STATISTIKA

STATISTICS AND ECONOMY JOURNAL

VOL. **98** (2) 2018

CZECH STATISTICAL OFFICE

CONTENTS

ANALYSES

120th ANNIVERSARY

INFORMATION

About Statistika

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

Publisher

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

Contact us

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

Cyclical-Adjusted External Balance of Goods and Services in the Czech Republic

Zdeněk Pikhart [1] | *University of Economics, Prague, Czech Republic*

Abstract

Article develop the cyclical adjustment methodology of the balance of goods and services, which is beneficial especially for a small open economy. Cyclical adjustment procedure is applied and adjusted to the Czech data by quantifying cyclical determinants of foreign trade such as domestic and foreign output gap and trade prices deviations from the trend. Results show that cyclical adjustment of external trade balance provide a useful tool for assessing the sustainability of trade development and thus help to better analyze current situation and forecast short-term future development.

Keywords	***JEL code***
Business cycle, cyclical adjustment, external balance of trade	*E32, F32, F44*

INTRODUCTION

The turnover and balance of foreign trade represents significant aggregates for the assessment of the internal and external balance of the economy. The balance of goods and services enters the use of the gross domestic product, but is also one of the main items on the current account of the balance of payments. Its development is thus closely linked to the overall macroeconomic situation of the country. For the purposes of macroeconomic analyses and forecasts, it is useful to adjust the balance of foreign trade for the business cycle, that is, to break down the balance of goods and services to the structural and cyclical components.[2]

The openness of the economy, captured by the ratio of foreign trade turnovers to GDP, is determined by the size and geographical position of the economy, barriers to trade, infrastructure, education, cultural proximity of states, the institutional framework, political risks and other factors. Higher international division of labour usually leads to narrower specialization and deeper use of comparative advantages as one of the factors of labour productivity growth.

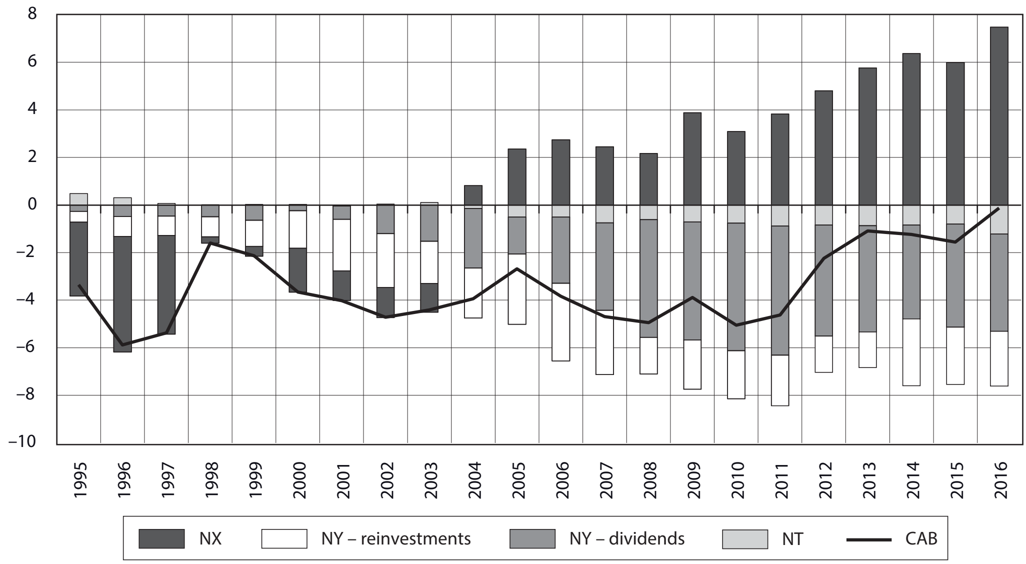
However, if we focus on the balance of foreign trade, its structural component is related to the intertemporal position of the current account of the balance of payments in the broader economic

[1] University of Economics, Prague, Faculty of Finance and Accounting, W. Churchill Sq. 1938/4, 130 67 Prague 3, Czech Republic. E-mail: xpikz00@vse.cz. Also the Ministry of Finance of the Czech Republic, Letenská 15, 118 10 Prague 1, Czech Republic. E-mail: zdenek.pikhart@mfcr.cz.

[2] This is a significantly extended version of the text published in the Macroeconomic Forecast (Ministry of Finance, Box 3.1, 2018).

convergence process. A young, transforming economy strongly attracts foreign capital, and the resulting excess of investment over savings leads to an increase in imports of goods and services and in turn, to a negative balance of foreign trade. An increasing production of foreign companies in the economy gradually improves the balance of goods and services, and the deficit of primary income with a predominant share of reinvested earnings is simultaneously increasing. In the later stages of the investment cycle, the balance of foreign trade is characterized by significant surpluses, which are to a large extent offset by the outflow of dividends abroad. The Czech economy has gone through these stages in the past and, at present, the positive balance of foreign trade even exceeds the negative balance of primary income, as a result of which the Czech Republic is approaching countries with net exports of savings (see Figure 1).[3]

Figure 1 Current account balance – national accounts (% of GDP)



Note: NX = net exports of goods and services, NY = net primary income, NT = net secondary income, CAB = current account balance.
Source: Czech Statistical Office

The identification of structural and cyclical determinants of the external balance of trade is definitely relevant for policy-making and has indeed attracted the academic interest, with a number of theoretical models flourishing in the literature after the pioneering works by Sachs (1981) and Buiter (1981), later extended by the classic inter-temporal approach of Obstfeld and Rogoff (1995). Several empirical applications of these models have drawn on the national accounting identity between the current account balance and the difference between national saving and investment and have suggested a variety of fundamental determinants of current account positions (Faruqee and Debelle, 1996; Blanchard and Giavazzi, 2002; Chinn and Prasad, 2003; Gruber and Kamin, 2007; Ca'Zorzi et al., 2009).

First currently used method how to distinguish between structural and cyclical factors proceeds from quantifying structural drivers of foreign trade as demography, openess, barriers, convergency and leaves the cyclical determinants as a residual (Cheung et al., 2010). Similarly, the IMF, within its External Balance Assessment procedure, estimates the structural component of the current account

[3] According to the national accounts statistics; the balance of payments statistics shows that the Czech Republic has been a net exporter of savings already since 2014.

on the basis of a reduced-form panel regression that relates the observed balance to a set of variables including both structural and cyclical factors, summarized by the domestic output gap relative to the world output gap (Phillips et al., 2013).

Second group of methods approachs the distribution of trade balance on the contrary by quantifying cyclical deviations from equilibrium balance of goods and services through trade elasticities (Wu, 2008; Kara and Sarikaya, 2013; Haltmaier, 2011 and 2014), often using effects of foreign and domestic output gaps and real exchange rate deviations. But there are many country specific factors which have to be taken into the account such as different import intensities of gross domestic product components, relevant foreign trading partners, assymetric effects of price level deviation, equilibrium real exchange rate appreciations in economic convergence process and, last but not least, unique estimate of trade elasticities. Fabiani et al. (2016) documented different sizes of trade elasticities for multiple economies carried out by regression estimates. Thus, this article applies and develops the approach of substracting cyclical components which best fits to the data of the Czech Republic.

1 DATA AND METHODOLOGY

In addition to the aforementioned structural foreign trade factors, the external position of the economy is significantly influenced by cyclical fluctuations of domestic and foreign economies. These are related to deviations of the koruna price levels (P) of traded goods and services resulting from the deviations of prices on foreign markets and of the nominal exchange rate. Price deflators of exports and imports of goods and services from the national accounts are used. Functional relationships within the nominal balance of goods and services (NX) are captured by Formula (1). Exports of goods and services (EX) are the function of foreign demand at constant prices of 2010[4] (Y_F), and imports of goods and services (IM) depend on the development of import-intensive exports and domestic demand at constant prices of 2010 (Y_{DOM}).

$$NX = \left[EX \rightarrow f\left(Y_F\right)\right] - \left[IM \rightarrow f\left(EX, Y_{DOM}\right)\right] \tag{1}$$

The structural component of exports (EX*) and imports (IM*) is quantified in Formulas (2) and (3) as actual exports and imports adjusted for the percentage deviation[5] of the variables from their equilibrium levels while taking into account the specific elasticities in foreign trade. The cyclically-adjusted balance of foreign trade (NX*) is then the difference in structural values of exports and imports (4). Seasonally adjusted data from quarterly national accounts (Czech Statistical Office, 2018) have been used:

$$EX^* = EX\left[1 - \varepsilon^{EX,Y_F}\left(Y_F - Y_F^*\right) - \varepsilon^{EX,P_{EX}}\left(P_{EX} - P_{EX}^*\right)\right] \tag{2}$$

$$IM^* = IM\left[1 - \varepsilon^{IM,EX}\left(EX - EX^*\right) - \varepsilon^{IM,Y_{DOM}}\left(Y_{DOM} - Y_{DOM}^*\right) - \varepsilon^{IM,P_{IM}}\left(P_{IM} - P_{IM}^*\right)\right] \tag{3}$$

$$NX^* = EX^* - IM^* \tag{4}$$

Figures 2–5 show the development of domestic and foreign demand and foreign trade prices with their deviations from trend values. Statistical approach by using of the Hodrick-Prescott filter with recommended λ = 1 600 for quarterly data was applied to seasonally adjusted time series from national

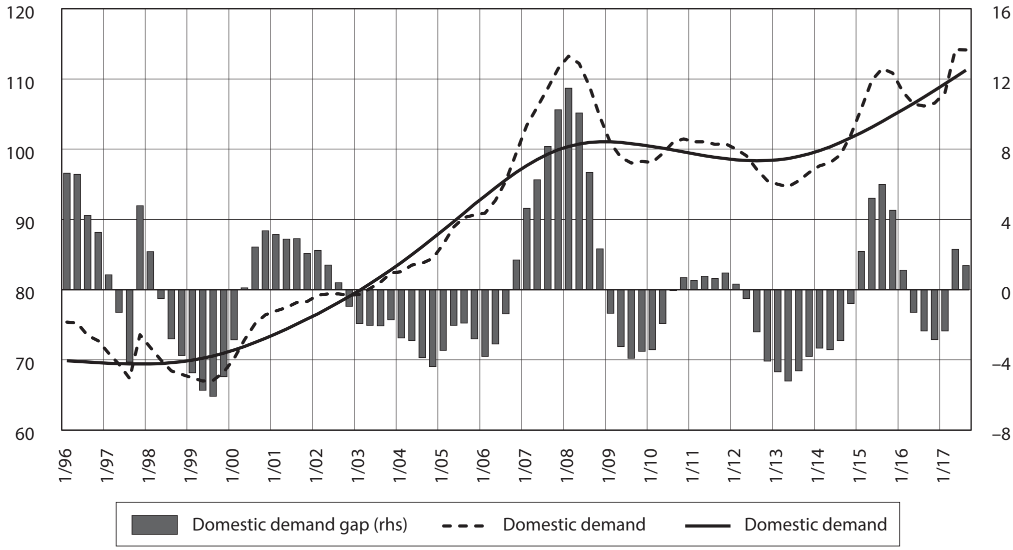
---

4 The real gross domestic product of the Euro Area was used, constant prices of 2010.

5 $\left(Y_F - Y_F^*\right); \left(P_{EX} - P_{EX}^*\right); \left(EX - EX^*\right); \left(Y_{DOM} - Y_{DOM}^*\right); \left(P_{IM} - P_{IM}^*\right)$ are interpreted as differences of logarithms, i.e., approximately a percentage deviation from trend values.

