdot \sqrt{\exp(\sigma^2) - 1}}{\xi + \exp\left(\mu + \frac{\sigma^2}{2}\right)}, \tag{14}$$

see [4]. The moment characteristics of the distribution shape include the skewness coefficient of the three-parametric lognormal distribution

$$\sqrt{\beta_1(X)} = [\exp(\sigma^2) + 2] \cdot \sqrt{\exp(\sigma^2) - 1} \tag{15}$$

and the kurtosis coefficient of this distribution

$$\beta_2(X) = \exp(4\sigma^2) + 2\exp(3\sigma^2) + 3\exp(2\sigma^2) - 3. \tag{16}$$

2.1 Moment method for parameter estimation

Moment estimations of parameters μ, σ^2 and ξ of the three-parametric lognormal distribution are obtained by putting into equation three sample moments and the corresponding moments of this theoretic distribution. As we are estimating three parameters, we need three moment equations. Therefore we put into equation the arithmetic mean $\bar{x}$ obtained from the sample and the expected value of the three-parametric lognormal distribution (8), the second sample central moment m_2 is put in equation with the variance of three-parametric lognormal distribution (12) (given the large extent of the sets, it is not necessary to distinguish between the sample variance and the sample second central moment, as with such large extents of sample, their values are practically identical) and the sample third central moment m_3 is put into equation with the third central moment of the three-parametric lognormal distribution

$$\bar{x} = \tilde{\xi} + \exp\left(\tilde{\mu} + \frac{\tilde{\sigma}^2}{2}\right), \tag{17.1}$$

$$m_2 = \exp(2\tilde{\mu} + \tilde{\sigma}^2)\cdot[\exp(\tilde{\sigma}^2) - 1], \tag{17.2}$$

$$m_3 = \exp\left(3\tilde{\mu} + \frac{3}{2}\tilde{\sigma}^2\right)\cdot[\exp(\tilde{\sigma}^2) - 1]^2\cdot[\exp(\tilde{\sigma}^2) + 2], \tag{17.3}$$

see [20].

By solving the system of moment equations (17) we obtain the moment estimates of the parameters of the three-parametric lognormal distribution

$$\tilde{\sigma}^2 = \ln\left[\sqrt[3]{1 + \frac{1}{2}b_1 + \sqrt{\left(1 + \frac{1}{2}b_1\right)^2 - 1}} + \sqrt[3]{1 + \frac{1}{2}b_1 - \sqrt{\left(1 + \frac{1}{2}b_1\right)^2 - 1}} - 1\right], \tag{18.1}$$

$$\tilde{\mu} = \frac{1}{2}\ln\frac{m_2}{\exp(\tilde{\sigma}^2)\cdot[\exp(\tilde{\sigma}^2) - 1]}, \tag{18.2}$$

$$\tilde{\xi} = \bar{x} - \exp\left(\tilde{\mu} + \frac{\tilde{\sigma}^2}{2}\right). \tag{18.3}$$

The moment method of parameter estimation does not guarantee maximum efficiency of the estimate, however, in the case of wage distribution we handle very large samples, and therefore any consistent method of parameter estimation yields satisfactory results and the moment method of parameter estimation can be used.

2.2 Quantile method of parameter estimation

The quantile method of estimation of the three-parametric lognormal distribution parameters uses three sample quantiles, namely the $100P_1\%$ quantile x^V_{P1}, 50% quantile (median) $x^V_{0,50}$ and the $100(1 - P_1)\%$ quantile $x^V_{(1-P1)}$, $0 < P_1 < 1$. These three sample quantiles are equated with the corresponding quantiles of the three-parametric lognormal distribution (10), by which we create a system of three quantile equations

$$x^V_{P1} = \xi^* + \exp(\mu^* + \sigma^* u_{P1}), \tag{19.1}$$

$$x^V_{0,50} = \xi^* + \exp(\mu^*), \tag{19.2}$$

$$x^V_{(1-P1)} = \xi^* + \exp(\mu^* - \sigma^* u_{P1}). \tag{19.3}$$

From the system of quantile equations (19) we obtain quantile estimates of the parameters of the three-parametric lognormal distribution

$$\sigma^{2^*} = \left[\frac{\ln \dfrac{x^V_{P1} - x^V_{0,50}}{x^V_{0,50} - x^V_{(1-P1)}}}{u_{P1}} \right]^2, \tag{20.1}$$

$$\mu^* = \ln \frac{x^V_{P1} - x^V_{(1-P1)}}{\exp(\sigma^* u_{P1}) - \exp(-\sigma^* u_{P1})}, \tag{20.2}$$

$$\xi^* = x^V_{0,50} - \exp(\mu^*), \tag{20.3}$$

see [20].

3 DEVELOPMENT OF WAGE DISTRIBUTIONS FOR MEN AND WOMEN IN THE CZECH REPUBLIC IN 2002–2008 AND PREDICTION FOR 2009 AND 2010

Table 1 shows the development of the medial value of gross monthly wages in the Czech Republic in 2002–2008 total for men and women together and also separately for men and women, including the prediction of the development of the medial value of gross monthly wages for 2009 and 2010. Tables 2–4 show the development of extreme deciles and extreme percentiles of gross monthly wages in the Czech Republic in 2002–2008. This development is also shown in Graphs 1–3. Table 2 shows that in the total group of all workers in 2002, 80% medium gross monthly wages were in the interval between CZK 9 243 and CZK 27 754, i.e., 10% of all wages were lower than or equal to CZK 9 243 and 10% of all wages were equal to or higher than CZK 27 754. As it is also shown in Graph 1, the interval between the first and ninth decile is slightly widening until 2008, when it can be said that 80% medium wages in the Czech Republic were in the interval from CZK 12 761 to CZK 40 548, i.e. 10% of gross monthly wages in the Czech Republic were not higher than CZK 12 761 and 10% of gross monthly wages were not lower than CZK 40 548. As for the development of extreme percentiles, Table 2 shows that 98% medium wages in 2002 were in the interval from CZK 6 365 to CZK 47 172, while the interval the first and the last percentile widens considerably

Table 1 Development of the sample medial value of gross monthly wages in the Czech Republic in 2002–2008 and prediction of the development of medial value of gross monthly wages for 2009 and 2010 (in CZK)

Group	Year								
	2002	2003	2004	2005	2006	2007	2008	2009	2010
Total	18 389	19 835	21 139	22 317	23 358	25 530	27 183	**28 815**	**30 703**
Men	20 014	21 799	22 964	24 295	25 440	27 770	29 802	**31 448**	**33 513**
Women	16 148	17 548	18 554	19 562	20 555	22 130	23 027	**24 189**	**25 325**

Source: own research

Table 2 Development of sample extreme deciles and extreme percentiles of gross monthly wages in the Czech Republic in 2002–2008 total for men and women (in CZK)

Deciles and percentiles	Year						
	2002	2003	2004	2005	2006	2007	2008
$\tilde{x}_{0,01}$	6 365	6 680	3 090	4 135	8 070	2 587	3 324
$\tilde{x}_{0,10}$	9 243	9 814	10 235	10 634	11 248	12 127	12 761
$\tilde{x}_{0,90}$	27 754	29 590	31 082	33 292	35 229	37 904	40 548
$\tilde{x}_{0,99}$	47 172	47 719	56 369	56 852	57 326	86 461	88 866

Source: own research

over the time until 2008, when 98% medium wages were situated in the interval from CZK 3 324 to CZK 88 866. These values are usually considered to be the limits of minimum and maximum wages. However, it must be reminded here that the calculations are based on data organized in a table of interval frequency distribution with open extreme intervals. Therefore, the estimated values of extreme percentiles have to be handled with caution and should be understood only as orientation values. The same applies to the estimated values of extreme percentiles in Tables 3 and 4.

Table 3 shows the development of extreme percentiles of gross monthly wages of men and Table 4 describes the development of extreme deciles and extreme percentiles of gross monthly wages of women, both for the Czech Republic in 2002–2008. Table 3 demonstrates that in 2002 80% gross monthly wages of men were in the interval from CZK 10 778 to CZK 31 101, Graph 2 shows that the interval between the extreme deciles of gross monthly wages of men was then widening gradually until 2008, when 80% gross monthly wages of men lied in the interval from CZK 14 822 to CZK 46 364, in 2002 10% gross monthly wages of men did not exceed the amount of CZK 10 778 and 10% gross monthly wages did not fall below CZK 31 101, while in 2008 10%

Table 3 Development of sample extreme deciles and extreme percentiles of gross monthly wages in the Czech Republic in 2002–2008 for men (in CZK)

Deciles and percentiles	Year						
	2002	2003	2004	2005	2006	2007	2008
$\tilde{x}_{0,01}$	7 066	7 497	6 427	8 054	8 369	4 430	5 367
$\tilde{x}_{0,10}$	10 778	11 459	11 945	12 348	12 930	13 962	14 822
$\tilde{x}_{0,90}$	31 101	34 564	34 819	37 211	39 381	42 870	46 364
$\tilde{x}_{0,99}$	48 047	48 417	57 514	57 808	58 104	90 671	92 333

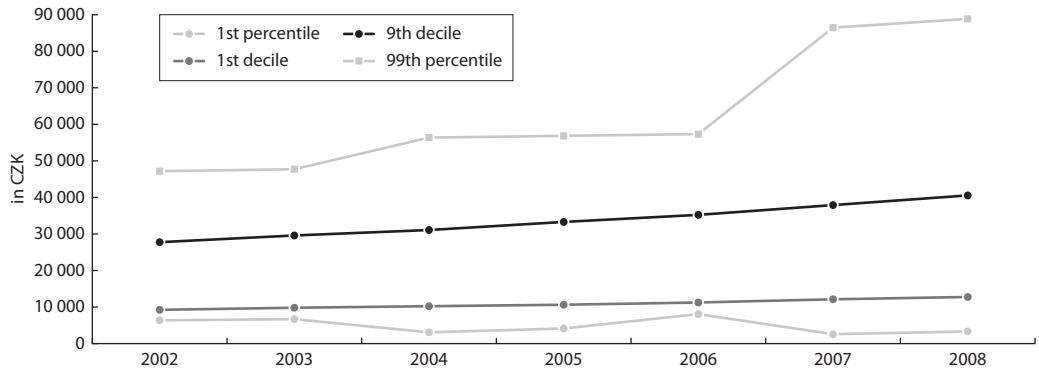
Source: own research

Table 4 Development of sample extreme deciles and extreme percentiles of gross monthly wages in the Czech Republic in 2002–2008 for women (in CZK)

Deciles and percentiles	Year						
	2002	2003	2004	2005	2006	2007	2008
$\tilde{x}_{0,01}$	6 098	6 277	1 835	2 442	6 458	1 652	2 199
$\tilde{x}_{0,10}$	8 296	8 819	9 056	9 445	10 178	10 855	11 403
$\tilde{x}_{0,90}$	23 292	24 637	25 776	27 503	29 082	31 201	33 398
$\tilde{x}_{0,99}$	43 339	44 883	50 776	52 509	54 054	68 415	73 470

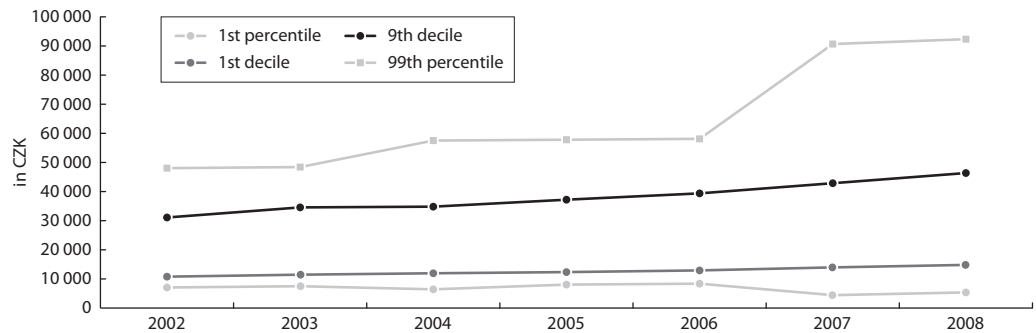
Source: own research

Graph 1 Development of sample extreme deciles and percentiles of gross monthly wages (in CZK) **in the Czech Republic in 2002–2008 overall for men and women**



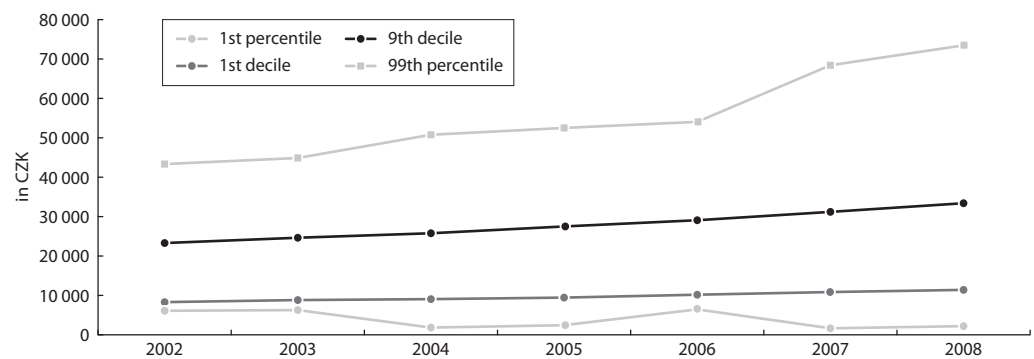
Source: own research

Graph 2 Development of sample extreme deciles and percentiles of gross monthly wages (in CZK) **in the Czech Republic in 2002–2008 for the set of men**



Source: own research

Graph 3 Development of sample extreme deciles and percentiles of gross monthly wages (in CZK) **in the Czech Republic in 2002–2008 for the set of women**



Source: own research

gross monthly wages of men did not exceed CZK 14 822 and 10 % gross monthly wages of men did not fall below CZK 46 364. In 2002, 98% men in the Czech Republic received gross monthly wages in the interval from CZK 7 066 to CZK 48 047, while in 2008, 98% men in the Czech Republic had their gross monthly wages in the interval from CZK 5 367 to CZK 92 333. However, these limits should be only used for orientation, as explained above.