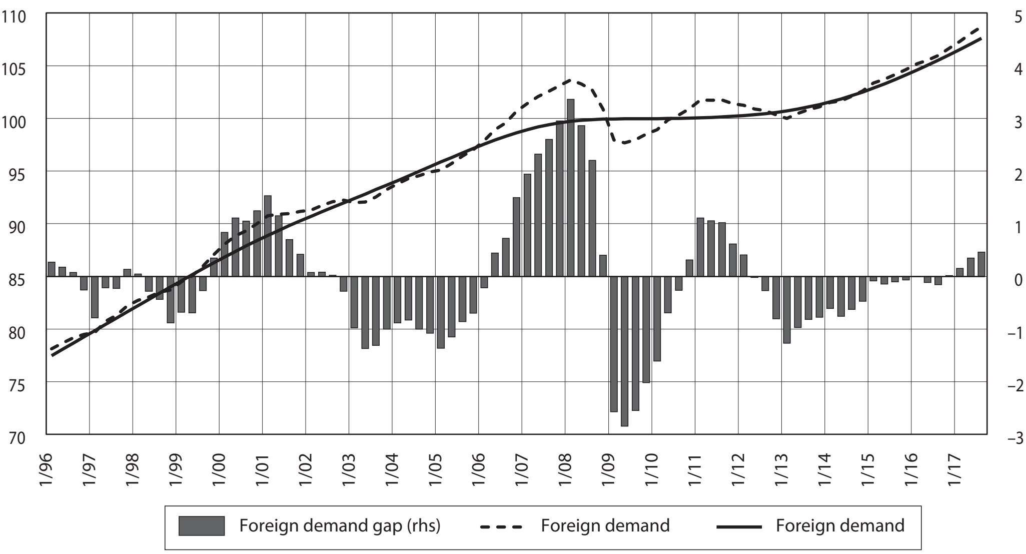
accounts (Hodrick and Prescott, 1997). Filtering was performed on quarterly data from 1996 to 2021, adding values from the current Macroeconomic Forecast of the Ministry of Finance (2018) to address the endpoint problem. Domestic demand is represented by import intensive real gross fixed capital formation, since final consumption expenditure of households and government sector remain statistically insignificant in variety of models.

Figure 2 Domestic demand gap (Ø2010 = 100, in %)



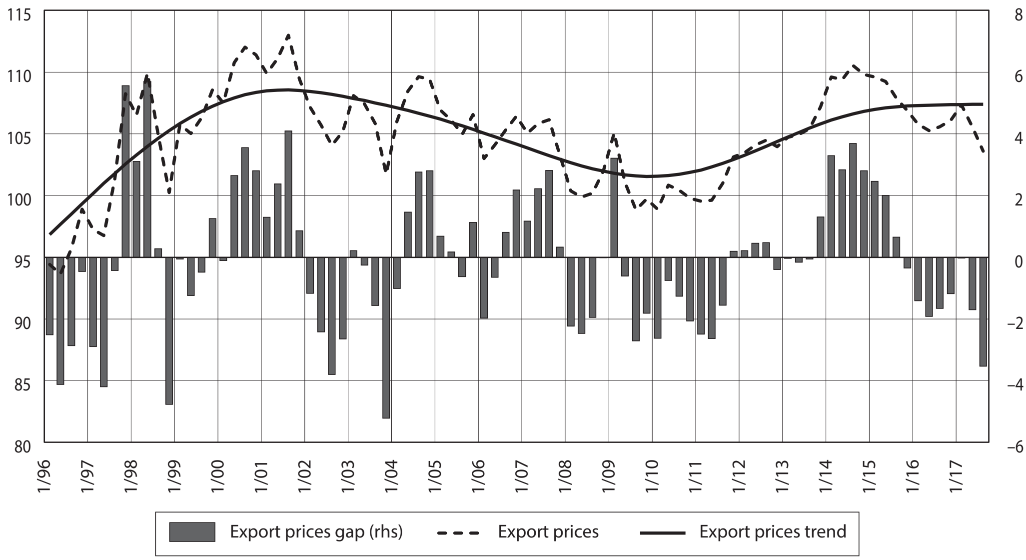
Source: Czech Statistical Office, own construction, (rhs = right hand side)

Figure 3 Foreign demand gap (Ø2010 = 100, in %)



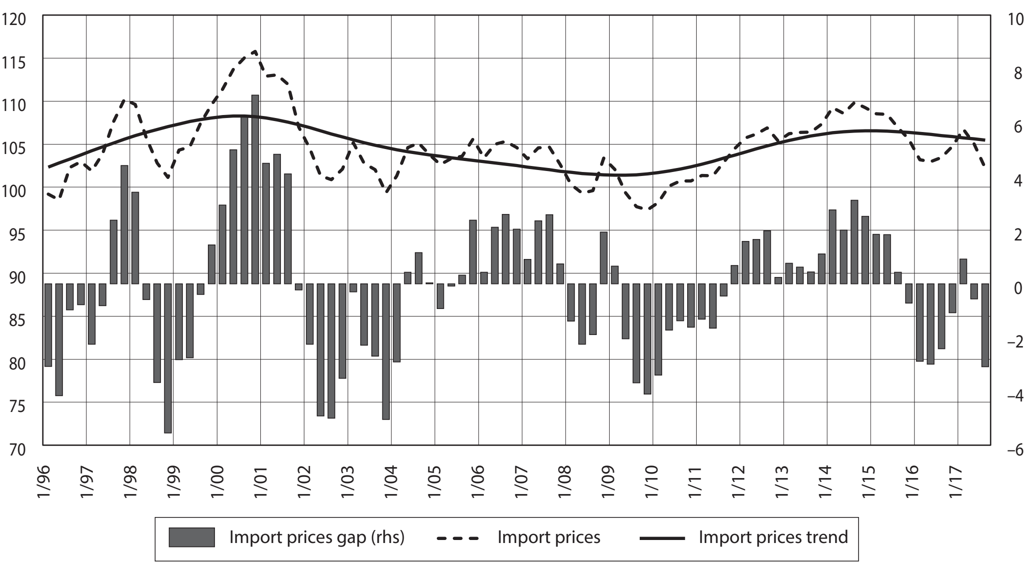
Source: Eurostat, own construction, (rhs = right hand side)

Figure 4 Export prices gap (Ø2010 = 100, in %)



Source: Czech Statistical Office, own construction, (rhs = right hand side)

Figure 5 Import prices gap (Ø2010 = 100, in %)



Source: Czech Statistical Office, own construction, (rhs = right hand side)

Table 1 shows description statistics for gap variables used in estimates. Quarterly data from the first quarter 1996 to the third quarter 2017 are used. All time series are stationary according to augmented Dickey-Fuller statistic.

Table 1 Data description

Variable	abr.	mean	st. dev.	Jarque-Bera	ADF
nominal export gap	ex_gap	0.13	6.49	5.42*	–4.64***
nominal import gap	im_gap	0.07	6.32	3.46	–4.24***
euro area output gap	ea_og	–0.03	1.17	6.69**	–4.34***
real gross fixed capital formation gap (constant prices of 2010)	gfcf_gap	–0.04	3.93	10.63***	–4.32***
export price deflator gap	pex_gap	–0.07	2.29	0.90	–4.54***
import price deflator gap	pim_gap	–0.06	2.68	0.88	–4.59***

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

In order to estimate volume and price elasticities in foreign trade regression estimates have been carried out by the method of least squares. All coefficients are statistically significant at 1% level of significance. Residual autocorrelation and heteroscedasticity is not present in the models. Results show very high dependence of exports deviations from the trend on foreign output gap. Relatively strong is also price elasticity of exports gap. Both effects affecting exports are then indirectly reflected on the imports side, since export gap is the strongest factor of the imports deviation. The import intensity of total gross fixed capital formation at 23% is significantly below the value from the input-output tables,[6] but might be influenced by EU investment cycle, which is closely related to government and construction investment with reasonably lower import intensity.

Table 2 Estimates

Endogenous variable:	ex_gap		Endogenous variable:	im_gap	
ea_og	3.52		ex_gap	0.72	
	[0.77]	***		[0.05]	***
pex_gap	1.00		gfcf_gap	0.23	
	[0.16]	***		[0.08]	***
ex_gap(–1)	0.79		pim_gap	0.37	
	[0.06]	***		[0.14]	***
			im_gap(-1)	0.51	
				[0.12]	***
Observations	86		Observations	86	
Adj. R-sq	0.86		Adj. R-sq	0.93	
Breusch	1.17		Breusch	0.52	
Harvey	1.00		Harvey	0.17	
AIC	4.74		AIC	3.88	
Durbin-Watson stat.	1.86		Durbin-Watson stat.	2.00	
RMSE	2.40		RMSE	1.58	
TC	0.19		TC	0.13	

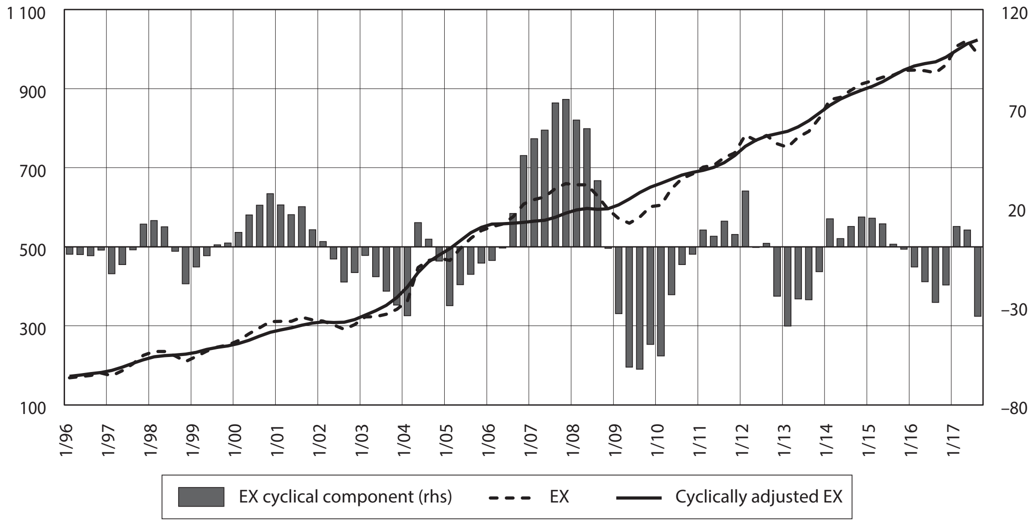
Note: J Estimates of the coefficients with standard errors in parenthesis are given. Adj. R-sq denotes the adjusted coefficient of determination. Breusch and Harvey represent tests of heteroscedasticity under the null of homoscedastic residuals. AIC is the value of Akaike information criterion. RMSE is a root mean square error. TC is Theil inequality coefficient. (*, **, *** denote rejection of the null at 10%, 5% and 1% level of significance, respectively).
Source: Own construction

[6] Import intensity of gross fixed capital formation is 44% based on symmetric input-output tables 2010.