Table 4 shows that in 2002, 80% gross monthly wages of women were in the interval from CZK 8 296 to CZK 23 292, i.e., 10% women did not have gross monthly salary higher than CZK 8 296 in that year, and the gross monthly wages of 10% women did not fall below CZK 23 292 in that year. Graph 3 illustrates that the interval between the extreme deciles of gross monthly wages of women was gradually slightly widening until 2008, when 80% women in the Czech Republic were paid gross monthly wage in the interval from CZK 11 403 to CZK 33 398, which means that the gross monthly wage of 10% women in the Czech Republic in 2008 did not exceed the amount of CZK 11 403 and 10 % women in 2008 had gross monthly wage at least in the amount of CZK 33 398. With a great caution, we can say that in 2002 98% women in the Czech Republic were paid gross monthly wage in the interval from CZK 6 098 to CZK 43 339, while in 2008 98% women in the Czech Republic had their gross monthly wage in the interval from CZK 2 199 to CZK 73 470.

Tables 5–7 illustrate the development of the basic descriptive moment characteristics of location (arithmetic mean), variability (standard deviation) and shape (skewness coefficient) of the frequency distribution of gross monthly wages in the Czech Republic in 2002–2008 including the predictions for 2009 and 2010, see [11]; which are based on the assumption that the current development continues. It is obvious that the location, absolute variability and skew of the distribution of gross monthly wages in the observed period have a growing trend in all three analyzed groups. Graphs 4 and 5 illustrate the development of the characteristics of the location of the medial and median values of gross monthly wages in 2002–2008, again including the predictions for 2009 and 2010, provided that the current development is sustained. It can be seen that the characteristics of the location of distributions of gross monthly wages continue to grow, both in the overall group of men and women, as well as in the separated groups of men and women.

Table 5 Development of arithmetic mean $\bar{x}$ (in CZK) standard deviation s_x (in CZK) and skewness coefficient $\sqrt{b_1(x)}$ of gross monthly wages in the Czech Republic in 2002–2008 total for men and women including the predictions for 2009 and 2010, and the parameter values of the three-parametric lognormal distribution estimated using the moment method of parameter estimation, including the predictions for 2009 and 2010

Year	Sample characteristics			Parameters estimated using the moment method		
	$\bar{x}$	s_x	$\sqrt{b_1(x)}$	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\varsigma}$
2002	17 437	8 321	1,817	9,491 969	0,264 377	2 311,688 3
2003	18 663	8 657	1,354	9,837 351	0,166 428	– 1 681,293 4
2004	19 698	9 804	1,614	9,779 148	0,220 813	–25,694 7
2005	20 738	10 180	1,481	9,905 938	0,192 775	–1 339,601 2
2006	21 803	10 477	1,419	9,979 491	0,179 767	–1 805,527 0
2007	23 883	13 776	2,338	9,733 688	0,376 546	3 509,923 8
2008	25 478	14 485	2,191	9,851 185	0,345 251	2 920,380 8
2009	**27 355**	**16 796**	**2,875**	**9,717 849**	**0,487 340**	**6 160,024 6**
2010	**29 428**	**19 182**	**3,516**	**9,645 006**	**0,609 149**	**8 484,567 7**

Source: own research

Tables 5–7 also include parameter estimations of the three-parametric lognormal distribution of gross monthly wages in the Czech Republic in 2002–2008 obtained using the moment method of parameter estimation, Table 11 contains the values of the known test criterion χ^2 and the values of the sum of all absolute deviations S of empirical and theoretical frequencies, provided that the distribution of gross monthly wages is the three-parametric lognormal distribution with the stated parameter values estimated using the moment method. With such large extent of samples, values of the test criterion χ^2 practically always lead to a refusal of the tested hypothesis on the

Table 6 Development of arithmetic mean $\bar{x}$ (in CZK) standard deviation s_x (in CZK) and skew coefficient $\sqrt{b_1(x)}$ of gross monthly wages in the Czech Republic in 2002–2008 for men, including the predictions for 2009 and 2010, and the parameter values of the three-parametric lognormal distribution estimated using the moment method of parameter estimation, including the predictions for 2009 and 2010

Year	Sample characteristics			Parameters estimated using the moment method		
	$\bar{x}$	s_x	$\sqrt{b_1(x)}$	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\varsigma}$
2002	19 267	8 757	1,400	9,814 123	0,175 870	–704,267 4
2003	20 588	9 109	1,213	10,001 729	0,138 386	–3 057,800 7
2004	21 791	10 401	1,510	9,904 654	0,199 674	–335,092 1
2005	22 884	10 757	1,364	10,046 728	0,168 503	–2 225,357 1
2006	23 923	11 080	1,262	10,156 320	0,148 098	–3 809,543 9
2007	26 501	14 927	2,199	9,877 357	0,347 021	3 325,133 0
2008	28 372	15 666	2,032	10,008 100	0,310 931	2 431,416 7
2009	**30 588**	**18 340**	**2,570**	**9,921 596**	**0,425 293**	**5 396,986 3**
2010	**33 082**	**21 090**	**3,062**	**9,880 774**	**0,524 119**	**7 673,648 3**

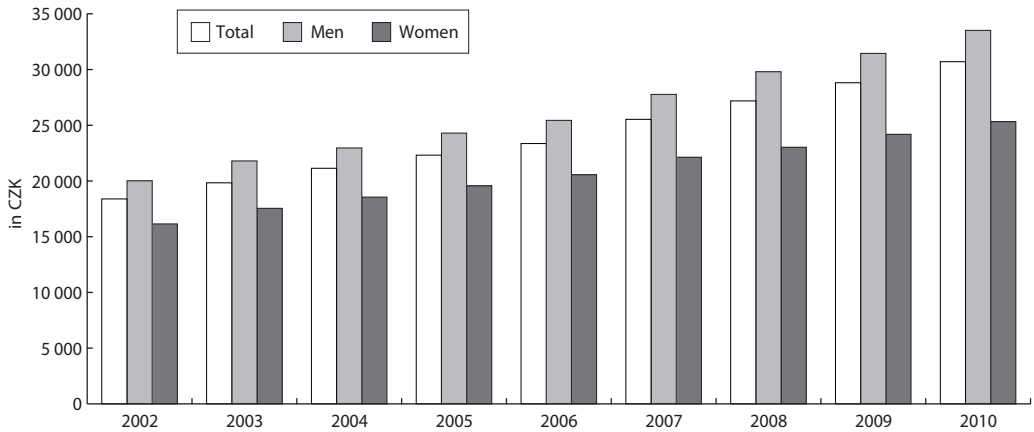
Source: own research

Table 7 Development of arithmetic mean $\bar{x}$ (in CZK) standard deviation s_x (in CZK) and skew coefficient $\sqrt{b_1(x)}$ of gross monthly wages in the Czech Republic in 2002–2008 for women, including the predictions for 2009 and 2010, and the parameter values of the three-parametric lognormal distribution estimated using the moment method of parameter estimation, including the predictions for 2009 and 2010

Year	Sample characteristics			Parameters estimated using the moment method		
	$\bar{x}$	s_x	$\sqrt{b_1(x)}$	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\varsigma}$
2002	15 088	6 945	1,766	9,340 691	0,253 414	2 157,049 1
2003	16 246	7 373	1,566	9,525 959	0,210 500	1 013,400 9
2004	16 942	8 177	1,733	9,523 956	0,246 219	1 465,763 0
2005	17 886	8 563	1,642	9,625 870	0,226 788	914,663 6
2006	18 958	8 843	1,658	9,647 974	0,230 221	1 578,258 7
2007	20 324	11 085	2,517	9,439 541	0,414 353	4 853,088 0
2008	21 585	11 641	2,442	9,519 993	0,398 566	4 949,391 3
2009	**22 973**	**13 203**	**3,068**	**9,410 412**	**0,525 282**	**7 089,223 6**
2010	**24 474**	**14 812**	**3,667**	**9,343 947**	**0,636 080**	**8 765,031 6**

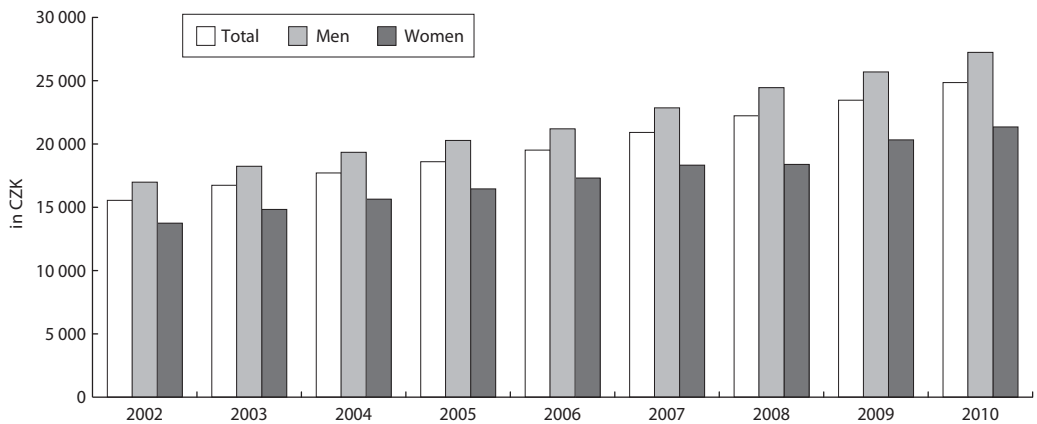
Source: own research

Graph 4 Development of the medial value of gross monthly wages (in CZK) **in the Czech Republic in 2002–2008, including the predictions for 2009 and 2010**



Source: own research

Graph 5 Development of the median gross monthly wages (in CZK) **in the Czech Republic in 2002–2008, including the predictions for 2009 and 2010**



Source: own research

assumed shape of distributions. This is caused by the fact that in the case of large sample extents, which are used for wage and income distributions, the test is strong enough to reveal the slightest deviations between the actually observed distribution and the model. Such hardly noticeable deviations have practically no importance for us and in such cases we usually "borrow" the model, see [21]. Tables 5–7 also show the predictions of the parameter values of the three-parametric lognormal distribution of gross monthly wages for 2009 and 2010 and Graphs 6 and 7 show the probability densities of the three-parametric lognormal distribution corresponding to these predicted parameter values.

Tables 8–10 present the development of quartiles of gross monthly wages in the Czech Republic in 2002–2008 including the predictions of this

Table 8 Development of quartiles (in CZK) **of gross monthly wages in the Czech Republic in 2002–2008 total for men and women including the predictions for 2009 and 2010, and parameter values of three-parametric lognormal distribution estimated using the quantile method of parameter estimation including the predictions for 2009 and 2010**

Year	Sample characteristics			Parameters estimated using the moment method		
	$\tilde{\tilde{x}}_{25}$	$\tilde{\tilde{x}}_{50}$	$\tilde{\tilde{x}}_{75}$	μ^*	$\hat{\sigma}^{2*}$	ξ^*
2002	11 944	15 545	20 215	9,663 345	0,148 549	−185,315 6
2003	12 728	16 735	22 224	10,152 882	0,217 827	−8 930,038 2
2004	13 416	17 709	23 077	10,561 268	0,109 631	−20 900,984 7
2005	14 063	18 597	24 470	10,469 561	0,147 272	−16 629,778 5
2006	14 717	19 514	25 675	10,558 943	0,137 760	−19 006,587 6
2007	15 769	20 910	27 545	10,609 483	0,143 087	−19 607,560 5
2008	16 853	22 225	29 404	10,526 808	0,184 906	−15 077,375 6
2009	**17 848**	**23 456**	**31 038**	**9,977 665**	**0,199 923**	**1 916,052 7**
2010	**18 995**	**24 849**	**32 934**	**9,962 441**	**0,229 164**	**3 634,481 8**

Source: own research

Table 9 Development of quartiles (in CZK) **of gross monthly wages in the Czech Republic in 2002–2008 for men including the predictions for 2009 and 2010, and parameter values of the three-parametric lognormal distribution estimated using the quantile method of parameter estimation, including the predictions for 2009 and 2010**

Year	Sample characteristics			Parameters estimated using the moment method		
	$\tilde{\tilde{x}}_{25}$	$\tilde{\tilde{x}}_{50}$	$\tilde{\tilde{x}}_{75}$	μ^*	$\hat{\sigma}^{2*}$	ξ^*
2002	13 415	16 985	22 604	9,188 668	0,452 625	7 199,144 0
2003	14 252	18 240	24 145	9,933 089	0,338 086	−2 360,410 4
2004	15 036	19 344	25 306	10,193 779	0,232 265	−7 392,304 1
2005	15 733	20 281	26 822	10,138 897	0,290 257	−5 027,572 6
2006	16 356	21 199	28 090	10,230 508	0,273 607	−6 537,666 0
2007	17 659	22 855	30 035	10,386 347	0,229 873	−9 558,722 0
2008	18 988	24 450	32 341	10,310 111	0,297 391	−5 584,489 9
2009	**20 171**	**25 692**	**33 843**	**9,747 471**	**0,333 616**	**8 581,098 5**
2010	**21 585**	**27 238**	**35 853**	**9,707 582**	**0,390 185**	**10 796,205 6**

Source: own research

development for 2009 and 2010, assuming that the current development is sustained. These tables also represent the estimated parameter values for the three-parametric lognormal distribution obtained using the quantile method of parameter estimation including the predicted parameter values for 2009 and 2010, which are also shown in Graphs 8 and 9. Values of test criterion χ^2 and values of the sum of absolute deviations S of the observed and theoretical frequencies for all intervals for the three-parametric lognormal curves obtained using the quantile method of parameter estimation are shown in Table 12.