2 RESULTS

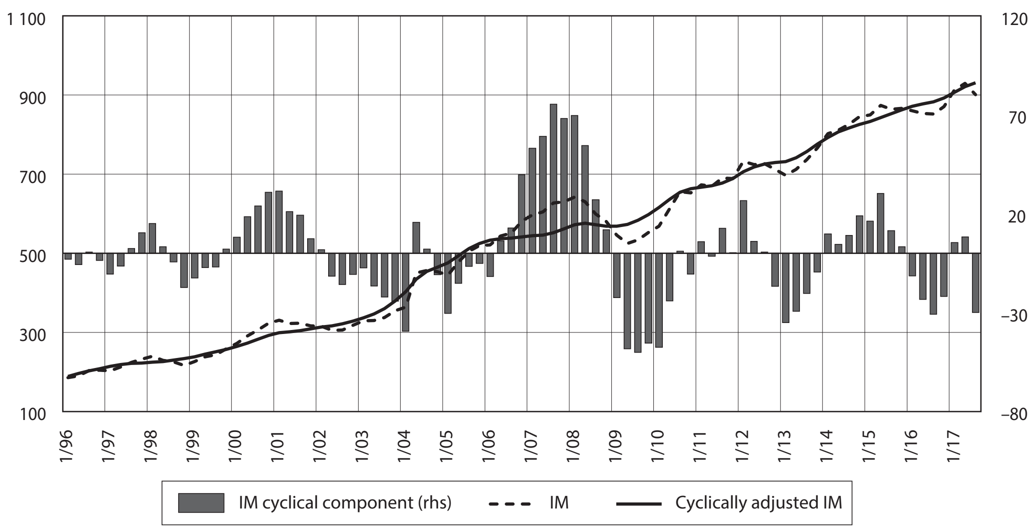
It follows from the breakdowns in Figure 6 and 7 that the Czech economy recorded the strongest cyclical fluctuations in exports and imports in the 2007–2008 boom and the subsequent decline in the recession in 2009.[7] The positive cyclical component of exports is evident in 2014 and 2015

Figure 6 Cyclically adjusted exports, (current prices, CZK bn)



Note: Adjusted values of export smoothed by the Hodrick-Prescott filter (λ = 1).
Source: Czech Statistical Office, own construction, (rhs = right hand side)

Figure 7 Cyclically adjusted imports, (current prices, CZK bn)



Note: Adjusted values of import smoothed by the Hodrick-Prescott filter (λ = 1).
Source: Czech Statistical Office, own construction, (rhs = right hand side)

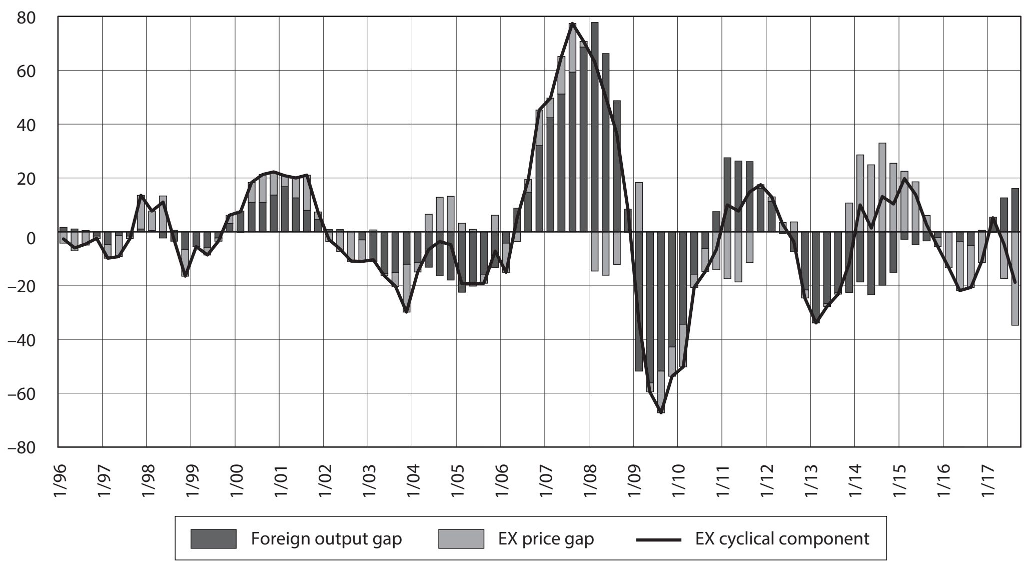
---

7 Presented data smoothed by the Hodrick-Prescott filter (λ = 1) for the sake of graphic readability.

in response to the devaluation of the koruna exchange rate by the CNB. On the import side, however, this effect was outweighed in 2015 by acceleration in import-intensive investment activity associated with dynamic end of allocation of EU funds from the previous financial perspective. For that reason, Figure 10 shows a significantly negative deviation of the balance of foreign trade in 2015.

Deviations of exports from the equilibrium levels are, for most of the time, created by the output gap abroad, which is also – indirectly – strongly reflected in imports. The deviation of prices in foreign trade from the trend is of a generally more pro-cyclical nature. However, in 2014, there is a clear positive contribution of the deviation of export prices stemming from weakening of the exchange rate after adoption of the CNB's exchange rate commitment. A relatively imminent reflection of the weakening of the koruna in the export deflator can be explained by price formation on export markets. Exported goods are sold at global prices, and devaluation will thus cause the koruna prices of exports and the profitability of exporters to jump. However, that effect gradually fades away over the horizon of 2 years due to the increased export supply, and pressures on reduction in foreign prices and trade margins of exporters at a stable exchange rate. It should be added that the crude oil price drop also affected the development during that period. Deviations of the balance of imports of goods and services from the equilibrium are determined mainly by export fluctuations, to a lesser extent by gross fixed capital formation and import prices.

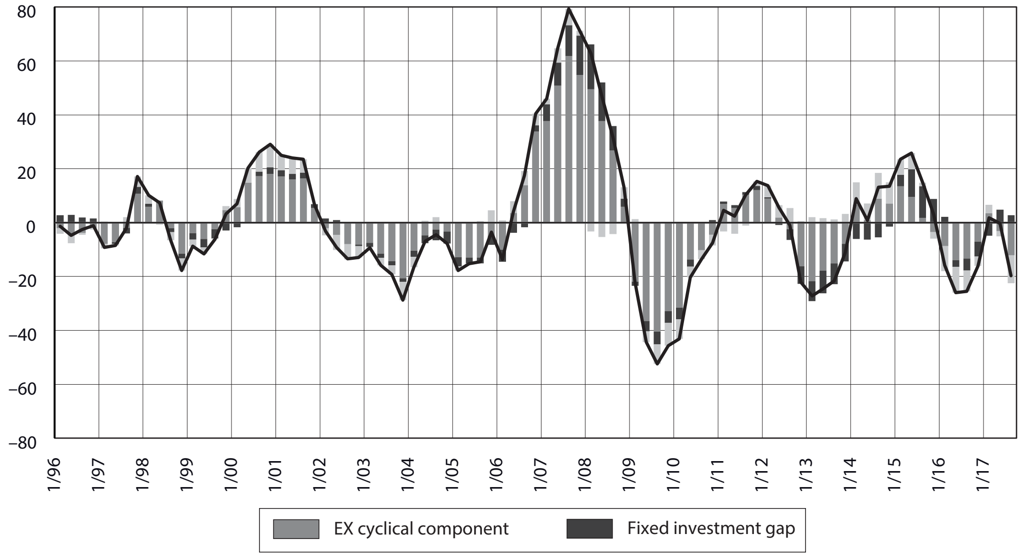
Figure 8 Contributions to exports cyclical component (current prices, CZK bn)



Source: Own construction

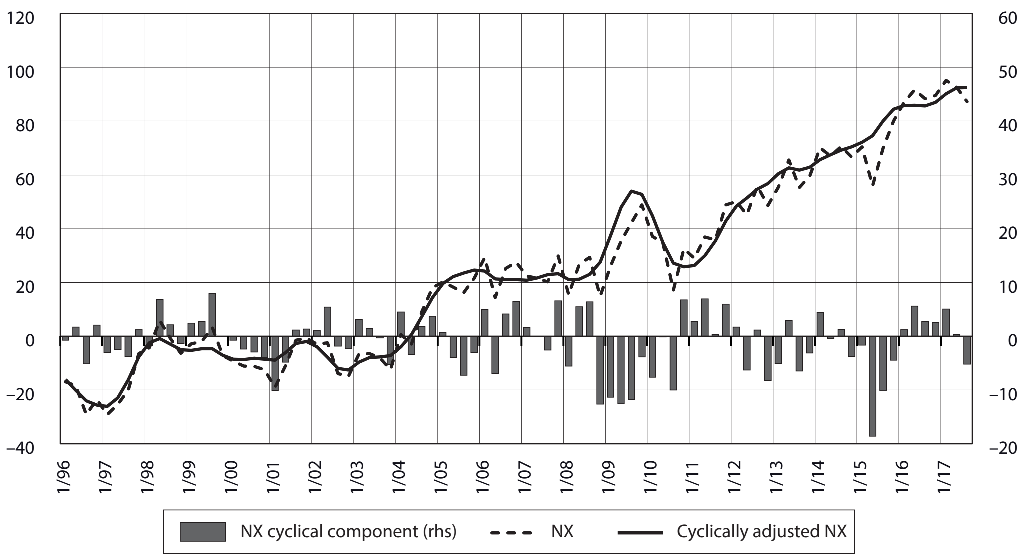
Exports, imports and the balance were close to their equilibrium values in mid-2017. However, the appreciation of the exchange rate after the discontinuation of the CNB's exchange rate commitment significantly reduced the nominal values of exports and imports and shifted the balance of goods and services below its equilibrium level in the third quarter of 2017 as a result of the synergy of the statistical effect (higher exports more influenced by the appreciation of the exchange rate than imports) and the classic economic effect (a stronger exchange rate curbs exports and promotes imports).