Table 10 Development of quartiles (in CZK) **of gross monthly wages in the Czech Republic in 2002–2008 for women including the predictions for 2009 and 2010, and parameter values of the three-parametric lognormal distribution estimated using the quantile method of parameter estimation, including the predictions for 2009 and 2010**

Year	Sample characteristics			Parameters estimated using the moment method		
	$\tilde{x}_{25}$	$\tilde{x}_{50}$	$\tilde{x}_{75}$	μ^*	σ^{2*}	ξ^*
2002	10 341	13 746	17 727	10,065 900	0,053 721	–9 781,211 9
2003	11 042	14 831	19 281	10,763 519	0,056 753	–32 433,466 5
2004	11 594	15 642	20 293	10,973 629	0,042 497	–42 673,921 9
2005	12 158	16 454	21 426	10,983 560	0,046 948	–42 443,890 0
2006	12 881	17 311	22 530	10,900 344	0,059 020	–36 884,048 9
2007	13 744	18 328	24 017	10,659 886	0,102 547	–24 284,072 8
2008	14 612	19 388	25 330	10,689 348	0,104 952	–24 498,268 7
2009	**15 512**	**20 326**	**26 446**	**10,023 877**	**0,126 647**	**–2 232,713 6**
2010	**16 499**	**21 347**	**27 679**	**9,937 197**	**0,156 760**	**661,328 8**

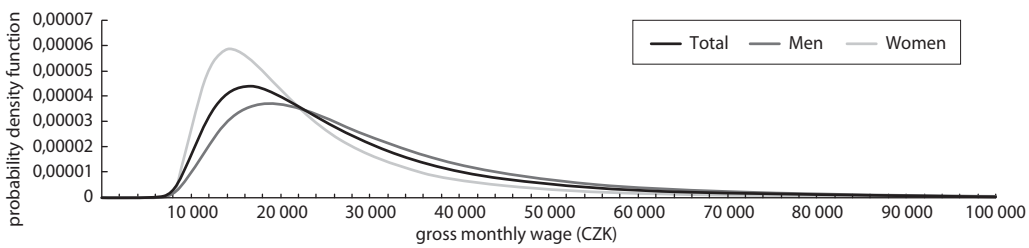
Source: own research

Table 11 Values of test criterion χ^2 and values of the sum of all absolute deviations S of the empirical and theoretical frequencies for the three-parametric lognormal curve obtained using the moment method of parameter estimation

		Set					
		Total		Men		Women	
Year	Criterion	χ^2	S	χ^2	S	χ^2	S
2002		25 576	114 691	34 003	105 690	9 696	45 033
2003		48 933	157 301	39 691	118 895	12 405	50 459
2004		67 967	226 646	46 447	142 165	23 653	94 724
2005		71 879	225 479	49 016	149 877	23 036	93 170
2006		86 368	248 955	56 518	167 314	27 491	95 430
2007		91 821	332 148	63 644	201 427	30 841	131 039
2008		94 290	341 796	68 516	212 555	29 195	132 684

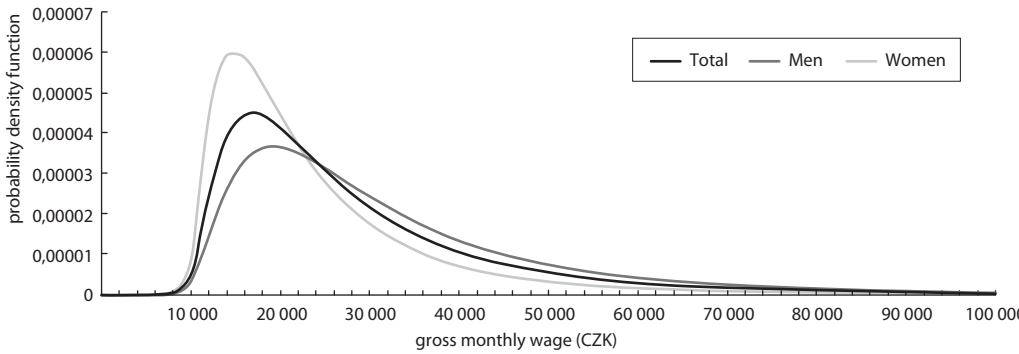
Source: own research

Graph 6 Densities of probability of the predictions of the three-parametric lognormal curves obtained using the moment method of parameter estimation for 2009



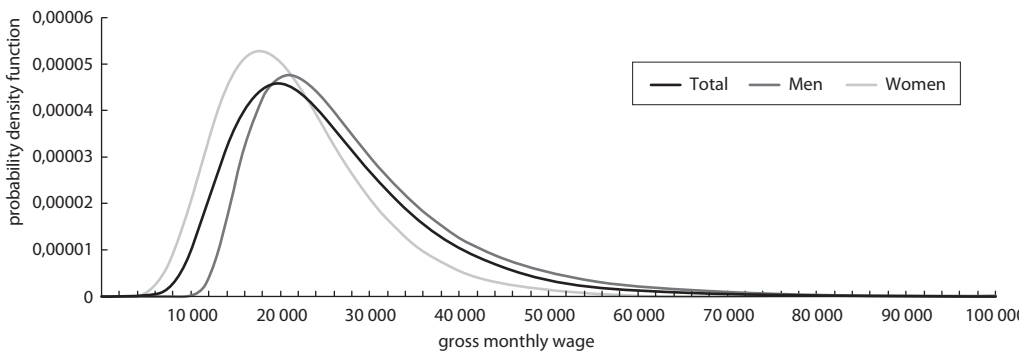
Source: own research

Graph 7 Densities of probability of the predictions of the three-parametric lognormal curves obtained using the moment method of parameter estimation for 2010



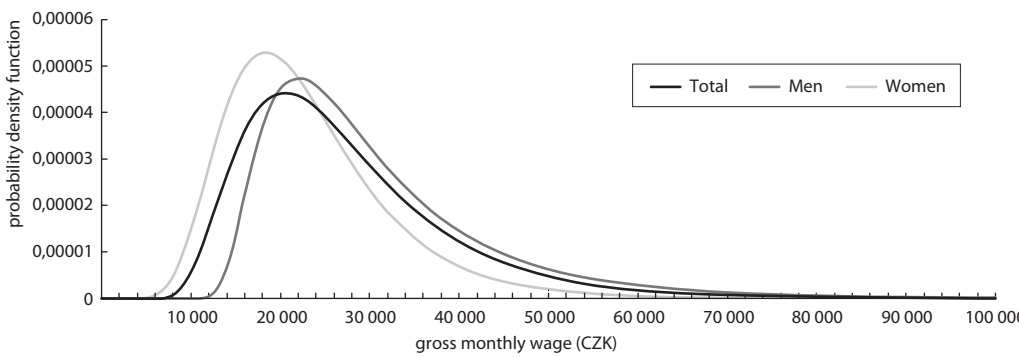
Source: own research

Graph 8 Densities of probability of the predictions of the three-parametric lognormal curves obtained using the quantile method of parameter estimation for 2009



Source: own research

Graph 9 Densities of probability of the predictions of the three-parametric lognormal curves obtained using the quantile method of parameter estimation for 2009



Source: own research

Table 12 Values of test criterion χ^2 and values of the sum of all absolute deviations S of the empirical and theoretical frequencies for the three-parametric lognormal curves obtained using the quantile method of parameter estimation

		Set					
		Total		Men		Women	
Year	Criterion	χ^2	S	χ^2	S	χ^2	S
2002		100 174	92 687	97 012	68 713	155 443	55 943
2003		365 753	547 743	165 094	270 380	230 874	291 421
2004		523 578	764 989	223 346	353 087	341 297	460 444
2005		541 900	798 876	226 467	369 835	365 954	491 762
2006		613 357	881 621	250 923	399 650	400 187	527 740
2007		617 231	930 856	272 421	436 956	344 095	502 185
2008		573 412	880 134	249 587	416 120	343 130	504 228

Source: own research

CONCLUSION

In the period 2002–2008, we can observe a significant growth of the level of gross monthly wages, both in the total set of men and women, as well as in the separate sets for both genders. This trend should continue also in the period 2009 and 2010. Throughout the observed period 2002–2008, the level of men's wages is considerably higher than the level of women's wages, and this difference continues to grow over the time and it is assumed that it will have a growing tendency also in the years 2009 and 2010.

As for the development of variability of gross monthly wages in 2002–2008, the absolute variability characteristics for this period are growing over the time, both in the total set as well as in the separate sets of men and women. Relative variability characteristics of all observed sets tend to stagnate. All the observed frequency distributions are characterized by positive skew, which means that in the given frequency distribution lower wages prevail over higher wages, which is typical for the distribution of wages and incomes. The skew of the frequency distribution is increasing over the time in all observed sets.

The parameters of three-parametric lognormal curves were estimated using two point estimation methods, namely the moment method and the quantile method. The results in Tables 11 and 12 indicate that the moment method has brought more accurate results than the quantile parameter estimation method.

Table 13 contains the prediction of the distribution of gross monthly wages for 2009 and 2010, assuming that the current development continues. These predictions are based on the moment method of parameter estimation, which yields more accurate results, as well as on the quantile method, which has yielded less accurate results. The question is to what extent the development of wage distributions will be affected by the current economic recession, see [13]; which has lead, for example, to considerable layoffs, concerning particularly workers with very low wages, which could paradoxically lead to a growth of the level of gross monthly wages accompanied by a decrease of the skew of frequency distribution. The impact of the economic recession on the development of gross monthly wages of men and women in the Czech Republic can be quantified more accurately when comparing the predictions of wage distributions with the wage distributions presented by the Czech Statistical Office. The respective data for 2009 have already been published, see Table 14.

By comparing the estimated shares of workers in gross salary wage bands for 2009 shown in Table 13 with the shares of employees in gross monthly wage bands in 2009 published by the Czech Statistical Office and shown in Table 14, we get an idea to what extent the predictions were fulfilled. The presented results reveal that in 2009 wages did not by far grow as fast as we had esti-

Table 13 Estimated shares of workers (in %) **in the bands of gross monthly wages** (in CZK) **for 2009 and 2010 by gender**

Method		Moment method						Quantile method					
Set		Total		Men		Women		Total		Men		Women	
Interval	Year	2009	2010	2009	2010	2009	2010	2009	2010	2009	2010	2009	2010
till–10 000		1,80	0,15	1,13	0,16	2,39	0,26	1,42	0,60	0,00	0,00	4,27	2,23
10 001–12 000		4,92	2,75	3,08	1,70	8,04	5,41	3,06	2,00	0,26	0,00	5,51	4,22
12 001–14 000		7,39	6,46	5,11	4,09	11,16	10,7	5,32	4,14	2,06	0,44	7,98	6,94
14 001–16 000		8,55	8,45	6,53	5,96	11,57	11,94	7,30	6,24	5,07	2,83	9,73	9,11
16 001–17 000		4,39	4,47	3,57	3,41	5,48	5,73	4,18	3,75	3,58	2,66	5,22	5,06
17 001–18 000		4,34	4,47	3,67	3,57	5,16	5,41	4,40	4,05	4,09	3,39	5,29	5,22
18 001–19 000		4,23	4,37	3,71	3,64	4,80	5,03	4,53	4,25	4,45	3,96	5,25	5,26
19 001–20 000		4,08	4,22	3,70	3,66	4,43	4,65	4,57	4,37	4,67	4,36	5,13	5,20
20 001–20 000		3,90	4,04	3,65	3,62	4,07	4,26	4,54	4,40	4,76	4,60	4,93	5,05
21 001–22 000		3,70	3,83	3,56	3,55	3,72	3,89	4,45	4,37	4,75	4,71	4,67	4,83
22 001–23 000		3,50	3,62	3,45	3,45	3,39	3,55	4,31	4,28	4,65	4,70	4,38	4,57
23 001–24 000		3,29	3,40	3,32	3,34	3,09	3,23	4,13	4,15	4,50	4,61	4,07	4,28
24 001–25 000		3,09	3,19	3,19	3,21	2,81	2,93	3,93	3,99	4,30	4,46	3,74	3,96
25 001–26 000		2,89	2,98	3,04	3,07	2,55	2,67	3,71	3,80	4,08	4,27	3,42	3,65
26 001–28 000		5,21	5,38	5,64	5,70	4,41	4,63	6,71	6,99	7,44	7,88	5,89	6,36
28 001–30 000		4,51	4,67	5,05	5,13	3,64	3,83	5,78	6,13	6,46	6,93	4,73	5,19
30 001–32 000		3,90	4,04	4,50	4,59	3,00	3,18	4,90	5,29	5,53	5,99	3,74	4,17
32 001–36 000		6,27	6,54	7,49	7,71	4,55	4,87	7,51	8,32	8,63	9,49	5,15	5,90
36 001–40 000		4,67	4,93	5,80	6,06	3,15	3,44	5,11	5,87	6,08	6,82	3,01	3,58
40 001–50 000		7,18	7,78	9,35	10,06	4,42	4,99	6,50	7,89	8,34	9,68	2,99	3,82
50 001–60 000		3,62	4,12	4,94	5,61	1,99	2,38	2,30	3,06	3,46	4,25	0,70	1,02
60 001–70 000		2,97	3,66	4,20	5,16	1,47	1,91	1,13	1,69	2,17	2,89	0,20	0,35
80 001–and more		1,63	2,48	2,32	3,54	0,68	1,09	0,20	0,37	0,67	1,07	0,01	0,03
Mean		**27 355**	**29 428**	**30 588**	**33 082**	**22 973**	**24 474**	**25 720**	**27 425**	**28 798**	**30 780**	**21 801**	**23 034**
Median		**22 771**	**23 929**	**25 762**	**27 224**	**19 304**	**20 194**	**23 456**	**24 849**	**25 692**	**27 238**	**20 326**	**21 347**
Stand. deviation		**16 796**	**19 182**	**18 340**	**21 090**	**13 203**	**14 812**	**11 198**	**12 073**	**12 722**	**13 805**	**8 831**	**9 217**
Var. coefficient		**61,40**	**65,18**	**59,96**	**63,75**	**57,47**	**60,52**	**43,54**	**44,02**	**44,18**	**44,85**	**40,51**	**40,01**

Source: own research

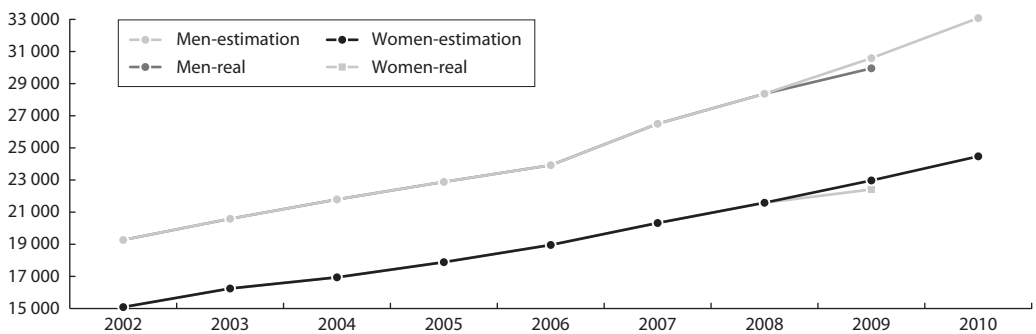
mated based on the development of wages until 2008 and it is therefore obvious that the economic recession has certainly lead to a considerable slowdown of wage growth. We can get a certain idea of the impact of the economic recession on the development of wages from Graph 10, again by comparing the prediction of the development of average gross monthly wages of men and women for 2009 with the values of average gross monthly wages presented by the Czech Statistical Office. It can be expected that the difference between the prediction and the values of average gross monthly wages published by the Czech Statistical Office for 2010 will be even bigger. Certain inaccuracies may also be caused by the fact that our calculations were based on data organized in frequency distributions, while the data presented by the Czech Statistical Office are calculated directly from microdata, which is reflected mainly in higher wages (upper percentiles).