Figure 9 Contributions to imports cyclical component (current prices, CZK bn)



Source: Own construction

Figure 10 Cyclically adjusted net exports (current prices, CZK bn)



Note: Adjusted values of export and import smoothed by the Hodrick-Prescott filter (λ = 1).
Source: Czech Statistical Office, own construction, (rhs = right hand side)

CONCLUSION

Following the cyclical adjustment of a general governement sector overall balance, the article contributed to the methodology of cyclical adjusting another important macroeconomic variable, especially in the small open economy – balance of goods and services. Rather country specific cyclical adjustment procedure has been applied to the Czech economy by quantifying cyclical determinants of trade balance. Like the government sector balance, foreign trade is also affected by the behaviour of output gap. In addition, foreign trade prices deviations stemming mostly from the volatility of the exchange rate also play a significant role.

The analysis outlined usefulness of cyclical adjusting of foreign trade balance, which provides an apropriate tool for the assesment sustainability of essential part of the external equilibrium and helps not only to analyze current situation more accurately, but also to better predict short-term future development.

ACKNOWLEDGMENT

Article was written with the support of the Czech Science Foundation project No. 18-12340S Anti-cyclical policies and external equilibrium in a model of inflation targeting.

References

BLANCHARD, O. AND GIAVAZZI, F. Current account deficits in the euro area: the end of the Feldstein-Horioka puzzle? *Brookings Papers on Economic Activity*, 2002, No. 2, pp. 147–186.

BUITER, W. Time preference and international lending and borrowing in an overlapping-generations model. *Journal of Political Economy*, 1981, No. 89, pp. 769–797.

CA' ZORZI, M., CHUDIK, A., DIEPPE, A. Current account benchmarks for Central and Eastern Europe: a desperate search? *ECB Working Paper*, 2009, No. 995.

CHEUNG, C., FURCERI, D., RUSTICELLI, E. Structural and Cyclical Factors behind Current Account Balances. *OECD Working Papers*, 2010, No. 775.

CHINN, M. AND PRASAD, E. Medium-term determinants of current accounts in industrial and developing countries: an empirical exploration. *Journal of International Economics*, 2003, No. 59, pp. 47–76.

CZECH STATISTICAL OFFICE. *Quarterly National Accounts* [online]. Prague: CZSO, 2018. [cit. 2.2.2018]. < https://www.czso.cz/csu/czso/hdp_cr>.

FABIANI, S., FEDERICO, S., FELETTIGH, A. *Adjusting the external adjustment: cyclical factors and the Italian current account.* Banca d´Italia, Ocasional papers, 2016, No. 346.

FARUQEE, H. AND DEBELLE, G. What Determines the Current Account? A Cross-Sectional and Panel Approach. *IMF Working Paper*, 1996, No. 96/58.

GRUBER, J. AND KAMIN, S. Explaining the global pattern of current account imbalances. *Journal of International Money and Finance*, 2007, No. 26, pp. 500–522.

HALTMAIER, J. Cyclically Adjusted Current Account Balances. *International Finance Discussion Papers*, Board of Governors of the Federal Reserve System, 2014, No. 1126.

HALTMAIER, J. Empirical Estimates of Trend and Cyclical Export Elasticities. *International Finance Discussion Papers,* Board of Governors of the Federal Reserve System, 2011, No. 1030.

HODRICK, R. AND PRESCOTT, E. Postwar U.S. Business Cycles: An Empirical Investigation. *Journal of Money, Credit, and Banking,* 1997, No. 29(1), pp. 1–16.

KARA, H. AND SARIKAYA, C. *Cyclically Adjusted Current Account Balance.* Türkiye Cumhuriyet Merkez Bankasi, Research notes in economics, 2013, No. 18.

MINISTRY OF FINANCE OF THE CZECH REPUBLIC. *Macroeconomic Forecast of the Czech Republic*. January, 2018. ISSN 1804-7971.

OBSTFELD, M. AND ROGOFF, K. *The intertemporal approach to the current account*. Handbook of International Economics, 1995, No. 3, pp. 1731–99.

PHILLIPS, S. et al. The External Balance Assessment (EBA) Methodology. *IMF Working Paper*, 2013, No. 272.

SACHS, T. The current account and macroeconomic adjustment in the 1970s. *Brookings Papers on Economic Activity*, 1981, No. 1, pp. 201–268.

WU, Y. Growth, Expansion of Markets, and Income Elasticities in World Trade. *Review of International Economics*, 2008, Vol. 16, Iss. 4, pp. 654–671.

On Extending Composite Leading Indicators by International Economic Series

Lenka Vraná [1] | *University of Economics, Prague, Czech Republic*

Abstract

Composite leading indicators (CLIs) are recognized as eligible tools for business cycle analysis. When the Organization for Economic Co-operation and Development (OECD) constructs CLI, its composition depends on national data only. However, European economies are often small and open and therefore their business cycles relate to situations in other countries. The approach described in this paper reflects these characteristics. The international CLIs for Austria, the Czech Republic, Germany, Poland and Slovakia are constructed and the leading influences on these countries are discussed.

The methodology of the CLI construction is described in detail by several organizations. It, therefore, comes as a surprise, that there are no publicly available software programs, R packages or Python libraries, that would support the whole computational process or its automation. A new Python-based framework is proposed to fill this gap and it is demonstrated on the international CLI construction. It is introduced for the very first time in this paper and it enables users to quickly analyze and visualize larger volumes of data than any other available solution.

Keywords	***JEL code***
Business cycle analysis, leading indicators, automation	*C32, E32*

INTRODUCTION

Although we are now in the phase of economic expansion, the public attention has been focused on the possibilities of forecasting the business cycle movements since the recent Great Recession in 2007. One of the methods used for the business cycle analysis is based on the study of composite indicators (CI) which combine several individual economic time series. The series can be divided into groups of leading, coincident and lagging ones with regard to the reference time series (usually gross domestic product (GDP) or industrial production index).

The composite leading indicator (CLI) draws most of the attention because it should be able to predict the future states of economic activity – when the economy is going to switch from the expansion phase into the contraction phase or vice versa. Astolfi et al. (2016, p. 15) study the performance of the real-time CLI warnings during the Great Recession in 2007 and state that "in both cases, at peak and trough,

[1] Faculty of Informatics and Statistics, Department of Statistics and Probability, W. Churchill Sq. 4, Prague 3, Czech Republic, E-mail: lenka.vrana@vse.cz.

the OECD (Organization for Economic Co-operation and Development) was able to signal the approaching turning points thanks to the continuous monitoring of the CLI growth rates, which initially recorded a significant reduction and then turned negative. With hindsight, the CLI for the OECD area as a whole peaked in June 2007, hence seven months ahead of the corresponding peak for GDP, which took place in December 2007".

The construction of composite indicators usually follows methodology created by OECD or the Conference Board. Most of the analyses of the Czech business cycle use the OECD methodology which is employed in this paper as well. The OECD methodology according to Gyomai and Guidetti (2012) consists of five steps: 1. *pre-selection phase*, which is passed only by long time series of indicators that have justified economic relationship with the reference series, broad coverage of economic activity and high frequency of observations, 2. *filtering phase*, when the time series are seasonally adjusted and de-trended, 3. *evaluation phase*, when only the best individual indicators with the strongest relationship with the reference series are selected to be included in the composite indicator, 4. *aggregation phase*, when the composite indicators are created and 5. *presentation of the results*. For the detailed description of these processes, see section 3.

OECD publishes CLI for most of its 35 member countries and for some partner countries (e.g., Brazil, India, the People's Republic of China). It also compiles the CLI of the whole G7, NAFTA, Euro area, European OECD and all OECD countries. However, the composition of each country's CLI depends on the national input data only. For example, Czech OECD CLI consists of Czech individual economic indicators. The European economies are nevertheless often small and open and therefore their business cycles relate to the situation in other countries.

This paper aims to construct the international CLI, which is based on input data from multiple countries and assess:

- if considering international data changes the structure of OECD CLIs,
- whether international CLIs can be used to analyze the relationships between business cycles of several countries,
- how can these relationships be interpreted and visualized.

This is not the first time when the researchers use international data to construct national CLI; however, this is the first case, to the best of my knowledge, where the structure of international CLI is visualized on maps and used to interpret the relationships between the countries.

Such analyses cannot be performed in any of the publicly available software programs. Therefore, the new Composite Indicators Framework (CIF) is proposed and introduced for the very first time in this paper. This framework is described and compared with the current solutions in section 2.

Section 4 utilizes the framework and demonstrates the international CLI construction on data from 5 countries: Austria, the Czech Republic, Germany, Poland and Slovakia. These countries were selected to simplify the interpretations and visualizations of the results. However, deploying the proposed framework guarantees that the international CLI could be easily based on all European data or data available all over the world.

1 BRIEF HISTORY OF COMPOSITE BUSINESS CYCLE INDICATORS

In 1930's two American economists, Arthur Frank Burns and Wesley Clair Mitchell, worked on new methods how to measure business cycles and determine recessions. In 1946 they published work on business cycle indicators which contained one of the first lists of leading, coincident and lagging indicators as well as a set of instructions how to track the cycle. Burns and Mitchell's methodology spread worldwide in the following years. It was executed manually and required lots of personal judgment and therefore it wasn't quite objective.

In 1969 Ilse Mintz proposed a new definition of business cycles, which became later known as growth or deviation cycles. The deviation cycle was obtained by removing trend from the reference time series and could be interpreted as output gap (the difference between the actual and potential economy output). Mintz also brought up new terms: speedups and slowdowns of the economy instead of expansions and contractions known from Burns and Mitchell's classical cycles.

In 1971 Gerhard Bry and Charlotte Boschan introduced their algorithm to automate the turning points detection. It was one of the first programmed approaches that were published and, with the fast development of information technologies, was then widely implemented.

OECD, Conference Board and other organizations still use Bry-Boschan algorithm with only slight changes. In their first proposal, Bry and Boschan used a 12-month moving average, Spencer curve and a short-term moving average of 3 to 6 month to detect the turning points. Nowadays, none of these are necessary because some other techniques (like Hodrick-Prescott filter) are used to smooth the time series without shifting the turning points.

The OECD methodology was described in detail by Gyomai and Guidetti (2012). For more information on the Conference Board methodology see its Business Cycle Indicators Handbook (2001) or Ozyildirim et al. (2010). General findings on composite indicators as well as detailed remarks on OECD, Conference Board and other methodologies were elaborately summarized in Eurostat (2017).

Authors all around the world used OECD or Conference Board methodologies and proposed their own improvements to the specific parts of these processes. Svatoň (2011) proposed Granger causality test to limit the number of candidate series during the pre-selection phase. Zarnowitz and Ozyildrim (2006) compared the application of phase-average trend with Hodrick-Prescott and Baxter-King band pass filters during the filtering phase of composite indicators construction. Nilsson and Gyomai (2011) then added Christiano-Fitzgerald filter into the comparison. The evaluation phase also drew plenty of suggestions: Hamilton (1989) introduced Markov switching approach later modified by Levanon (2010) to compare the recession signal across many indicators. Bruno and Otranto (2004) studied combinations of parametric and non-parametric methods and their impact on business cycle dating. Gallegati (2014) proposed wavelet-based composite indicator, which provided early warning signals of turning points. The aggregation phase was also subject of research: Zhou et al. (2009) introduced a mathematical programming approach to optimize the weights of individual indicators for human development index and the similar technique could be used to analyze business cycles as well.