Table 14 Shares of workers (in %) **in bands of gross monthly wages** (in CZK) **in 2009 by gender, published by the Czech Statistical Office**

Interval	Set		
	Total	Men	Women
till–10 000	3,02	2,03	4,32
10 001–12 000	4,69	2,49	7,54
12 001–14 000	6,30	3,66	9,72
14 001–16 000	7,61	5,70	10,10
16 001–17 000	4,30	3,72	5,05
17 001–18 000	4,50	4,27	4,79
18 001–19 000	4,59	4,45	4,76
19 001–20 000	4,67	4,62	4,74
20 001–20 000	4,78	4,85	4,69
21 001–22 000	4,54	4,70	4,34
22 001–23 000	4,34	4,54	4,09
23 001–24 000	4,03	4,30	3,67
24 001–25 000	3,84	4,10	3,51
25 001–26 000	3,64	3,91	3,28
26 001–28 000	6,25	6,66	5,71
28 001–30 000	4,98	5,66	4,09
30 001–32 000	4,01	4,72	3,08
32 001–36 000	5,67	6,90	4,07
36 001–40 000	3,64	4,53	2,48
40 001–50 000	4,83	6,15	3,11
50 001–60 000	2,13	2,84	1,20
60 001–70 000	1,79	2,43	0,94
80 001–and more	1,88	2,77	0,72
Mean	**26 677**	**29 953**	**22 414**
Variation coefficient	**0,97**	**1,05**	**0,66**

Source: own research

Graph 10 Development of average gross monthly wages in the Czech Republic by gender in 2008–2009 and prediction of gross monthly wages for 2009 and 2010



Source: www.czso.cz + own research

References

[1] BARTOŠOVÁ, J. Logarithmic-Normal Model of Income Distribution in the Czech Republic. *Austrian Journal of Statistics.* 2006, Vol. 35, No. 23, pp. 215–222. ISSN 1026-597x.

[2] BARTOŠOVÁ, J. Pravděpodobnostní model rozdělení příjmů v České republice. *Acta Oeconomica Pragensia.* 2007, Vol. 15, No. 1, pp. 7–12. ISSN 0572-3043.

[3] BARTOŠOVÁ, J. Analysis and Modelling of Financial Power of Czech Households. *Aplimat – Journal of Applied Mathematics.* 2009, Vol. 2, No. 3, pp. 31–36, ISSN 1337-6365.

[4] BARTOŠOVÁ, J., BÍNA, V. Modelling of Income Distribution of Czech Households in Years 1996–2005. *Acta Oeconomica Pragensia.* 2009, Vol. 17, No. 4, pp. 3–18. ISSN 0572-3043.

[5] BENÍŠEK, J. 2007. Výpočet bezpečnostní přirážky a úloha směrodatné odchylky v aktuárské praxi. *E + M Ekonomie a Management.* 2007, Vol. 10, No. 3, pp. 104–108. ISSN 1212-3609.

[6] BÍLKOVÁ, D., BUDINSKÝ, P., VOHÁNKA, V. *Pravděpodobnost a statistika.* 1. vyd. Plzeň: Vydavatelství a nakladatelství Aleš Čeněk, 2009. 639 s. ISBN 978-80-7380-224-0.

[7] BÍLKOVÁ, D. Pareto Distribution and Wage Models. *Journal of Applied Mathematics.* 2009, Vol. 2, No. 3, pp. 37–46. ISSN 1337-6365.

[8] BÍLKOVÁ, D. Paretovo rozdělení a vývoj mzdových rozdělení v České republice v letech 2001–2006. *Statistika.* 2009, Vol. 89, No. 1, pp. 32–52. ISSN 0322-788x.

[9] BÍLKOVÁ, D. Application of Lognormal Curves in Modeling of Wage Distributions. *Journal of Applied Mathematics.* 2008, Vol. 1, No. 2, pp. 341–352. ISSN 1337-6365.

[10] BÍLKOVÁ, D. Modelování mzdových rozdělení v České republice v letech 2004 a 2005 s využitím logaritmicko-normálních křivek a křivek Pearsonova a Johnsonova systému. *Statistika.* 2008, Vol. 88, No. 2, pp. 149–166. ISSN 0322-788x.

[11] BÍLKOVÁ, D. *Příjmová rozdělení: modelování v letech 1956–1992 a předpovědi pro roky 1995 a 1997.* Doktorská disertační práce. Praha: Vysoká škola ekonomická, 1996. 187 s.

[12] BÍLKOVÁ, D. Vývoj příjmových rozdělení v letech 1956–1992 a jejich předpovědi pro rok 1995 a 1997. *Politická ekonomie.* 1995, Vol. 43, No. 4, pp. 510–531. ISSN 0032-3233.

[13] BÍLKOVÁ, D. O vlivu procesu transformace na předpovědi rozdělení příjmů. *Statistika.* 1995, Vol. 32, No. 5, pp. 197–205. ISSN 0322-788x.

[14] BÍLKOVÁ, D. K modelování příjmových rozdělení lognormálními křivkami. *Statistika.* 1994, Vol. 31, No. 11, pp. 453–467. ISSN 0322-788x.

[15] CYHELSKÝ, L. *Úvod do teorie statistiky.* 2. vyd. Praha: SNTL/ALFA, 1981. 352 s. 04-318-81.

[16] HÁTLE, J., HUSTOPECKÝ, J., NOVÁK, I. *Modelování a krátkodobá předpověď příjmových rozdělení* [Výzkumná zpráva č. 66]. Praha: Výzkumný ústav sociálně-ekonomických informací a Vysoká škola ekonomická v Praze, 1975. 95 s.

[17] JÍLEK, J., FRIEDLAENDER, J., MORAVOVÁ, J., BÍLKOVÁ, D. Household Incomes, Expenditures and Their Changes in the Last Years. *Prague Economic Papers.* 1995, 4 (1), pp. 41–64. ISSN 1210-0455.

[18] JÍLEK, J., MORAVOVÁ, J., BÍLKOVÁ, D., FRIEDLAENDER, J. *Příjmy a výdaje domácností v posledních letech* [Výzkumná zpráva]. Praha: Nadace pro výzkum sociální transformace START, 1995. 90 s.

[19] JOHNSON, N. L., KOTZ, S., BALAKRISHNAN, N. *Continuous Univariate Distributions.* Vol. 1, 2nd edition. USA: John Wiley & Sons, 1994. 756 p. ISBN 0-471-58495-9.

[20] KOTZ, S., BANKS, D. L. READ, C. B. *Encyclopedia of Statistical Sciences.* Update Volume 2. USA: Wiley-Interscience, 1998. 745 p. ISBN 978-0471119395.

[21] MORAVOVÁ, J. FRIEDLAENDER, J., BÍLKOVÁ, D. Změny příjmů a výdajů sociálních skupin obyvatelstva v období transformace. *Statistika.* 1995, Vol. 32, No. 3, pp. 97–117. ISSN 0322-788x.

[22] NOVÁK, I. *Vývoj mzdových rozdělení v národním hospodářství a průmyslu ČSSR v létech 1959–1964 a možnosti jejich extrapolace.* Habilitační práce. Praha: Vysoká škola ekonomická, 1964. 119 s.

[23] PACÁKOVÁ, V., SIPKOVÁ, L. Generalized Lambda Distributions of Household\s Incomes. *E + M Ekonomie a Management.* 2007, Vol. 10, No. 1, pp. 98–107. ISSN 1212-3609.

[24] ČSÚ, 2009. <www.czso.cz>.

What Could Fuzzy Logic Bring to Statistical Information Systems?

Miroslav Hudec[a] | *INFOSTAT, Bratislava*

Abstract

The aim of the paper is to present the applicability of the fuzzy logic for statistical information systems in order to improve work with statistical data. The improvement offers the approximate reasoning in order to solve problems in a way that more resembles human logic. The paper examines the fuzzy logic approach, emphasizes situations when the two-valued (crisp) logic is not adequate and offers solutions based on fuzzy logic. The first step of using data is its selection from a database. Although the Structured Query Language (SQL) is a very powerful tool, it is unable to satisfy needs for data selection based on linguistic expressions and degrees of truth. For this purpose the fuzzy generalised logical condition (GLC) was developed. Later researches have shown that the GLC formula is suitable for other processes concerning data, namely data classification and data dissemination.

Key words:

fuzzy database query, fuzzy generalised logical condition, applicability of fuzzy query

INTRODUCTION

The following statement holds for many cases: If something is precisely defined it is easy to implement but it could be far from reality; if something contains ambiguities and uncertainties in its definition it is harder to implement but it is better connected to reality. This statement is generally true but there are methodologies capable to easier implement ambiguities and uncertainties.

One of these methodologies is fuzzy set and fuzzy logic. "Fuzzy logic allows us to bring the operation of information systems closer to the working methods of humans." [6]. Users frequently deal with terms such as high unemployment rate, the majority of, low migration level, etc. These terms include a certain vagueness or uncertainty that information systems based on two-valued logic {true, false} do not understand and therefore cannot use. The fuzzy set theory works with the gradation, uncertainty and ambiguity described by linguistic expressions when sharply defined criteria could not be created. More about this kind of vagueness and uncertainty can be found in [24].

Statistics is a promising area to develop and implement the fuzzy logic concept. Large data collections are mainly stored in relational databases. Users need this data for the analysis, decision making or just to find interesting information. In many cases the rigid or precise definition of an analysed task cannot be created or the user wants to obtain additional information.

The first step of using data is in many cases its selection from relational databases. The SQL is used for this purpose. Although the SQL is a powerful tool, it uses the two-valued logic in the selection process. It causes that the small er-

[a] *INFOSTAT, Dubravska cesta 3, 845 24 Bratislava, Slovakia, e-mail: hudec@infostat.sk*

ror in data values or in cases when the user cannot unambiguously define the criterion by crisp boundaries may involve some inadequately selected or non-selected data. As a solution, the use of fuzzy logic is proposed and described in the paper. This area of research is not new, but there are still many possibilities for the improvement of existing approaches and for creating new approaches. Some fuzzy query implementations have been designed e.g. [2] and [22]. Although there are some variations according to the particularities of different implementations, the answer to a fuzzy query sentence is generally a list of records, ranked by the degree of matching [3]. Issues and perspectives of fuzzy querying can be found in [14]. In order to bring benefit of the fuzzy logic to database users and to make an easy to use querying tool the generalized logical condition (GLC) for database queries was created in [9]. The implementation of the GLC for statistical databases was discussed in [10].

Later researches have shown that the GLC formula can be used for other processes concerning data. In our researches we are mainly interested in information systems improvements for data classification and data dissemination areas. Application for municipalities classification by fuzzy expert systems in the Slovak Republic was proposed in [12]. Another way of classification has been found during work on fuzzy database queries. The GLC could be used for data classification by generating queries from fuzzy rules [11]. Generating queries from fuzzy rules is an active research field. Some fuzzy classification implementations have been proposed in [3] and [21]. The last two papers contain information about other researches and implementations in this field.

The data dissemination becomes a very important area especially for organisations that provide wide variety of data to professionals and broad audience. One of dissemination channels is the internet. Many articles and books about website design and appropriate design of tables and graphs can be found, for example in [19] and [5] respectively. In our paper advantages of the fuzzy approach in the process of finding and selecting adequate data for statistical websites are pointed out.

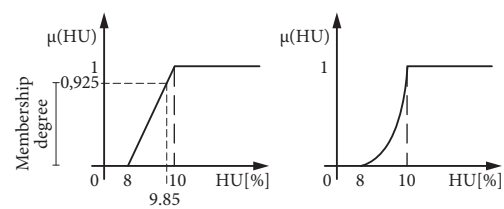
The paper is organised as follows: The crucial differences between crisp and fuzzy logic are reviewed in Section 2. SQL and fuzzy database queries are examined in section 3. The disadvantage of SQL is pointed out and an alternative approach based on the fuzzy GLC is analysed. In the last part of this section the proposed concept is demonstrated by a case study. Section 4 emphasises that developed GLC and fuzzy queries can be used for many other purposes. The usefulness of the GLC for data classification and dissemination is discussed in this section. Finally some conclusions are drawn.

1 FUZZY LOGIC

The core of both crisp logic and fuzzy logic is the idea of a set. In the crisp set theory an element belongs or does not belong to a set. For example, consider a set called high unemployment (HU) defined as follows: HU = {x| unemployment(x) ≥ 10%} where x is a region. It means that region with 9.9% unemployment does not belong to the HU but region with 10% belongs. These constraints are drawback when the boundaries between values of some attributes are continuous.

The concept of fuzzy sets was initially introduced in [23] where was observed that more often than not, precisely defined criteria of belonging to a set could not be defined. The fuzzy set and logic theory brings a paradigm in work with the gradation, uncertainty and ambiguity described by linguistic expressions. This gradation is described by a membership function μ valued in the interval [0, 1]. The HU example can be presented by fuzzy sets shown in Figure 1. User could define that the unemployment equal and higher than 10% is HU, the unemployment smaller than 8% definitely is not HU and unemployment between 8% and

Figure 1 Fuzzy sets for high unemployment concept



Source: own research

10% partially belongs to the HU concept. The closer is the unemployment to 10% the stronger it belongs to the high unemployment concept. The fuzzy set gives answer for the following question: How compatible is 9.85% unemployment with the HU concept? The answer is 0.925 or territorial unit with 9.85% unemployment rate is a very strong member of the HU concept. If territorial unit's unemployment is 9%, it is a moderate member of the HU concept.

2 DATABASE QUERIES

2.1 SQL and its limitation

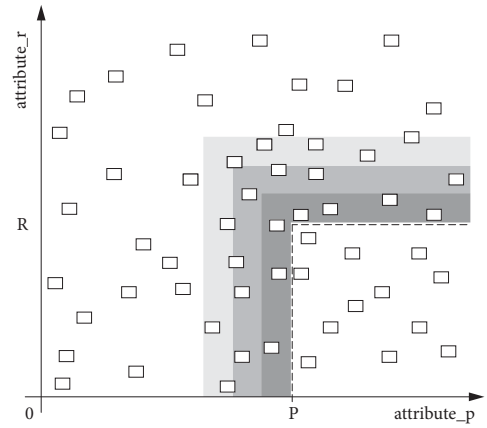
The SQL was initially developed in [4]. Since then the SQL has been used in many relational databases and information systems for data selection. The use of SQL may be regarded as one of the mayor reasons for the success of relational databases in the commercial world [20].