2 COMPUTATIONAL FRAMEWORKS

The methodology of CLI construction is described in detail by several organizations. It, therefore, comes as a surprise, that no publicly available software program supports the whole computational process or its automation. A new framework is proposed to fill this gap and it is demonstrated in this paper for the very first time. The construction of the international CI, performed in this paper, would otherwise be very difficult. The new framework enables users to analyze and visualize larger volumes of data than any other available solution.

In this section, software programs suitable for constructing composite indicators and the newly proposed Composite Indicators Framework (CIF) are described and compared.

2.1 Available software programs

This paper overviews some of the existing software solutions suitable for analyzing business cycles: CACIS, EViews, Python and R (in alphabetical order). These are selected as they cover most of the tasks required to construct composite indicators.

Python and R are free and general software environments very popular in data science. They provide a great selection of libraries (Python) and packages (R) with many functions and methods, so users do not need to write them from the scratch. Moreover, it is also possible (although sometimes time-consuming) to program the missing pieces.

EViews is oriented mainly on time series analyses and forecasting. It provides a basic graphical user interface, but also requires the knowledge of its own programming language. It is the only commercial software discussed in this paper.

OECD offers its own Cyclical Analysis and Composite Indicators System (CACIS). It is designed directly to compute the composite indicators and, therefore, it provides the most exhaustive pallet of functions. However, the publicly available version has not changed much since 2010, and its user interface and generated visualization are obsolete.

CACIS and EViews run only on Windows operating system.

2.2 Newly proposed computational framework

None of the existing solutions provides all the tasks required to analyze the business cycle. Because the switching between several software programs is uncomfortable, slow and makes the automation of the process impossible, the new solution needs to be proposed. This paper introduces the new framework attempting to fill the gap in currently available software repertoire.

CACIS was developed to analyze the cycle, but it is now obsolete and does not meet certain basic requirements like loading the data directly from OECD API, adjusting the graphs or automating the whole process. EViews provides even less of the specified tasks, and its results are only slightly more controllable by the user. Therefore, the new framework could be based either on Python or R. After some experiments, Python was selected as it provided elegant syntax, better performance and its existing libraries were more compatible with each other.

If users miss some functions, the Python-based interface of CIF guarantees that they can write them by themselves and easily integrate them into the computing process. For example, creating CLI from international data and visualizing them as leading influence maps (as is described in section 4.3) would not be possible with any other described framework.

The current version of CIF is available on GitHub[2] (internet platform for sharing, collaboratively developing and documenting code) as a Python file, which nowadays contains thousands of lines of code in 33 functions to support and automate the entire process of CI construction. See the GitHub repository for the complete list of available functions and minimal functional pipeline to help the users start analyzing with CIF for the first time. CIF is soon going to be available as a classical Python library installable via pip command.

Table 1 overviews the selected software solutions and CIF and evaluates them in the fields most essential for constructing composite indicators.

EViews, Python and R load data from versatile data sources (databases as well as data files). Python and R can also communicate with other applications via application programming interface (API), which enables them to download data directly from organizations like OECD when connected to the internet. However, they do not process such data automatically so additional steps are needed to transform it into general data table format. CACIS can load data from excel and csv files only, so it requires lots of manual work while downloading and preparing input data.

The new framework (CIF) focuses mainly on the automation of the construction. It is designed to save the time of the users, so they can just load the input data and the result is delivered with their minimal

[2] Available at: <https://github.com/LenkaV/CIF>.

effort and without manual intervention. Alternatively, the user can specify only the country of the interest and the available data are downloaded directly from OECD API (other APIs will be added in the future). These functions allow the users to quickly compare results across many countries and eliminate the time needed for the bothersome data transformations.

Table 1 Overview of software functions necessary for composite indicators construction (available = fully supported, generic = supported, but some adjustments needed, x = not supported)

actions	CACIS	EViews	Python	R	CIF
loading data from files	excel or csv only	available	available	available	available
loading data from databases	x	available	available	available	available
loading data from generic API	x	x	available	available	available
loading data from OECD API	x	x	x	x	available
convert quarterly to monthly time series	available	available	x	available	available
seasonal adjustment and outlier detection (TRAMO/SEATS)	available	available	available	available	available
de-trending and smoothing (Hodrick-Prescott filter)	available	available	available	available	available
normalization	available	generic	generic	generic	available
turning-point detection (Bry-Boschan alg.)	available	x	x	quarterly data only	available
turning-point matching	available	x	x	x	available
aggregation	available	x	x	x	available
visualization	available	generic	generic	generic	available
custom development	x	x	available	available	available
evaluation (ex-post)	available	x	x	x	available
evaluation (real time)	x	x	x	x	available
automation	x	x	generic	generic	available
logs	x	x	generic	generic	available

Source: Own construction

All the presented software solutions can perform the time series transformations: seasonal adjustments, detrending, smoothing and normalization. The other tasks which are necessary to construct CI (turning points detection, turning points matching and aggregation) are offered only by CACIS, but often require substantial manual interventions. R also contains package BCDating for turning points detection, but it works only with quarterly time series.

Python and R offer a vast number of visualization libraries and packages and, therefore, enable almost any type of diagrams. EViews also provide some visualization capabilities but (compared to the previously mentioned solutions) they are limited. None of these programs contains the exact charts needed to illustrate the cycle analyses and CLI construction. As CACIS is the only existing program developed directly to analyze the business cycle, it contains the necessary visualizations, but it doesn't enable the users to alter them in any way. In contrast, newly proposed CIF provides a great variety of fully adjustable charts and other diagrams that can accompany the cycle analysis.

CACIS offers only the ex-post analysis of CI performance. The ex-post analysis usually overestimates the quality of constructed indicator and therefore should be accompanied by the real-time analysis, which considers also the historical revisions of economic series and lags between the events and their publication. For more information on the real-time evaluation, see Astolfi et al. (2016). CIF offers both the ex-post and the real-time quality assessment.

CIF also thoroughly records the whole computation process and saves the logs for later examination. One of CIF's main goals is to enable users to run the analysis automatically, without any manual interventions to the process.

3 METHODOLOGY

This section describes the OECD methodology which is employed in this paper with slight changes in pre-selection and evaluation phases. These modifications enable to construct international composite indicators in section 4.

3.1 Pre-selection

When constructing the composite indicators, the eligible individual series has to be selected first. According to Gyomai and Guidetti (2012) only long-time series of indicators that have the justified economic relationship with the reference series, broad coverage of economic activity, high frequency of observations (preferably monthly), that were not subject to any significant revisions and are published soon, should pass the pre-selection phase.

The easiest way when deploying CIF is to automatically download the whole table of main economic indicators supplied by OECD.[3] When OECD constructs its CLI, it takes into account only data from the analyzed country. The other way would be to consider also individual indicators from other countries (e.g., neighboring ones). The latter approach is described in section 4 herein.

The construction of composite indicators is highly dependent on the selection of reference time series. Usually, GDP or index of industrial production is used as reference series. GDP should respond to the cyclical movements better but it is quarterly statistics and it needs to be converted to the monthly estimates. OECD had used the industrial production index until March 2012 and then switched to the adjusted monthly GDP, which is also used in this paper.

3.2 Filtering

The second phase of the composite indicator construction is called the filtering. The main task of this stage is to decompose the individual time series and find their cyclical component. This means that the series have to be seasonally adjusted with their trend component removed.

The quarterly series need to be converted to a monthly frequency. Gyomai and Guidetti (2012, p. 6) describe, that OECD uses simple linear interpolation: "This conversion from quarterly to monthly is achieved via linearly interpolating quarterly series and aligning them with the most appropriate month of the quarter, depending on the nature/construction of the quarterly series (…), for most series this is the central month of the quarter (...)."

OECD uses TRAMO module from TRAMO/SEATS provided by the National Bank of Belgium to identify outliers, seasonally adjust the series and provide short horizon stabilizing forecasts before detrending the series (OECD, 2010). CIF utilizes the X-13ARIMA-SEATS Seasonal Adjustment Program developed by the United States Census Bureau because it is already integrated into Python in the Statsmodels library.

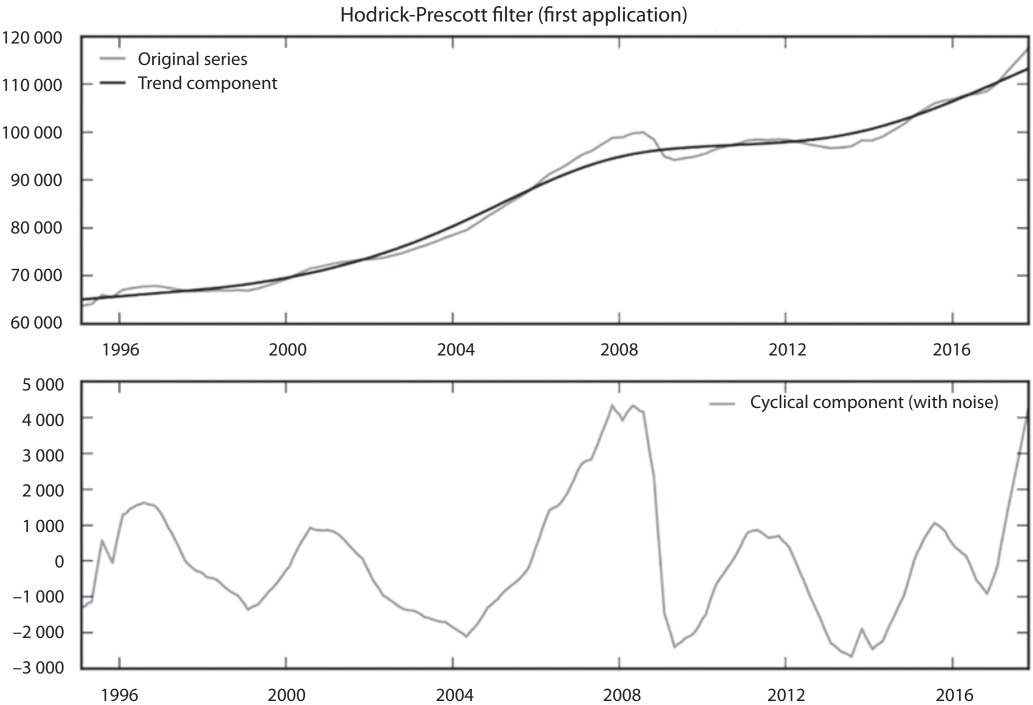
[3] Available at: <http://www.oecd-ilibrary.org/economics/data/main-economic-indicators_mei-data-en>.

Hodrick-Prescott filter divides the series into two parts (τ_t – trend component and c_t – the cyclical component) and optimizes expression:

$$\min_{\tau_t}\left[\sum_t\left(y_t-\tau_t\right)^2+\lambda\sum_t\left(\tau_{t+1}-2\tau_t+\tau_{t-1}\right)^2\right]. \tag{1}$$

It minimizes the difference between the trend and the original series and smooths the trend as much as possible at the same time. The λ parameter prioritizes the latter from the two contradictory goals – the higher the λ, the smoother the trend.