Generally speaking, users search databases in order to obtain data needed for analysis, decision making or just to satisfy their curiosity. Situations when constraints of crisp logic in querying processes may occur are examined by the following example:

```
select <attribute(s) list>
from <table(s) list>
where attribute_p > P and attribute_r < R;
```

The best way how to describe limitations of a SQL query is the graphic mode shown in Figure 2. Values P and R delimit the space of selected data. The user cannot obtain any information about records that are close to meet the query criterion (areas marked with grey shadows). The area marked with the darkest grey shadow contains records that almost meet the intent of the query. It means that the record would not be selected even if it is extremely close to meet the query criterion. Records belonging to shadowed areas could be potential customers and direct marketing could attract them or territorial units which almost satisfy criterion for some financial support for example. In case of no data is selected by SQL, there is not any information concerning possible records that almost meet the query criterion. This is the penalty paid to use the crisp logic in selection process.

Figure 2 The result of the SQL query



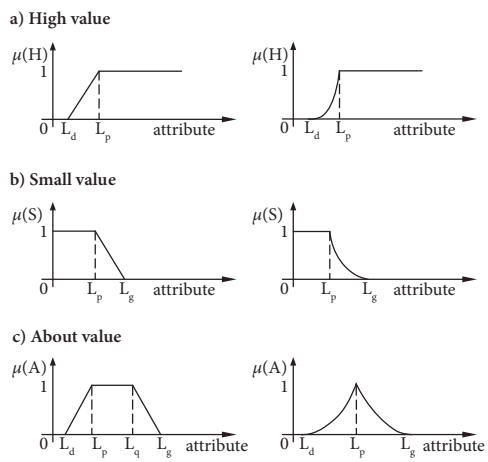
Source: own research

The new way of evaluating the *WHERE* clause of a SQL query and its further applicabilities are explained in next subsections.

2.2 Fuzzy query idea

SQL conditions in queries contain these comparison operators: > , < , = , ≠ and *between* when numerical attributes are used. These crisp logical comparison operators are adapted for fuzzy queries in the following way: operator > (greater than) was improved with fuzzy set "High value" (Figure 3a), operator < (less than) was improved with fuzzy set "Small value"

Figure 3 Fuzzy sets



Source: own research

(Figure 3b) and operator = (equal) was improved with fuzzy set "About value" (Figure 3c). Operator ≠ is the negation of the operator = so this operator is not further analysed. Analogous statement is valid for the operator *between* because it is similar to the operator = from the fuzzy point of view.

For the further reading it is important to define the Query Compatibility Index (QCI). The QCI is used to indicate how the selected record satisfies a query criterion. The QCI has values from the [0, 1] interval with the following meaning: 0 – record does not satisfy the query, 1 – record fully satisfies the query, interval (0, 1) – record partially satisfies the query with a distance to the full query satisfaction.

According to the above mentioned facts, the example of a fuzzy query e.g. to find appropriate areas for tourism has the following form:

```
select district
from table
where air_pollution is Small and number_of_
sunny_days is High;
```

The meaning of a fuzzy query is obvious at first glance because it is expressed with linguistic expressions. The shape of membership function (μ(x)) (Figure 3) can be adjusted according to user's requirements without changing the meaning of a query. In this example the fuzzy set with $L_p = 10$ and $L_g = 15$ units of measured pollutant and shape as from Figure 3b) on the left side describes the small air pollution concept. The fuzzy set with $L_d = 140$ and $L_p = 150$ days and shape as from Figure 3a) on the left side describes the high number of days with sunshine concept. The result of this query is in the table 1.

In this example μ(P) denotes the membership degree to the small pollution fuzzy set and μ(S) denotes the membership degree to the high number of days with sunshine fuzzy set. The QCI (calculated as a minimum) represents membership degree to the small air pollution and high number of sunny days concept. If user wants to make some activity in the appropriate district and it is not possible to realise it in districts D7, user can choose the district D5 that almost satisfies the intent of the query. Again, if it is not possible to choose the district D5 the next choice is D2 and so on. It is important to emphasize that ranking is not done by one indicator, their linear combination by weighted coefficients, etc. Both indicators have the same importance and ranking is done according to the satisfying the concept created in the query criterion.

2.3 Fuzzy query realisation

The starting point of our research was the following premise: To make easy to use data selection by concepts and to access to relational databases in the same way as SQL does. Suggested querying process consists of two main steps. In the first step all records that have the membership degree to the condition defined by linguistic expressions in the *WHERE* clause greater than zero (QCI > 0) are selected from database. For this purpose the GLC for the *WHERE* part of the SQL was created and described in [9]. The GLC has the following structure:

$$\text{WHERE} \bigotimes_{i=1}^{n} (a_i \circ L_{xi}) \tag{1}$$

where *n* denotes number of attributes with fuzzy constraints in a *WHERE* clause of a query,

$$\otimes = \begin{cases} and \\ or \end{cases}$$

where *and* and *or* are fuzzy logical operators, and

Table 1 Areas conductive to the tourism

District	Pollution (P)	μ (P)	Number of sunny days (S)	μ (S)	QCI	
D2	9.5	1	147	0.7	0.7	←
D3	11	0.8	145	0.5	0.5	
D5	10.2	0.96	149	0.9	0.9	←
D7	8.2	1	161	1	1	
D8	14.1	0.18	160	1	0.18	

Source: own research

$$a_i \circ L_{ix} = \begin{cases} a_i > L_{di}, & \text{a}_\text{i} \text{ is High} \\ a_i < L_{gi}, & \text{a}_\text{i} \text{ is Small} \\ a_i > L_{di} \text{ and } a_i < L_{gi}, & a_i \text{ is About} \end{cases} .$$

where a_i is a database attribute, L_d is the lower bound and L_g is upper bound of a linguistic expression described by fuzzy set. Two types of fuzzy set for High, Small and About expressions are shown in Figure 3.

In this step lower and/or upper bounds of linguistic expressions (fuzzy sets) are used as parameters for database query criteria. Let take the *WHERE* clause from the previous query:

> where air pollution is Small and number of sunny days is High.

Parameters L_p and L_g are used to define meaning of the subcriterion "air pollution is Small". User could state that district with measured pollutant less than 10 units fully belongs to the analysed concept and the parameter L_p set this state: $L_p = 10$ units. District with air pollution between 10 and 15 partially belong to the concept air pollution is Small. The closer is the air pollution to 10 units the stronger it belongs to the small air pollution concept. User could state that district with air pollution higher than 15 units does not belong to small air pollution concept and the parameter L_g is used to set this state: $L_g = 15$ units. Similar discussion holds for the subcriterion "number of sunny days is High".

According to the parameters of fuzzy sets and the GLC (1), fuzzy query is converted into the following SQL structure:

> where air pollution < 15 and number of sunny days > 140.

This *WHERE* clause ensure that query selects all records with QCI > 0 from a database.

In the second step the chosen analytical form of the fuzzy set is used to calculate the membership degree of each selected record to appropriate fuzzy set e.g. pollution value to concept of low pollution and number of sunny days to concept of high number of sunny days. Finally, the QCI value for each selected record is calculated. When a classical query contains more than one condition in the *WHERE* clause *and* and *or* logical operators are used. In classical case there exists only one logical function for *and* and *or* operators because the subcriterion is satisfied (value 1) or not (value 0). In fuzzy logic there exist many functions describing *and* operator (these functions are called t-norms) and *or* operator (these functions are called t-conorms) because each of subcriterions can be fully or partially satisfied. More about t-norm and t-conorm function could be found in e.g. [15]. For example the territorial unit satisfies the high number of days with sunshine concept with 0.5 and the low air pollution with 0.8. Both conditions are partially satisfied so the {0, 1} logic is not useful. It is needed to combine membership degrees so that the total result of a query can be expressed. The following t-norm functions can be used [18] for logical *and* operator:

- minimum:

$$\text{QCI} = \min(\mu_\text{i}(\text{a}_\text{i})), \qquad \text{i} = 1,..., \text{n} \tag{2}$$

- product:

$$\text{QCI} = \prod_{\text{i}=1}^{\text{n}} (\mu_i(\text{a}_\text{i})), \tag{3}$$

- bounded difference (BD)

$$\text{QCI} = \max(0, \sum_{i=1}^{n} \mu_i(a_i) - n + 1) \tag{4}$$

The following t-conorm functions can be used [18] for logical or operator:

- max

$$\text{QCI} = \max(\mu_i(\text{a}_\text{i})), \qquad \text{i} = 1,..., \text{n} \tag{5}$$

- bounded sum (BS)

$$\text{QCI} = \min(1, \sum_{i=1}^{n} \mu_i(a_i)) \tag{6}$$

where $\mu_i(a_i)$ denotes the membership degree of the attribute a_i to the i-th fuzzy set. The min t-norm takes into account the lowest value of membership

degrees to fuzzy sets (0.5 in previous example). The product t-norm takes into account all membership degrees and balances the query truth membership value across each of conditions in the *WHERE* clause (0.4 in previous example). The whole process concerning data selection by the GLC can be found in [9].

2.4 Case study

Data from the Urban and Municipal Statistical database [1] are used for case study. This database is in official use at the Statistical Office of the Slovak Republic. In this case study, districts with high length of road and small area size are sought. The high road infrastructure density is analysed as an illustrative example. The query has the following form:

```
select district
from table
where roads is High and area is Small;
```

The road length indicator is represented by High fuzzy set with these parameters $L_d = 150$ km and

Table 2 Result of fuzzy query

District	Roads [km]	Area [km²]	μ (Road)	μ (Area)	QCI
Bratislava I	335.1	9.6	1	1	1
Senec	269.1	359.9	1	1	1
Piešťany	305.6	381.1	1	1	1
Myjava	563.9	327.4	1	1	1
Púchov	320.9	375.4	1	1	1
Bytča	231	281.6	1	1	1
Kysucké Nové Mesto	269.9	173.7	1	1	1
Detva	567.2	449.2	1	1	1
Žarnovica	366.6	425.5	1	1	1
Považská Bystrica	**324.5**	**463**	**1**	**0.913**	**0.913**
Sabinov	**220.8**	**483.5**	**0.888**	**0.777**	**0.777**
Šaľa	**206.9**	**355.9**	**0.713**	**1**	**0.713**
Poltár	**207.4**	**476.1**	**0.713**	**0.826**	**0.713**
Ilava	**205.8**	**358.5**	**0.7**	**1**	**0.7**
Dolný Kubín	**197.8**	**491.8**	**0.6**	**0.721**	**0.6**
Žiar nad Hronom	*249.8*	*517.6*	*1*	*0.549*	*0.549*
Zlaté Moravce	*226.4*	*521.2*	*0.95*	*0.525*	*0.525*
Hlohovec	*187.1*	*267.2*	*0.463*	*1*	*0.463*
Pezinok	*176.9*	*375.5*	*0.338*	*1*	*0.338*
Bánovce nad Bebravou	*172.5*	*461.9*	*0.275*	*0.921*	*0.275*
Partizánske	*168*	*301.2*	*0.225*	*1*	*0.225*
Tvrdošín	*164.9*	*478.9*	*0.188*	*0.807*	*0.188*
Svidník	*164.5*	*549.6*	*0.175*	*0.336*	*0.175*
Nové Mesto nad Váhom	*528.5*	*580*	*1*	*0.133*	*0.133*
Gelnica	*163.6*	*584.4*	*0.175*	*0.104*	*0.104*
Krupina	*334.9*	*584.9*	*1*	*0.101*	*0.101*
Levoča	*157.1*	*357.2*	*0.088*	*1*	*0.088*
Spišská Nová Ves	*388.9*	*587.4*	*1*	*0.084*	*0.084*
Topoľčany	*371.8*	*597.7*	*1*	*0.015*	*0.015*

Source: [1]

$L_p = 230$ km and the shape as from Figure 3a) on the left side. The Small fuzzy set with parameters $L_p = 450$ km^2 and $L_g = 600$ km^2 and shape as from Figure 3b) on the left side describes the district area indicator.

The result of fuzzy query is shown in Table 2. The value of min t-norm (2) is used for the calculation of the QCI. The Table 2 shows nine districts fully satisfying the query; one district is extremely close to satisfy the query (marked with the stronger bold text) and another five districts are close to meet the query criterion. These five records are marked with the lighter bold text. It means for example that even small changes in attributes could imply that another district fully satisfies the query. If SQL were used, this additional valuable information would remain hidden. Territorial units are distinguished according to gradation of belonging to the concept (the query criterion).

If SQL were used, the criterion would be as follows:

where roads > 230 and area < 450.

The result of this criterion is shown in Table 3. The difference between information in the Table 2 and Table 3 is obvious. Records marked as bold and italic in Table 2 are not selected by the SQL query and the last three rows from the Table 2 is not possible to calculate because SQL query selects data only whereas fuzzy query selects data and calculates additional information.

Fuzzy queries reduce the risk of obtaining empty answer. In situation when no data is selected by the SQL, fuzzy query can inform that there are some records that almost meet the query criterion. It means that all data marked with bold and italic text in Table 2 are selected only and the distance to full query satisfaction for these records is calculated. It is not need to rearrange the fuzzy query in order to select some records.

2.5 Some fuzzy query characteristics

The SQL is a very powerful and useful query language, but only a query language. In this research the core of SQL remains intact and the extension is done to improve the querying process. Adding some flexibility to the SQL increases effectiveness and comprehensibility of the data selection. As metadata are used to explain the meaning of figures, linguistic expressions are used to explain the meaning of a query. The fuzzy approach improves the SQL with approximate reasoning. The intent of query based on fuzzy logic is not to select more data but to select better data. The advantages of this approach for users are as follows [10]:

- the connection to the database and the data accessing do not have to be modified,
- users do not need to learn a new query language,
- the querying process supports the (quasi) natural language,
- presenting of obtained data is in similar way as from SQL but with additional valuable information,
- users see data "behind the corner" (grey areas on Figure 2 and bold text in Table 2).

Database querying languages based on the fuzzy logic need additional calculations in comparison with SQL counterpart. This constatation also holds for the methodology suggested in this article. The first additional step consists of conversion from linguistic expressions to the SQL structure. The second additional step, activated immediately after records are selected from database, consists of QCI calculation for each of selected record. This additional amount of calculation is balanced with additional information obtained from databases.