Figure 1 First application of Hodrick-Prescott filter on the Czech gross domestic product (in US dollars, monthly estimates, seasonally adjusted) with high lambda parameter to remove the trend

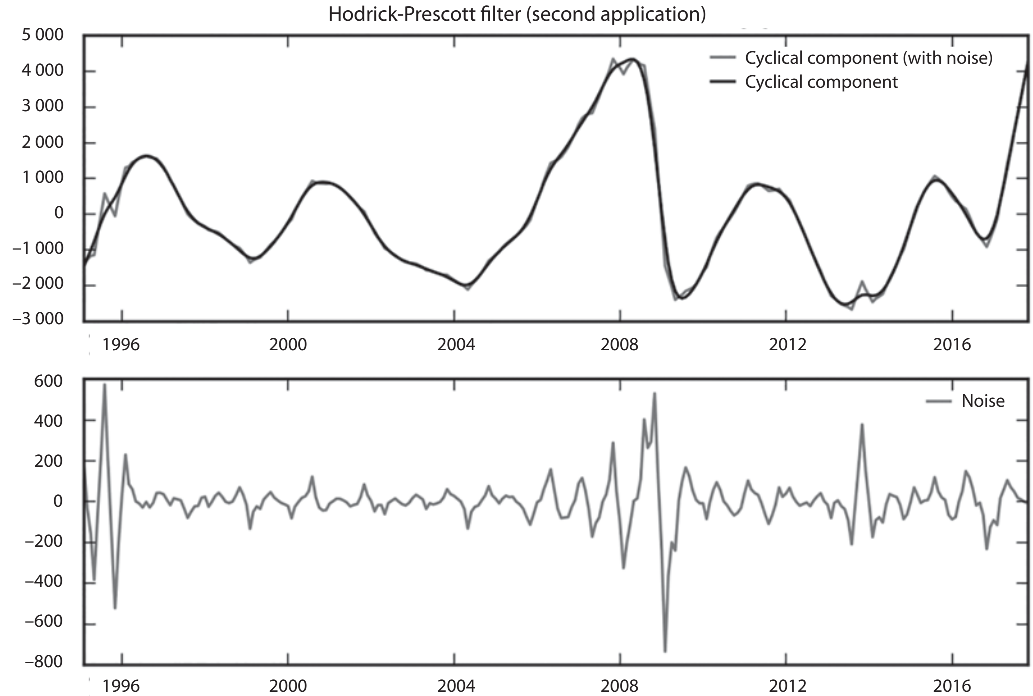


Source: Data from OECD (2017), processed by CIF

Hodrick-Prescott filter deals with the series as with the system of sinusoids and it keeps in the trend only those with low frequency (high wave length). According to OECD the business cycles last 10 years at maximum, therefore the fluctuations with lower wave length should be kept in the cycle component. Ravn and Uhlig (2002) recommend setting the λ parameter equal to 129 600 for monthly series. Nilsson and Gyomai (2011) use the Hodrick-Prescott filter twice: first with high λ to find the trend and then with low λ to smooth the cycle component. They also confirm that Hodrick-Prescott filter gives clear and steady turning point signals. Figures 1 and 2 illustrate the effects of the filter on Czech GDP.

After the trend component is estimated, it is subtracted from the original data set (this is called the deviation cycle then) and the series are normalized.

Figure 2 Second application of Hodrick-Prescott filter on detrended Czech gross domestic product with low lambda parameter to smooth the cyclical component



Source: Data from OECD (2017), processed by CIF

3.3 Evaluation

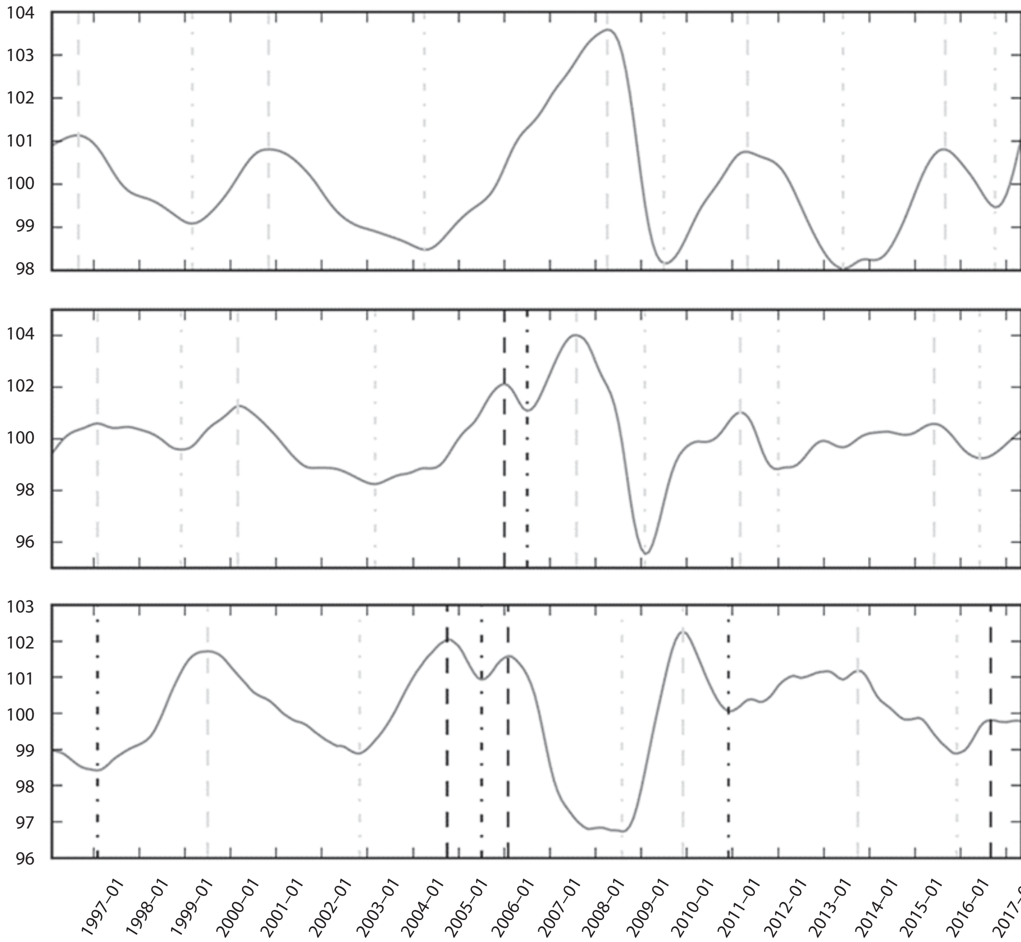
After the cycle components of all the individual indicators are found, turning points are detected. Not every peak or trough of the cycle is considered as the turning point though. OECD uses Bry-Boschan algorithm (Bry and Boschan, 1971) to determine the turning points.

The cyclical components of all the individual indicators are compared to the reference series. OECD uses several methods how to evaluate their relationship: the average lead (lag) times between the turning points, number of extra and missing cycles and cross correlations. Then the selected individual indicators are divided into groups of leading, coincident and lagging ones according to their characteristics.

Figure 3 presents one of the visualizations created by CIF to compare the turning points detected in individual and reference series. It shows the normalized cyclical component of the reference series (Czech GDP) and its turning points found by Bry-Boschan algorithm in the upper chart. The middle and the lower charts depict the normalized cyclical components of Czech share prices and unemployment, respectively. The unemployment shows typical counter-cyclical behavior and the series needs to be inverted before the next steps. The extremes of the individual series are marked in gray, if the corresponding turning points are detected in the reference series, or in black, when these are false signals. The corresponding turning point must occur in the neighborhood of the reference series extreme to be considered as matched: with the maximal lead of 24 months or maximal lag of 9 months according to the Eurostat (2017). CIF also enables to mark each matched extreme in a different color (not all in gray) to better distinguish their chronology and the missingturning points.

There is a discrepancy between the scales of y-axes, which is intentionally kept in this chart type. This plot should serve mainly to compare the turning points and not the level of the series because the amplitude of the composite indicator (or its normalized component series) can be interpreted only as the confidence of the CLI outlook and never to analyze the level of the economy (Eurostat, 2017).

Figure 3 Turning points detected by Bry-Boschan algorithm in selected Czech indicators. All series have been seasonally adjusted, detrended and normalized. Upper chart: Gross domestic product[4] (reference series). Middle chart: Share prices.[5] Lower chart: Unemployment of persons aged 15–24.[6] Extra turning points are marked in black.



Source: Data from OECD (2017), processed by CIF

[4] Detected dates of the turning points: 1996-09 (Peak), 1999-03 (Trough), 2000-11 (P), 2004-04 (T), 2008-04 (P), 2009-07 (T), 2011-05 (P), 2013-06 (T), 2015-09 (P), 2016-10 (T).

[5] Detected dates of the turning points: 1997-02 (P), 1998-12 (T), 2000-03 (P), 2003-03 (T), 2006-01 (P), 2006-07 (T), 2007-08 (P), 2009-02 (T), 2011-03 (P), 2012-01 (T), 2015-06 (P), 2016-06 (T).

[6] Detected dates of the turning points: 1997-02 (T), 1999-07 (P), 2002-11 (T), 2004-10 (P), 2005-07 (T), 2006-02 (P), 2008-08 (T), 2009-12 (P), 2010-12 (T), 2013-10 (P), 2015-12 (T), 2016 09 (P).

After each indicator is compared with the reference series, only the best-performing ones are selected into the composite indicator. The number of selected indicators may differ across the countries and depends on the criteria setup. OECD (2010, p. 31) defines these criteria quite laxly: "Ideally, potential component series should have a mean lead greater than 2 and a correlation at peak greater than 0.5 (with a peak lead equal or greater than 2). (…) Furthermore, users should bear in mind that a series provides valuable information if it does not flag too many extra cycles and does not miss too many turning points". Unlike some classical models like linear regression, the composite indicator quality does not depend on the number of input individual indicators, and its performance may decrease with additional indicators.

As the turning point detection and evaluation are the parts of the CI construction, which is covered the least in existing software programs, some authors try to avoid it by using only cross correlations which are much easier to compute. However, the Eurostat (2017, p. 286) states, that "the location of the peak of the cross correlation function is a good alternative indicator of average lead time. Whereas the correlation value at the peak provides a measure of how well the cyclical profiles of the indicators match, the size of correlations cannot be the only indicators used for component selection". As CIF contains the proper turning points detection, it can help the researchers to avoid similar quality endangering shortcuts.

3.4 Aggregation

In this phase, the selected individual indicators are aggregated into the leading, coincident or lagging composite indicators. This paper focuses on the leading composite indicator, which can help to predict the next regression or expansion of business cycle.

Different weighting schemes can be utilized during aggregation. However, OECD does not use any weights so all the input series have equal impacts on the constructed CLI. Another possibility during this phase is to lag-shift the input series with the longer lead, so their signals do not get neutralized by series with the shorter lead. This can lead to signals with shorter-lead, but enhanced quality.

CLI is published when at least 60% of selected input indicators are available. The chain linking method is used to prevent jumps and discontinuities when new series are added. For more details, see Eurostat (2017).

3.5 Presentation of the results

OECD publishes the final CLI in 3 forms:

- the amplitude adjusted CLI, which can be compared with normalized values of the cyclical component of the reference series,
- the trend restored CLI, which can be compared with the original values of the reference series,
- the 12-month growth rate of CLI, which can be compared with the 12-month growth rate of the reference series.

This paper goes further and analyzes the structure of the constructed CLI thoroughly. The choropleth maps are suggested as the tool to assess the leading influences between countries.

4 INTERNATIONAL LEADING COMPOSITE INDICATORS

OECD publishes CLIs for most of its 35 member countries and for some partner countries. It also compiles the CLI of the whole G7, NAFTA, Euro area, European OECD and all OECD countries. However, the composition of each country's CLI depends on the national input data only. For example, Czech OECD CLI consists of Czech individual economic indicators. The European economies are nevertheless often small and open and therefore their business cycles relate to the situation in the surrounding countries.

This is not the first time when the authors use international data to construct national CLI. For example, the authors of the Czech CLIs often include German economic series in their indicators, for more details see Svatoň (2011) or Vraná (2013). However, this is the first case, to the best of my knowledge, when the structure of international CLI is visualized on maps and used to interpret the relationships between the countries.

The international CLI construction is demonstrated on data from 5 countries: Austria (AUT), the Czech Republic (CZE), Germany (DEU), Poland (POL) and Slovakia (SVK). Each business cycle is compared with available indicators from all of these countries and the best matching ones are selected as its CLI elements.

These countries were selected to simplify the interpretations and visualizations of the results. However, deploying the newly proposed framework guarantees that the international CLIs could be easily based on all European data or data available all over the world.

The data for this paper were downloaded[7] from OECD API via CIF.

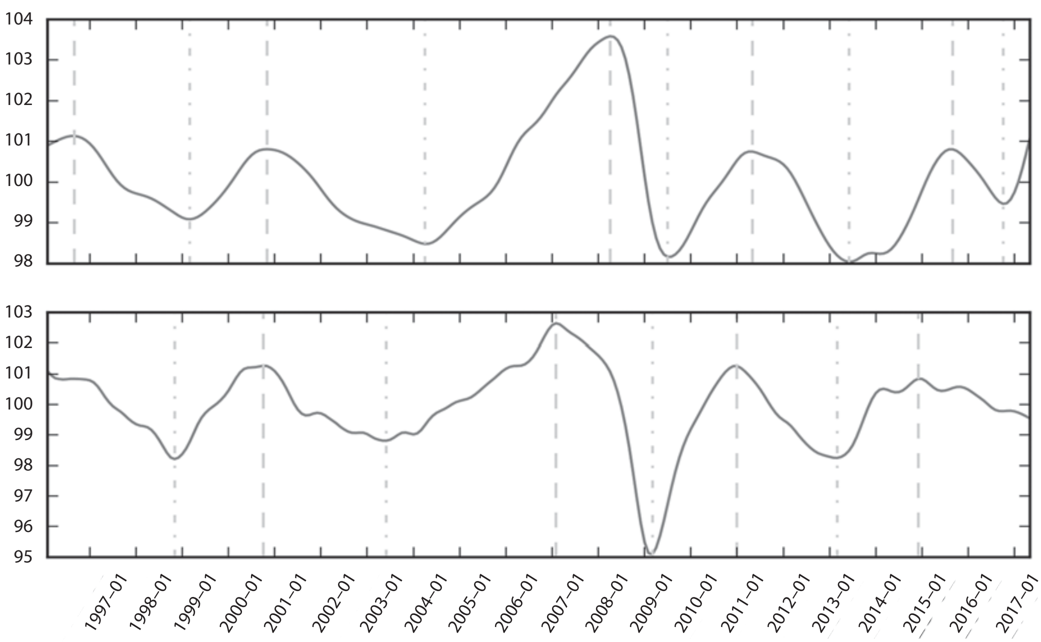
Another important data source for this chapter is GADM (Global Administrative Areas) spatial database, which provides mapping files (Hijmans et al., 2015).

4.1 Construction and basic characteristics

The analysis follows the OECD methodology and was performed completely with the CIF described in section 2.2. Only minor parts of the pre-selection and evaluation phases need to be altered for this use case:

- the pre-selection phase contains input data from multiple countries,
- the number of individual economic indicators selected to be aggregated into CLI is fixed to 15 during the evaluation phase.

Figure 4 Comparison of turning points of Czech GDP (upper chart) and OECD CLI[8] (lower chart)



Source: Data from OECD (2017), processed by CIF

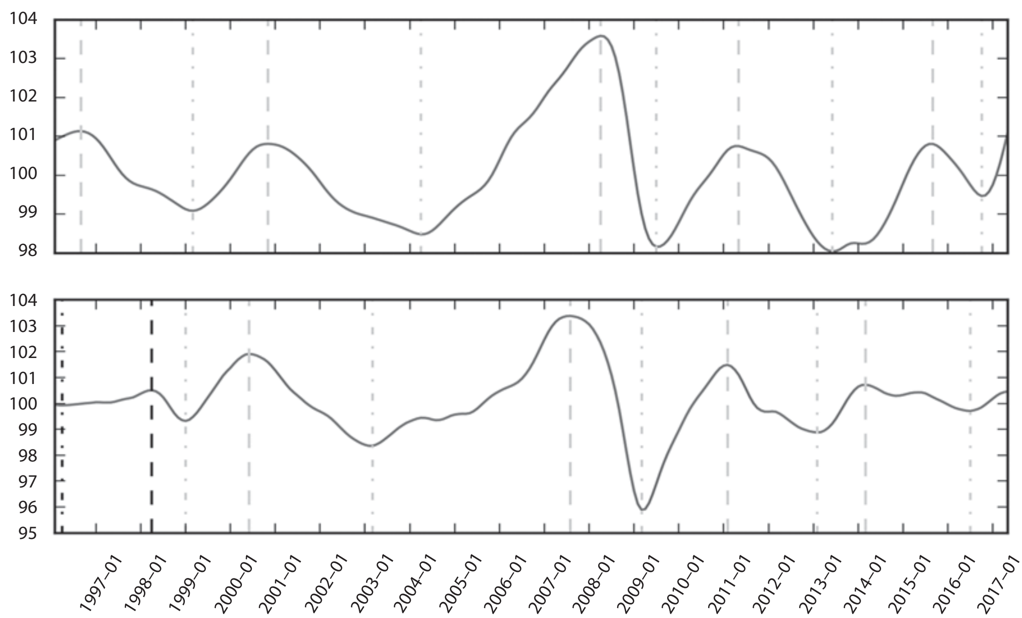
7 On the 17th September 2017.

8 Detected dates of the turning points: 1998-11 (T), 2000-10 (P), 2003-06 (T), 2007-02 (P), 2009-03 (T), 2011-01 (P), 2013-03 (T), 2014-12 (P).

The number of selected individual indicators is fixed to enable the comparison of international influences between several countries. Usually, this kind of prerequisite is not necessary because the quality of constructed CLI does not depend on the number of selected indicators as was explained in section 3.3. However, it is the structure, not the performance quality, which is the main focus of this paper.

Figures 4 and 5 display OECD CLI and international CLI of the Czech business cycle, respectively. Each figure shows reference series (GDP) with detected turning points in the upper chart and CLI with the matched (or unmatched) turning points in the lower chart. If the CLI contains any extra turning points, they are marked in black.

Figure 5 Comparison of turning points of Czech GDP (upper chart) and international CLI[9] (lower chart). 2 extra turning points detected in the beginning of the international CLI are marked in black.



Source: Data from OECD (2017), processed by CIF

This paper aims to analyze and visualize the leading influence between multiple countries. Its primary goal is not the improvement or assessment of current OECD CLI quality. However, the comparison with the OECD results should not be avoided entirely: Tables 2 and 3 show basic statistics of OECD CLIs and international CLIs, respectively. They summarize the number of missing and extra turning points, mean and medium lead time of turning points, maximum and location of the peak of the cross correlation function and the cross-check (the difference between the correlation peak location and the median lead). Eurostat (2017, p. 286) states that "the lead at which the highest correlation occurs should not be too different from the median lead if the composite leading indicator is to provide reliable information about approaching turning points and the evolution of the reference series."

9 Detected dates of the turning points: 1996-04 (T), 1998-04 (P), 1999-01 (T), 2000-06 (P), 2003-03 (T), 2007-08 (P), 2009-03 (T), 2011-02 (P), 2013-02 (T), 2014-03 (P), 2016-07 (T).

Table 2 Basic characteristics of OECD CLIs

	missing	extra	mean lead time	median lead time	cross correlation maximum	cross correlation peak location	cross-check
AUT	0	2	3.00	3.00	0.77	6	3.00
CZE	1	0	6.78	4.00	0.83	6	2.00
DEU	0	5	5.00	5.00	0.77	6	1.00
POL	1	0	9.85	11.00	0.41	12	1.00
SVK	0	5	7.40	9.50	0.78	1	8.50

Source: Data from OECD (2017), processed by CIF

Table 3 Basic characteristics of international CLIs

	missing	extra	mean lead time	median lead time	cross correlation maximum	cross correlation peak location	cross-check
AUT	0	2	6.83	6.00	0.85	8	2.00
CZE	0	2	7.70	4.50	0.82	5	0.50
DEU	0	2	6.60	7.00	0.88	6	1.00
POL	0	1	4.64	3.50	0.79	7	3.50
SVK	0	2	5.25	6.00	0.75	5	1.00

Source: Data from OECD (2017), processed by CIF

The CLI is computed when at least 60% of individual indicators are available. The international time series tend to vary in lengths substantially. E.g., OECD provides the first German main economic indicators from January 1955 and first Czech ones not sooner than January 1990. That means that if German CLI is composed mainly of Czech economic indicators (which of course is a hypothetical situation), it could not be calculated sooner than January 1990 (and probably even later, as only a few Czech indicators are available right from the beginning of this timespan). Therefore, the length of the OECD CLI and international CLI may differ and their lengths are adjusted to the shorter one of the two to enable comparison.

Tables 2 and 3 report that the Austrian and German international CLIs show longer leads (and the German one also gives less false signals) than the OECD CLIs, which is based on their national data only. The Czech international CLI also displays improvement in the leading time measured by mean and median, but not by cross correlation (for comparison, see also Figures 4 and 5). The Slovak OECD CLI seems to perform better in mean and median lead time, but the difference between median and location of the peak cross correlation is high and, therefore, the international CLI would probably provide more stable results. The only OECD CLI clearly outperforming the international CLI is the Polish one with more than 3 times longer median lead time (although with the lower correlation coefficient). However, even the Polish CLI could still be improved by adding international data from other countries, then those selected in this paper.