Table 3 Result of SQL query

District	Roads [km]	Area [km^2]
Bratislava I	335.1	9.6
Senec	269.1	359.9
Piešťany	305.6	381.1
Myjava	563.9	327.4
Púchov	320.9	375.4
Bytča	231	281.6
Kysucké Nové Mesto	269.9	173.7
Detva	567.2	449.2
Žarnovica	366.6	425.5

Source: [1]

There is no competition between querying based on crisp and fuzzy logic. A fuzzy database query provides flexibility for the inclusion of records that are close to meet the query criterion (potential candidates) and to calculate additional valuable information. SQL database queries are useful when clean and exact boundary between selected and non selected data is required.

3 FUZZY QUERIES AND THEIR FURTHER APPLICABILITY IN STATISTICS

The statement that fuzzy querying engines gives new possibilities for data selection has proven in the previous chapter. This chapter draws attention to applicability of fuzzy queries for a broader usage. In this section the improvements of data classification and data dissemination by fuzzy logic and the GLC are examined.

There exist many other applications of fuzzy logic and fuzzy queries. Some of them are:

- case based reasoning realised inside relational database using fuzzy query approach [16],
- extension of fuzzy query for data mining and knowledge discovery [17],
- fuzzy logic approach can be used in GIS for many purposes from finding locations to spatial data analysis. One example can be found in [13].

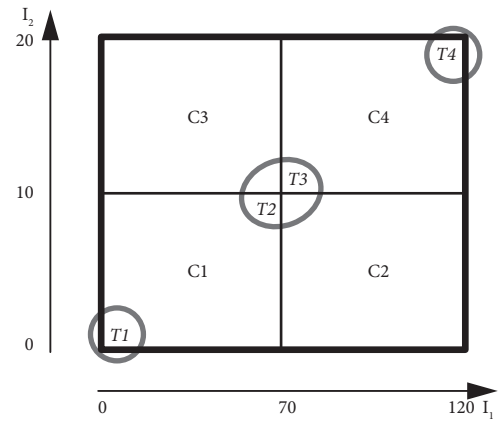
The above mentioned areas where fuzzy logic and fuzzy queries could be used are mentioned only to point out how wide are possibilities to use fuzzy logic in information systems.

3.1 Data classification

3.1.1 Crisp and fuzzy classification comparison

In classification by crisp tools, classes have sharp boundaries. If values of indicators are similar for two objects (customers, territorial units), they are similar too but it could imply that objects may fall into different classes. The classification diagram presented in Figure 4 shows this situation in graphical mode. Objects are divided into four classes from class C1 (the smallest) to class C4 (the biggest). This method treats the top rated object T4 in the same way as T3. Units T2 and T3 have

Figure 4 Crisp classes



Source: own research

similar indicators values. However, T2 and T3 are treated in different classes.

Expert systems offer a good support for classification but limitations of crisp logic may occur. The following question arises: How to solve this problem without additional calculation from user's point of view? The answer is fuzzy logic. In fuzzy classification classes do not have sharp boundaries and a classified object can belong to more than one overlapping class. Belonging to a fuzzy class depends of the membership degree to the relevant class.

The fuzzy approach gives two main ways for solving classification tasks: fuzzy systems and generating fuzzy queries from previously created fuzzy rules. The first way is an extension of expert systems by fuzzy sets and fuzzy logic. Fuzzy systems and their applicability are examined in details in [18]. The fuzzy inference system (FIS) from the Mat Lab software was used to create and solve municipalities classification model [7] and [12].

By reason that the emphasis of this paper is on the GLC and its applicability, classifications by fuzzy systems are not further considered in the paper. The idea for classification by the GLC has been found during work on fuzzy database queries. Researches have shown that the GLC formula (1) could be used in data classification [11]. Queries are equivalent with the *IF* part of the rules and result of the query are records that

fully or partially belong to the output class representing the *THEN* part of the rules. Classification by the fuzzy system and by the GLC has the same rule base structure but ways how these rules are calculated in order to obtain solution is different.

3.1.2 Classification by the GLC

"Fuzzy queries sentences are structured definitions of fuzzy concept. Under this assumption, fuzzy queries can be automatically generated by fuzzy rule based classifiers" [3]. This paper illustrates the classification using above described fuzzy queries and the GLC. The difference is in the added clause *CLASSIFY_INTO*. The *CLASSIFY_INTO* clause specifies the name of the output class to which selected records satisfying query are classified. This membership degree is also membership degree to the appropriate output class. The structure of a query is as follows:

classify_into [class$_c$]
select <attribute(s) list>
from <table(s) list>
where $\left[\bigoplus_{j=1}^{m}\bigotimes_{i=1}^{n}(a_i \circ L_{xij})\right]_C$

where the logical operator ⊗ from (1) describes *IF* part of the rule, ⊕ is the logical *or* operator that merges those *m IF* parts of rules that have the common *THEN* part or the same output class *c*, *n* is the number of attributes inside the *IF* part of the rule.

Object can belong to more than one class with different membership degrees. The rank of object is calculated by the aggregation of class coefficient where object belongs and its membership degree to these classes respectively using the following equation:

$$R_O = \sum_{l=1}^{L} \mu_{Ocl} K_l \tag{7}$$

where L is number of classes, μ_{Ocl} is the membership degree of object O to class C_l and K_l is the parameter describing class C_l.

Advantages of this approach are as follows [10]:

- queries select only those records that will be classified. Records that do not belong to any class are not needlessly selected;
- data preparation to the adequate input vector or matrix like for fuzzy systems is not needed;
- presentation of results in a useful and understandable form for example in the xls format could be easy implemented.

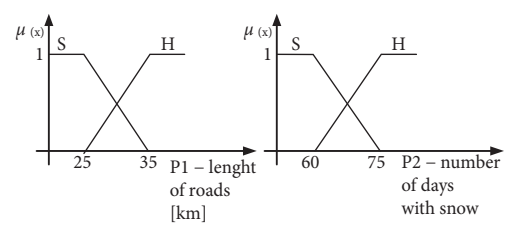
3.1.3 Case study

The classification feasibility of the GLC is illustrated with the purpose of rough planning of road maintenance requirements in winter. Detailed classification model which contains three indicators fuzzified into five fuzzy sets and 93 rules in the rule base can be found in [12] where the model was solved by fuzzy system using the Mat Lab software. In order to present classification by fuzzy queries and the GLC on illustrative example, the model is reduced. Data from the same database as for the fuzzy query case study was used. In this example two indicators are included and fuzzified into two fuzzy sets: length of roads in kilometres (Road) and number of days with snow (Snow). These sets are shown in Figure 5.

This example contains four fuzzy rules with the following structure:

- if Road is Small and Snow is Small Then Maintenance is Small;
- if Road is Small and Snow is High Then Maintenance is Medium;
- if Road is High and Snow is Small Then Maintenance is Medium High;
- if Road is High and Snow is High Then Maintenance is High.

Figure 5 Fuzzy sets small (S) and high (H) for Roads and Snow indicators



Source: own research

According to the rule base and the GLC (1) four fuzzy queries are created. The query for the Small output class has the following form:

```
classify_into Small
select municipality
from table
where roads is Small and snow is Small;
```

The percentage of requirements parameter (K) for the winter road maintenance can be associated with each fuzzy output class: for instance class S (Small) gets 10%, class M (medium) gets 35%, class MH (Medium High) gets 65% and municipalities from class H (high) gets 90% from considered amount of resources. Table 4 shows ranking results for some municipalities.

The fuzzy classification allows softer classification and ranking among municipalities (municipalities marked with the bold text in Table 4). In case of crisp classification these municipalities may be classified into different classes and it will cause difference between required needs and obtained resources. To avoid this disadvantage the user has to create very high number of output classes and rules if he wants to use crisp classification tools.

Territorial units that partially belong to more than one class are treated in all classes where they have partially membership. If data values of attributes are similar for territorial units, they are similar treated and get nearly the same percentage of resources for example.

The similar reason holds for choosing between fuzzy and crisp classification and between fuzzy and crisp selection. Fuzzy classification provides flexibility for the classification with the gradation of belonging to overlapped classes. Crisp classification is useful when clean and exact boundary between output classes is required.

Although mathematics based on fuzzy sets has greater expressive power than classical mathematics based on crisp sets, the usefulness depends critically on our capability to construct appropriate number of fuzzy sets, describe their membership functions and create all relevant fuzzy rules.

3.2 Data dissemination

Dissemination of statistical data targeting web-based audience is one of important tasks of statistical organisations. This poses a significant challenge to the statistical organisation to provide the suitable website design, accuracy, timeliness, and reliability of data and metadata. Metadata (especially descriptive metadata and metadata assisting in the navigation and search) are important elements for data dissemination. The metadata facilitate legibility and apprehensibility of the disseminated data and ensure the correct interpretation of presented data.

These metadata are also used for creation of database queries, more precisely in the projection (which columns from tables are included in query)

Table 4 Result of fuzzy classification

Municipality	Coefficient of needs (R)
Banská Bystrica	0.9
Banská Štiavnica	0.9
Zvolen	0.9
Detva	0.9
Donovaly	**0.845**
Lučenec	**0.75**
Cerovo	**0.68**
Fiľakovo	0.65
Rimavská Sobota	0.65
Horný Tisovník	**0.515**
Jasenie	0.35
Banský Studenec	0.35
Kremnica	0.35
Podhorie	0.35
Sliač	0.35
Skerešovo	**0.33325**
Leváre	**0.31675**
Pôtor	**0.31675**
Jelšava	**0.26675**
Vinica	**0.25**
Rapovce	**0.21675**
Bottovo	**0.16675**
Radzovce	0.1
Hostice	0.1
Nenince	0.1
Dudince	0.1

Source: own research

and the selection (which conditions have to be satisfied to extract a record from the database). In the projection phase the user chooses interesting indicators (the *SELECT* statement of a query). In the selection phase the user is limited by the properties of the crisp logic in the *WHERE* clause of a query. It means that the record either satisfies the intent of a query or do not satisfies it. This logic does not permit any other possibility. In some cases this property of the crisp logic is desirable. For example, the user wants to select all municipalities belonging to the district A. The meaning and logic of this query is two-valued (municipality fully belongs or fully does not belong to the district A).

In cases when the logic of a query cannot be limited by crisp logic, the fuzzy approach gives a solution. For example, when the user wants to find towns with good living conditions, the user can describe preferences by linguistic expressions. The output of a query is softly ranked towns according to the previously created preferences.

In [21] fuzzy classification interpreter and editor have been implemented as Java Servlets. A similar approach could be applied for fuzzy selection. The GLC was tested on desktop application. The idea of broadening this approach to websites might be very perspective and is under consideration. The interesting candidate for testing and implementing data dissemination by fuzzy queries is the population and housing censuses in Slovakia on the website [8]. The essence of fuzzy queries is reducing or eliminating the communication barrier between the human and the computer during querying process. Another reason for this development is the fact that the goal of many websites is to target broad audience. Many users of websites are not familiar with limitations of SQL and they expect data selection process to be closer to working methods of humans. Providing a query by linguistic expressions gives natural way for database queries creation and websites could become more user friendly oriented in processes of data selection.

CONCLUSION

It is proven in our research that the proposed fuzzy logic approach can improve work with statistical information systems. If crisp sets and sharp boundaries in queries are used the result may involve some inadequately selected data, e.g. in cases when the user cannot unambiguously define the criteria by crisps values. The SQL requires the crisp specification of a query criterion, while for users a query is better describable in terms of a natural (or quasi) natural language with ambiguities and uncertainties. This is one of reasons why the research has started with database querying improvements by fuzzy logic. As an output of this research, the GLC was created. In this way, queries based on linguistic expressions on client side are supported and are accessing relational databases in the same way as the SQL. No modification of databases has to be undertaken.

The goal of query based on fuzzy logic is not to select more data but to select more representative data. The fuzzy logic approach is not only more natural for users, but it is also more powerful. Data is selected according to the gradation of satisfying a query criterion. Database querying languages based on fuzzy logic demand additional calculations in comparison with SQL counterpart. This additional amount of calculation is balanced with additional information obtained from database.

The software for fuzzy selection based on the GLC has been developed on prototype level. More precisely, required items needed for case studies were realized. The stress was on research and creation of equations to describe the fuzzy logic and its potentiality.

Later researches have shown that the GLC can be used for data classification and dissemination. Fuzzy classification approach gives users the possibility to include the approximate reasoning into the classification problem by creating fuzzy *IF-THEN* rules. These fuzzy rules are converted into fuzzy queries and solved using the GLC. Data dissemination on websites is mentioned as an interesting field where queries based on the GLC could be realised.

It is important to point out that there is no competition between computing by crisp logic and computing by fuzzy logic. A fuzzy query provides flexibility when user cannot unambiguously define the criterion by crisps boundaries or user can not expressly prove why the chosen bound-

ary value is the best one. Moreover, selection of relevant entities from data sets is more flexible, allowing examination of records that clearly meet the criteria, as well as those that almost meet the given criteria. SQL database queries are useful when a clean and exact boundary between selected and non selected data is required and user is interested in data which clearly meet given criteria only. The similar statement holds for the classification. In dissemination the nature of searched data or information predetermines the use of SQL or fuzzy query. It is on the user to decide which approach is better for the particular task. Although mathematics based on fuzzy sets and fuzzy logic has greater expressive power than classical mathematics based on crisp sets and crisp logic, the usefulness depends critically on our capability to construct appropriate membership functions of linguistic expressions and create relevant fuzzy rules for each particular task.

Metadata are used to explain the meaning of the indicator and its values. The similar constatation can be told for fuzzy data selection: linguistic expressions are used to explain the meaning of a query.

References

[1] BENČIČ, A., HUDEC, M. *MOŠ/MIS–Urban and municipal statistics project and information system of the Slovak Republic.* SYM-OP-IS, XXI-32--XXI-35, 2002.

[2] BOSC, P., PIVERT, O. SQLf Query Functionality on Top of a Regular Relational Database Management System. In: Pons M, Vila M A and Kacprzyk J (eds.). *Knowledge Management in Fuzzy Databases.* Physica Publisher, Heidelberg, 2000. pp 171–190.

[3] BRANCO, A., EVSUKOFF, A., EBECKEN, N. *Generating Fuzzy Queries from Weighted Fuzzy Classifier Rules.* ICDM workshop on Computational Intelligence in Data Mining, 2005. 21–28.

[4] CHAMBERLIN, D., BOYCE, R. SEQUEL: *A Structured English Query Language.* ACM SIGMOD Workshop on Data Description, Access and Control, 1974. 249–264.

[5] FEW, S. *Show me the numbers – Design tables and graphs to enlighten.* Oakland: Analytic Press 2004.

[6] GALINDO, J., URRUTIA, A., PIATTINI, M. *Fuzzy Databases: Modeling, Design and Implementation.* Hershey: Idea Group Publishing 2006.

[7] HUDEC, M., VUJOŠEVIĆ M. *Fuzzy systems and neuro-fuzzy systems for the municipalities classification.* Eurofuse anniversary workshop on "Fuzzy for Better", 2005. 101–110.

[8] HUDEC, M., BÜCHLER, P. *Metadata and website design for statistical data dissemination. Management*, No 52, 2009. 23–30.

[9] HUDEC, M. *An Approach to Fuzzy Database Querying, Analysis and Realisation.* Computer Science and Information Systems Vol. 6, No. 2, 2009. 124–140.

[10] HUDEC, M. *Soft computing techniques for statistical databases.* Meeting on the Management of Statistical Information Systems, 2009. <http://www.unece.org/stats/documents/ece/ces/ge.50/2009/wp.22.e.pdf>.

[11] HUDEC, M., VUJOŠEVIĆ, M. *Selection and Classification of Statistical Data Using Fuzzy Logic.* NTTS Conferences on New Techniques and Technologies for Statistics, 2009. 186–195.

[12] HUDEC, M., VUJOŠEVIĆ, M. A fuzzy system for municipalities classification. *Central European Journal of Operations Research,* Vol.18, No. 2, 2010. 171–180.

[13] IOANNIDIS, C., HAZICHRISTOS, T. *A municipality selection proposal for the expansion of the Hellenic cadastre using fuzzy logic.* Spatial information management, experience and visions for the 21st century, 2000. <http://www.fig.net/com_3_athens/>.

[14] KACPRZYK, J., PASI, G., VOJTÁŠ, P., ZADROZNY, S. Fuzzy querying: Issues and perspectives. *Kybernetika*, Vol. 36, No. 6, 2000. 605–616.

[15] KLIR G., YUAN B. *Fuzzy sets and fuzzy logic, theory and applications.* Prentice Hall: New Jersey 1995.

[16] PORTINALE, L., VERRUA, A. *Exploiting Fuzzy-SQL in Case-Based.* Florida Artificial Intelligence Research Society Conference, 2001. 103–107.

[17] RASMUSSEN, D., YAGER, R.R. Summary SQL – A Fuzzy Tool for Data Mining. *Intelligent Data Analysis,* 1, 1997. pp. 49–58.

[18] SILER, W., BUCKLEY, J. *Fuzzy expert sytems and fuzzy reasoning.* New Jersey: John Wiley & Sons, 2005.

[19] *United Nations Statistical Commision and Ecconomic Commision for Europe. Best practices in designing websites for dissemination of statistics.* Conference of European statisticians, Methodological material, Geneva, 2001.

[20] URRUTIA, A., PAVESI, L. *Extending the capabilities of database queries using fuzzy logic.* Collecter-LatAm, 2004. <http://www.collecter.org/archives/2004_October/06.pdf>.

[21] WERRO, N., MEIER, A., MEZGER, C., Schindler, G. *Concept and Implementation of a Fuzzy Classification Query Language.* International Conference on Data Mining, 2005. 208–214.

[22] WANG, T.C., LEE, H.D., CHEN, C.M. *Intelligent Queries based on Fuzzy Set Theory and SQL.* Joint Conference on Information Science, Salt Lake City, 2007. 1426–1432.

[23] ZADEH, L. Fuzzy Sets. *Information and Control,* No. 8, 1965. 338–353.

[24] ZIMMERMANN, H-J. *Fuzzy Set Theory: And Its Applications.* Kluwer Academic Publishers: London, 2001.

Identification of Influential Points in a Linear Regression Model

Jan Grosz[a] | *Czech University of Life Sciences in Prague*

Abstract

The article deals with the detection and identification of influential points in the linear regression model. Three methods of detection of outliers and leverage points are described. These procedures can also be used for one-sample (independent) datasets. This paper briefly describes theoretical aspects of several robust methods as well. Robust statistics is a powerful tool to increase the reliability and accuracy of statistical modelling and data analysis. A simulation model of the simple linear regression is presented.

Keywords

leverage points, outliers, method of least squares – LSM, trimmed mean, confidence interval, regression model, data simulation, robust estimators

INTRODUCTION

Regression analysis, along with variance analysis, belongs to such mathematic-statistical methods, which can find a broadest usage in practical applications of various sciences. The main goal of regression analysis is finding of a real function f, which describes the relation of the dependent variable Y and a group of independent variables $X_1, X_2, \ldots, X_m$. This function is called the regression function and shall comply with the relation as follows:

$$Y = f(X_1, X_2, \ldots, X_m) + \varepsilon,$$

where ε is the random variable representing random deviations (errors) of the model.

Let us further limit to the linear class of functions, that is to deal with the model as follows:

$$Y = \beta_1 + \beta_2 X_2 + \ldots + \beta_m X_m + \varepsilon .$$

Parameters β_i are called linear regression coefficients and this paper is devoted to their estimators.

It is known that estimators of regression coefficients by means of the classical method of least squares are very sensitive to extreme points that means to the points, which are "standing out of the line" in a certain way. In practice, such data "is created" most frequently by an error when data is entered into the computer, or by potentially erratic filling in of the original source data. Therefore, it is of great importance to identify such points and eliminate them from the dataset because their presence – and there may be the only one such point – would substantially distort or even completely deteriorate the resulting values of regression analysis parameters. Such values are referred to as influential points (observances) and for the sake of simplicity are classified as:

- extreme points, called outliers (type E), occurring at the dependent variable, see Figure 3; and
- outlying leverage points (type V) occurring at the independent variables, see Figure 2.

[a] *Czech University of Life Sciences in Prague, Kamýcká 129, 165 21 Praha 6-Suchdol, e-mail: grosz@pef.czu.cz*

1 GENERIC MODEL OF LINEAR REGRESSION

Let us take the classical model of linear regression:

$$y_i = \sum_{j=1}^{m} x_{ij}\beta_j + \varepsilon_i, \quad i=1,\ldots,n, \quad n > m, \qquad (1)$$

where x_{ij} are given values of ith repetition of jth explanatory (independent) variable, εi are independent, random variables of normal distribution with zero mean value and variance σ^2 (so-called "white noise"), β_j means unknown regression coefficients, and y_i is a value of the regressand (or the dependent variable) at ith observation.

The matrix record is in the form

$$Y = X\beta + \varepsilon \qquad (2)$$

where $X = (x_{ij})$ is a matrix of order nxm and $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)'$, $\beta = (\beta_1, \ldots, \beta_m)'$, and $Y = (y_1, \ldots, y_n)'$ are column vectors.

Therefore, Y is a random vector, which has normal distribution with the mean value (vector) of $X\beta$ and the variance-covariance matrix σ^2. I_n, where I_n is a unit matrix of order n.

The basic goal of regression analysis is to estimate the vector β by minimising the sum of squares of observed points deviations from the regression line. In mathematic language it is finding of the minimum of the quadratic form of $S(\beta) = (Y - X\beta)'(Y - X\beta)$.

Therefore we seek:

$$\min (Y - X\beta)'(Y - X\beta). \qquad (3)$$

Let us say $\hat{\beta}$ is any solution of a linear equations system:

$$X'X\beta = X'Y. \qquad (4)$$

The system of (so-called normal) equations (4), which is yielded when solving the task (3) has always one solution, at least, because $L(X) = L(X'X)$. Here $L(X)$ refers to the linear envelope formed of columns of the matrix X – see [7], for instance.

In general, the linear envelope of a finite set of elements (vectors) of a vector space is defined as a set of all linear combinations of these vectors.

It holds:

$$(Y - X\beta)'(Y - X\beta) \geq (Y - X\hat{\beta})'(Y - X\hat{\beta})$$

In other words the quadratic form $S(\beta)$ takes its minimum in the point $\beta = \hat{\beta}$.

Here $S(\hat{\beta})$ represents the residual sum of squares of deviations observed from fitted values.

It is easy to show that the following relations are valid:

$$S(\hat{\beta}) = Y'Y - Y'X\hat{\beta} \qquad (5)$$

$$E(S(\hat{\beta})) = (n - h(X)).\sigma^2 \text{ and } D(Y - X\hat{\beta}) = D(Y) - D(X\hat{\beta}) \qquad (6)$$

The quantity $S(\hat{\beta})/(n - h(X))$ is therefore an unbiased estimator of the parameter σ^2. The symbol of $h(X)$ means the rank of the matrix X, $E(X)$ is a mean value of the random variable X, and $D(Y)$ is a variance-covariance matrix of the vector Y. Proof can be found in the publication [7] as well. Rather detailed publications dealing with matrix algebra are [6] and [7]. The next section mostly deals with the case m = 2 – that is the most frequently occurring issue of simple linear regression in practice.

2 IDENTIFICATION OF EXTREME POINTS

There are numerous methods, which can identify extreme points. Procedures given here are good to interpret and appropriate characteristics can be easily calculated within the environment of the spreadsheet software Excel – therefore they do not require any special statistical software. In author's experience they are highly effective and sensitive in discovering extreme points of input datasets.

2.1 Identification of leverage points

Let us assume hereinafter that the model (1) is a full rank model, that is $h(X) = m$ is valid.

In this case the solution of the system (4) is determined unambiguously and has the form:

$$\hat{\beta} = (X'X)^{-1}X'Y. \qquad (7)$$

It holds that $E\,\hat{\beta} = \beta$, and so the estimator is unbiased and has the least variance among such es-

timators. In such case we can call it the best linear unbiased estimator of the vector β.

Now, let us mark $\hat{Y} = X\hat{\beta}$ the "predicted" vector Y. If $\hat{\beta}$ in the equation is replaced with the expression (7) the yield is:

$$X\hat{\beta} = X(X'X)^{-1}X'Y = WY, \quad W = X(X'X)^{-1}X' \tag{8}$$

The matrix W is a square matrix of rank n having properties as follows:

(i) $W' = W$ (symmetry)

(ii) $W_2 = W$ (idempotency)

(iii) $W'X = X$

(iv) W is a hat matrix to $L(X)$

(v) $0 \le w_{ii} \le 1,\ i = 1, \ldots, n$

(vi) $\sum_{i=1}^{n} w_{ii} = \mathrm{m}$

(vii) Let us mark $\hat{Y} = (\hat{y}_1, ..., \hat{y}_n)$ a $\hat{\boldsymbol{e}} = Y - \hat{Y} = (\hat{e}_1, \ldots, \hat{e}_n)$ – the vector of residuals. Then $var\,(\hat{y}_i) = w_{ii}\sigma^2$ and $var\,(\hat{e}_i) = (1 - w_{ii})\sigma^2$

(viii) $\hat{y}_i = y_i w_{ii} + \sum_{j \ne i} w_{ij} y_j$.

(ix) Diagonal elements of w_{ii} of the hat matrix W represent – roughly – the distance of ith observation from the middle of other points concerning explanatory variables.

(x) Such point x_i can be considered an extreme point, for which:

$$w_{ii} > \frac{2m}{n},\ i = 1, \ldots,\ n. \tag{9}$$

The procedure as follows can be used to explain this boundary. Let us assume that row vectors of the matrix X form multivariate normal distribution. Then, testing the hypothesis that all rows have the mean value constant, the testing statistics

$$F = \frac{n-m}{m-1}\frac{w_{ii} - \frac{1}{n}}{1 - w_{ii}}$$ has Fischer's distribution

with $m - 1$ and $n - m$ degrees of freedom. If critical value of this statistics is roughly equal to 2, then $F > 2$ (which is the critical region to reject the hypothesis) when the relation (9) is approximately valid. Details can be found in [8] or [9].

2.2 Identification of extreme values – outliers

First, let us introduce the term of so-called trimmed mean α $(0 < \alpha < 0.25)$. This is an arithmetic average, which remains after $100^*\alpha$ % of the smallest and largest values are eliminated.

That means more precisely:
let us mark $y_{(1)} \le y_{(2)} \le \ldots\ldots \le y_{(n)}$ ordered original values of the dependent variable, $n_1 = \{\alpha^* n\}$, $n_2 = n - n_1$, (symbol $\{x\}$ means the nearest natural number higher or equal to x). Thus in total n_1 of the least and largest values are eliminated and the sample then contains $k = n_2 - n_1$ values. Then let us define α – the trimmed mean and α – trimmed variance as follows:

$$\bar{y}_\alpha = \frac{1}{k}\sum_{i=n_{1+1}}^{n_2} y_{(i)} \tag{10}$$

$$\sigma_\alpha{}^2 = \frac{1}{(k-1)}\sum_{i=n_{1+1}}^{n_2} (y_{(i)} - \bar{y}_a)^2. \tag{11}$$

In practice it is selected to be $0.05 \le \alpha \le 0.1$, ie. ca 10% – 20% of sample values are eliminated and the aforementioned characteristics of the mean value and variance are calculated using the rest of the sample values.

Further procedure is based on the modification to the known three sigma rule, which holds for normal distribution. Let us choose the confidence interval

$$(\bar{y}_a - 3\sigma_a,\ \bar{y}_a + 3\sigma_a)$$

and detect such points y_i, lying outside this interval, i.e. such y_i, which meet $| y_i - \bar{y}_\alpha | > 3\sigma_\alpha$, $i=1, \ldots, n$.

This way identified points can be considered extreme ones. These values need to be subject to further assessment. Verification, if these are erratic data (which is quite common case in data entering), or these are really extreme values, has to be carried out. In the first case the points are corrected, of course, in the second case such values may be either eliminated from the dataset and then to calculate the vector β estimator using the common method of least squares; or we can chose some other method. This provides a certain guarantee that the dataset got cleaned of "suspicious values".

2.3 Identification of influential points

Ali S Hadi (1992) proposed the following (additive) statistics, which tests influential points in the model of linear regression as follows:

$$H_i = \frac{w_{ii}}{1 - w_{ii}} + \frac{m}{1 - w_{ii}} \frac{d_i^2}{1 - d_i^2}, \text{ where}$$

$$d_i = \hat{e}_i / \sqrt{S(\hat{\beta})} \qquad (12)$$

is so-called normalized residual. $i = 1, 2, \ldots, n$.

The first summand in (12) represents a portion of influence of the explanatory variable, the second addend then represents influence of the dependent variable. The test therefore consists in the fact an influential point is either of E or V type, respectively. High values of H_i prove that ith observance represents an influential point while there is no exact limit determined in this case. Recommendation is to set preliminary critical value to 1.

3 DATA SIMULATIONS – ILLUSTRATIVE EXAMPLES

For the purpose of quality verification of the aforementioned methods a simulation experiment was carried out by means of a random number generator for the model of simple linear regression with parameters as follows:

$n = 30$ (sample size), $m = 2$ (number of parameters), $\beta = (3,7)'$, $\sigma^2 = 4$

Therefore the model (1) has the shape:

$$y_i = 3x_i + 7 + \varepsilon_i, \quad i = 1, \ldots, 30. \qquad (13)$$

The explanatory variable xi was generated from uniform distribution $R(20, 30)$ and ε_i has normal distribution with zero mean value and standard deviation 2.

Table 1 Generated data (13)

I	y_i	x_i	i	y_i	x_i
1	77.03	23.20	16	81.18	25.03
2	92.16	28.48	17	74.86	22.85
3	86.23	26.00	18	74.17	22.29
4	84.13	25.26	19	82.46	24.15
5	78.98	22.86	20	95.61	29.21
6	89.43	27.20	21	70.16	22.33
7	84.13	26.34	22	70.41	20.83
8	71.25	21.41	23	75.99	23.28
9	70.44	21.58	24	83.18	25.22
10	89.07	26.78	25	67.65	20.35
11	78.88	24.29	26	89.90	28.08
12	71.55	21.17	27	77.91	23.84
13	88.01	25.98	28	69.66	20.45
14	75.70	22.52	29	77.06	24.11
15	90.54	27.71	30	95.50	29.63

Source: own research

The estimates of the regression coefficient parameter β and standard deviation σ of the model (13) obtained by the method of least squares were $\hat{\beta} = (3.017; 6.769)'$ and $\hat{\sigma} = 2.6$.

Data (13) was subsequently "contaminated" with influential values of E and V types this way:

V: $x_1' = 40$ and $x_3' = 4$ (original values were $x_1 = 23.2$ and $x_3 = 26$).

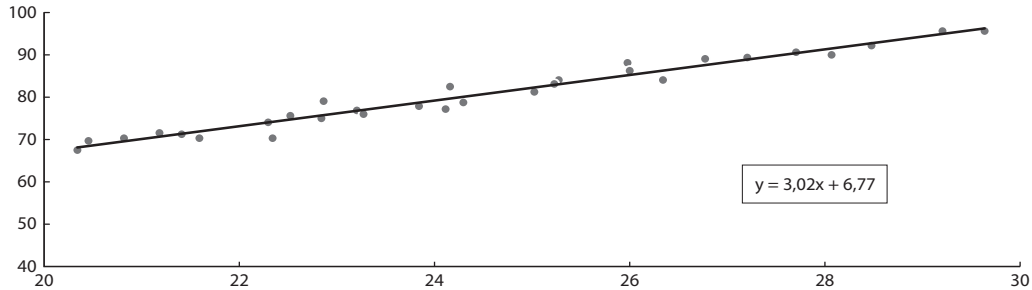
Diagonal elements of the matrix W are used for the detection of extreme values, as derived above; if

$$w_{ii} > \frac{2m}{n},$$

then such a point may be considered extreme value. In our case this critical value is 0.13, while $w_{11} = 0.31$ and $w_{33} = 0.51$, and for the other $w_{ii} < 0.07$, so x_1', x_3' can be considered extreme values.

In the calculation of the hat matrix W in the Excel environment functions are used as follows:

Figure 1 Linear regression of a dataset (13)



Source: own research

TRANSPOSITION *(A)* – carries out transposition of the given matrix *A ´*;

MATRIX.PRODUCT*(A; B)* – result is a product of matrixes AB; and

INVERSION *(A)* – calculates the inversion matrix A^{-1} (if there is any).

The inversion matrix calculation is of sufficient accuracy; nevertheless this function has its limitations (especially for matrixes of higher orders, for instance with n > 50).

Table 2 Values of diagonal elements of the hat matrix *W*

I	w_{ii}	i	w_{ii}
1	0.321	16	0.034
2	0.054	17	0.036
3	0.508	18	0.038
4	0.035	19	0.033
5	0.036	20	0.062
6	0.044	21	0.038
7	0.038	22	0.047
8	0.043	23	0.034
9	0.042	24	0.034
10	0.041	25	0.051
11	0.033	26	0.050
12	0.044	27	0.034
13	0.037	28	0.050
14	0.037	29	0.033
15	0.047	30	0.067

Source: own research

Further two values were replaced with extreme points this way:

E: $y_1' = 50$ and $y_2' = 140$

(original values were $y_1 = 77.02$ and $y_2 = 92.2$).

Subsequently, α – trimmed mean and variance for $\alpha = 0.05$ were calculated:

$n_1 = \{30*0.05\} = 2,$

$n_2 = 30 - 2 = 28,$

$k = 26,$

$\bar{y}_\alpha = 80.03, \sqrt{\sigma_\alpha{}^2} = 7.39$.

There were 4 points eliminated and the confidence interval was calculated in the form:

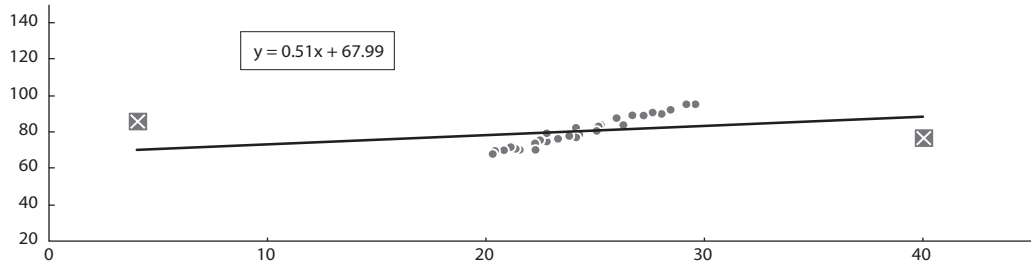
$$(57.8;\ 102.2) \qquad (14)$$

There are solely points y_1' and y_2' out of the interval (14) (see Figure 3 below).

What can also be seen in Figures 2 and 3 is the presence of outliers lead to significantly worse results than the presence of leverage points. Both the methods described can be employed for the detection of extreme (erratic) values for one-sample datasets.

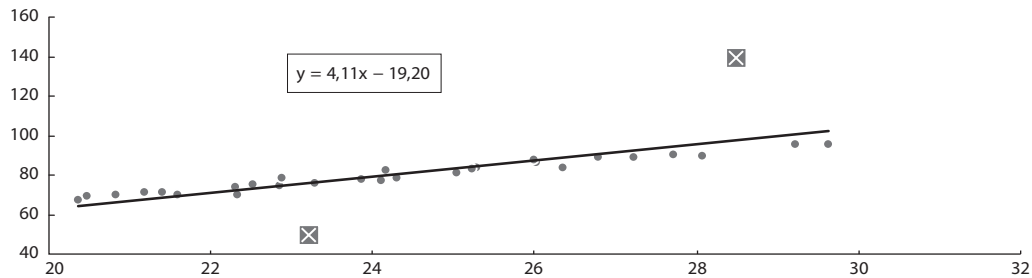
In order to verify the Hadi measure the original dataset was "contaminated" with extreme values of E and V types simultaneously: $x_1' = 40$ and $x_3' = 4$ and $y_1' = 50$ and $y_2' = 140$ (the way the first three pairs of the original values were replaced (13)). It is clear in Figure below how this modification deteriorated "proper" parameters of the regression function.

Figure 2 Linear regression of a dataset containing two extreme points



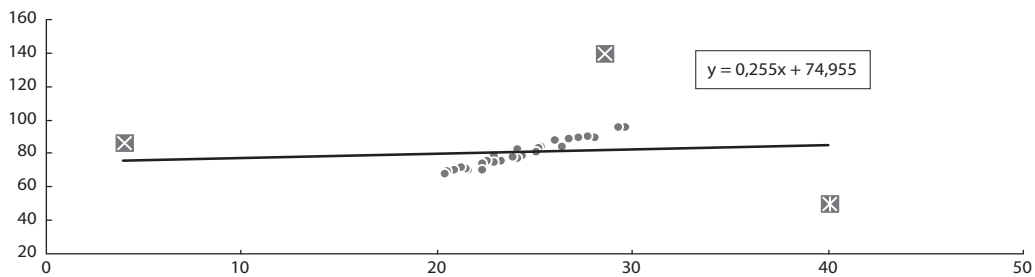
Source: own research

Figure 3 Linear regression of a dataset including two influential points



Source: own research

Figure 4 Linear regression of a dataset containing three influential points



Source: own research

It can be seen from Table 3 that the Hadi measure attains the highest value for H_3 and H_1, and further then for H_2, which indicates the presence of influential points. Other H_i are by an order of magnitude lower.

4 ROBUST METHODS

Methods, which reduce sensitivity to extreme values and simultaneously give high quality regression coefficient estimators, are called robust methods. A whole number of such methods were proposed due to fast progress in computer technology. These methods are theoretically described in a very detailed manner in the today already classical monograph [2], and newly can also be found in [5].

The most often applied procedure in the estimating of regression coefficients is M-estimators (maximum likelihood). It is such an estimator of $\hat{\beta}$, which minimises the sum of residuals using a suitable way chosen function ρ, which is convex, and there is a derivation ρ´. That means it is a certain generalisation of the method of least squares where

Table 3
The Hadi measure values

I	H_i	i	H_i
1	2.837	16	0.035
2	0.314	17	0.038
3	4.072	18	0.041
4	0.036	19	0.037
5	0.042	20	0.067
6	0.046	21	0.039
7	0.041	22	0.053
8	0.047	23	0.036
9	0.044	24	0.036
10	0.043	25	0.057
11	0.035	26	0.054
12	0.050	27	0.035
13	0.039	28	0.058
14	0.040	29	0.035
15	0.050	30	0.074

Source: own research

$\rho(x) = x^2$. The M-estimator therefore depends on the selection of the function ρ. Its drawback is the M-estimator eliminates solely effects of outliers and not those of leverage points.

Other applied estimator is the LTS-estimator (least trimmed squares estimator). This estimator is calculated by omitting a certain number of the smallest and largest residuals (similar way as in 3.2).

The LMS-estimators or LMedS-estimators (least median of squares estimators) are based on the idea of minimising the median of squared residuals. Generalisation of LMS and LTS estimators give birth to the S-estimator.

Statistical software SAS ver. 9.2 has, in its routine ROBUSTREG, four methods of estimators, including testing for the presence of outliers and leverage points: M-, LTS-, S-, and MM-estimators. Yet the identification of outliers is based on other methods than those mentioned here above.

CONCLUSION

In real applications one can often face the issue of identification and detection of extreme (and/or leverage) points, which are such points that in principal manner affect the dataset analysis. Such points are classified as outliers of values of the dependent variable, leverage points of the independent variable, or influential points of both the variables. It is right the presence of such points that results in often completely worthless regression parameters estimators using the method of least squares. Therefore the type of the analysed data contamination must be identified first. Three methods were chosen out of a number of existing methods as follows: detection by means of a projection matrix, "robust" confidence interval, and the Hadi measure. In author's experience these methods have worked very well in practice, namely in accuracy checking of PC entered data.

Some of the methods of the regression coefficient calculation by means of so-called robust methods are briefly described in section 5. These methods are implemented in the SAS system.

Remark: All necessary numeric calculations were carried out in the spreadsheet software EXCEL.

References

[1] BARNETT V., LEWIS T. *Outliers in Statistical data.* Wiley: New York, 1994.
[2] HUBER, P. J. *Robust Statistics.* John Wiley: New York, 1981.
[3] CHAJDIAK, J. *Štatistika v Exceli.* Statis: Bratislava, 2002.
[4] CHATTERJEE, S., HADI, A.S. *Regression analysis by example.* Wiley-Interscience: Hoboken, 2006.
[5] MARONNA R., MARTIN D.,YOHAI V. *Robust Statistics: Theory and Methods.* Wiley, 2006.
[6] NERING, D.E. *Linear algebra and matrix Tudory.* Wiley: New York, 1970.
[7] RAO, C. R. *Linear Statistical Inference and Its Applications.* Wiley: New York, 1965.
[8] RYAN, T.P. *Modern regression methods.* Wiley: Hoboken, 2009.
[9] ZVÁRA, K. *Regresní analýza.* Academia: Praha, 1989.

Papers
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 to12 000 words or up to 20 1.5-spaced pages.
The *Methodologies* 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 20 1.5-spaced pages.
The *Book Reviews* section brings reviews of recent books in the field of the official statistics. Reviews shall have up to 600 words or up to 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) – Introduction – 1 Literature Survey – 2 Methods – 3 Results – 4 Discussion – Conclusion – Annex – Acknowledgments – References – Tables and Figures

Authors and Contacts
Rudolf Novak[a,*], Jonathan Davis[b]
[a] Czech Statistical Office, Na padesátém 81, 100 82 Praha 10, Czech Republic
[b] Eurostat, European Commission, Rue Alcide de Gasperi, LU-2920 Luxembourg, Luxembourg
[*] Corresponding author. E-mail: rudolf.novak@domainname.cz, Phone: +555 555 555 555

Main Text Format
Arial 10 (main text), 1.5 spacing between lines. Page numbers in the lower right-hand corner. Underline text if necessary. Do not use **bold** or *italics*. Paper parts numbering: 1, 1.1, 1.2, etc.

Headings
1 FIRST-LEVEL HEADING (ARIAL 12, bold)
1.1 Second-level heading (Arial 10, bold)
1.1.1 Third-level heading (Arial 10, bold italic)

Footnotes
Footnotes should be used sparingly. Do not use endnotes. Do not use footnotes for citing references.

References in the Text
Place reference in the text enclosing authors' names and the year of the reference, e.g. "Novak (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. et WEYANT, J., eds. *Discounting and Intergenerational Equity.* Washington, D.C.: Resources for the Future, 1999.
[for a journal] COASE, R.H. The Problem of Social Cost. *Journal of Law and Economics.* 1960, 3 (October): 1–44.
[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 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.

Figures
Figure is any graphical object other than table. Attach each figure as a separate file. Indicate position of the figure by placing 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).

Pie charts, bar charts, trend lines, etc. 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.

Paper Submission
Please email your papers in *.doc and *.docx 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
Address:
Journal of Statistika
Czech Statistical Office
Na padesátém 81, 100 82 Praha 10
Czech Republic

E-mail: statistika.journal@czso.cz
Web: www.czso.cz/statistika_journal

Journal of Statistika
Czech Statistical Office
Na padesátém 81
100 82 Praha 10
Czech Republic

statistika.journal@czso.cz
www.czso.cz/statistika_journal