More than 5 countries should be analyzed to achieve the proper comparison of the OECD and international CLIs performance. Such analysis should not be based on a single time instant, but it should also involve historical data. However, the general comparison is not the goal of this paper and it remains for a future work. The aim of this paper – analysis of international influence – is the subject of the following section.

4.2 Leading influence maps

The international CLIs can be used to analyze the relationships, similarities and differences between business cycles of selected countries. Table 4 summarizes the structure of each constructed CLI (for the complete overview see the Appendix). The input number of selected individual economic indicators was artificially set to 15 as was explained in section 4.2, therefore the total equals 15 for each column. The row totals represent the frequency of the national individual indicators in all of the constructed CLIs. The higher this number is, the more common it is for the individual indicators of this country to appear in the leading indicators. This could also be interpreted as the economic lead or influence the country has when it is compared to the others.

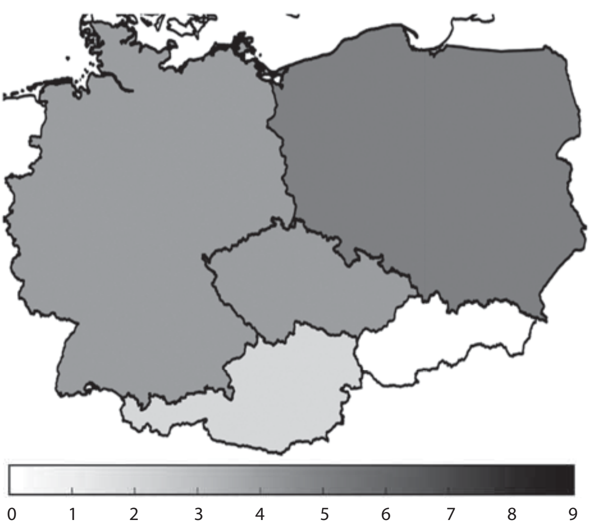
Table 4 Summary of international CLIs structures

		reference country					
		AUT	CZE	DEU	POL	SVK	Total
input data	AUT	3	2	1	2	4	12
input data	CZE	0	4	1	1	1	7
input data	DEU	9	4	6	4	5	28
input data	POL	3	5	6	6	5	25
input data	SVK	0	0	1	2	0	3
	Total	15	15	15	15	15	x

Source: Data from OECD (2017), processed by CIF

Data from Table 4 are visualized as choropleth maps to ease the interpretation of the results. Figure 6 shows the choropleth map of leading influences of the selected countries on the Czech business cycle. The darker the shade of area in the map, the higher number of its economic indicators appeared in the constructed CLI. For the maps of the rest of the analyzed countries, see Figure 7.

Figure 6 Visualization of the leading influences of neighboring countries on the Czech Republic business cycle. The darker the shade, the greater the influence measured by the number of individual economic series included in the international CLI.

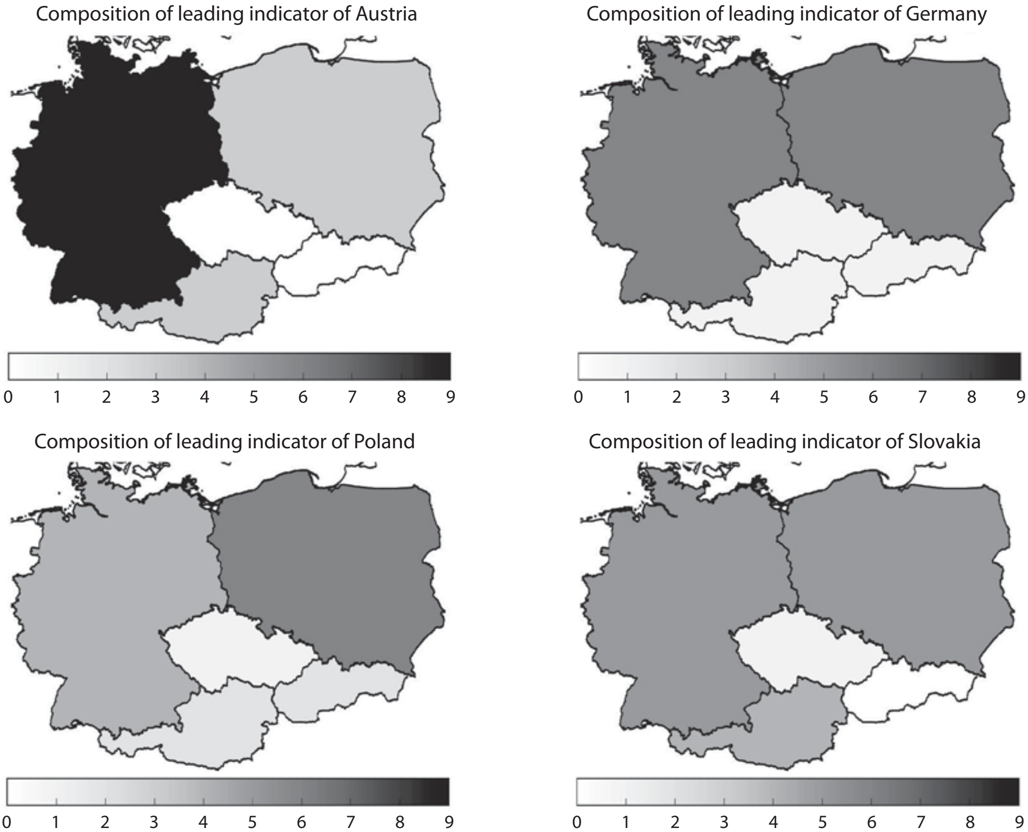


Source: Data from OECD (2017), processed by CIF

Germany and Poland are the most leading economies according to the appearance of their economic indicators in CLIs (28 and 25 times, respectively). The CLIs of these two countries also contain the highest ratio of their own national indicators (6 out of 15). The other extreme is Slovakia, whose CLI contains only foreign indicators (almost exclusively German, Polish and Austria ones).

The leading role of Poland could explain, why the Polish international CLI does not show up any improvement when compared with OECD CLI (as described in section 4.2). This is, however, in contradiction with the German CLI, which tends to perform better when international data are incorporated.

Figure 7 Visualizations of the leading influences between the business cycles of selected countries. The darker the shade, the greater the leading influence measured by the number of individual economic series included in the international CLI.



Source: Data from OECD (2017), processed by CIF

The influence of Germany on the Czech economy is not surprising as it is the key business partner of the Czech Republic. Tables 5 and 6 show values of Czech imports and exports in 2016 and Germany is number one in both. The role of Poland, which provides most of the leading individual indicators, is more surprising: Poland is of course in the top positions among the Czech import and export countries, but its values are only a fraction of the German ones. On the other hand, Slovakia occurs on the top ranks of Czech imports and exports as well, but not even one of its economic indicators appeared in Czech CLI.

Table 5 Neighboring countries by imports into the Czech Republic in 2016

country	rank	import value (thousands of CZK)	import ratio (%)
DEU	1	924 082 513	26.40
POL	3	288 884 681	8.30
SVK	4	177 637 683	5.10
AUT	7	101 370 620	2.90

Source: Czech Statistical Office (2017)

Table 6 Neighboring countries by exports from the Czech Republic in 2016

country	rank	export value (thousands of CZK)	export ratio (%)
DEU	1	1 286 717 667	32.40
SVK	2	331 354 077	8.30
POL	3	229 138 114	5.80
AUT	7	168 445 174	4.20

Source: Czech Statistical Office (2017)

Austria business cycle is led mainly by German indicators, which form more than half of its CLI. It displays no signs of the influence of Czech or Slovak economies.

CONCLUSION

This paper presented the OECD methodology of composite indicators construction and how to modify it by considering international input data. Three objectives were defined: (1) to assess whether the international data would change the structure of OECD CLIs, (2) whether the international CLIs could be used to analyze the relationships between business cycles of several countries and (3) how could these relationships be interpreted and visualized.

The OECD methodology was described in section 3 and followed during the rest of this paper with two modifications: the preselection phase included data from multiple countries and the number of selected component series was fixed to enable the comparison of international influences.

Section 4.2 presented the newly constructed international CLIs of Austria, the Czech Republic, Germany, Poland and Slovakia. Their performances were compared to the national CLIs published by OECD. All the CLIs, except the Polish one, tended to improve their leading performance after the international data were added. This confirmed, that the individual economic series from one country can contribute to predicting the business cycle movements of another country. The first of the three objectives of this paper was therefore met.

Section 4.3 discussed the structure of the international CLIs and analyzed the leading or lagging behavior of each country's business cycle. Germany and Poland were recognized as the most leading economies, Slovakia as the most lagging one. The choropleth maps were designed to visualize and easily interpret the leading influences between the analyzed countries. This section therefore gave the answers to the other two research questions. This was also the first time, to the best of my knowledge, when the structure of international CLI was visualized on maps and used to interpret the relationships between the countries.

Moreover, it was shown, on the example of the Czech Republic, that the leading influences revealed by this analysis were not driven solely by the country's international trade. Therefore, the modified CLI could serve as another indicator of international relationships.

The described approach could be extended to analyze the available data from countries all around the world and it could for example help to create clusters of regions with similar business cycle movements.

None of these analyses would be possible in any of the publicly available software programs. Therefore, a new computational framework was proposed and introduced for the very first time in section 2 of this paper. This framework is now available as an open-source project on GitHub platform and it will soon be available as Python library installable via pip command. Researchers from now on will not have to waste their time on deploying basic tasks, e.g., how to detect the turning points or evaluate and aggregate the series. They will be allowed (and encouraged) to download the new framework and start to collaborate on its future development.

ACKNOWLEDGEMENTS

This paper was written with the support of the Czech Science Foundation project No. P402/12/G097 DYME – Dynamic Models in Economics.

References

ASTOLFI, R., GAMBA, M., GUIDETTI, E., PIONNIER, P.-A. The use of short-term indicators and survey data for predicting turning points in economic activity: A performance analysis of the OECD system of CLIs during the Great Recession. *OECD Statistics Working Papers*, 2016/07. DOI: http://dx.doi.org/10.1787/5jlz4gs2pkhf-en.

BRUNO, G. AND OTRANTO, E. Dating the Italian business cycle: a comparison of procedures. *Istituto di Studi e Analisi Economica, Working Paper*, 2004, 41, 25.

BRY, G. AND BOSCHAN, C. Cyclical analysis of time series: selected procedures and computer programs. *National Bureau of Economic Research*, 1971, 216.

BURNS, A. F. AND MITCHELL, W. C. *Measuring Business Cycles* [online]. NBER, 1946. Retrieved from: <http://papers.nber.org/books/burn46-1>.

EUROSTAT (EUROPEAN COMMISSION). *Handbook on cyclical composite indicators for business cycle analysis.* Luxembourg: Publications Office of the European Union, 2017. DOI: http://dx.doi.org/10.2785/962890.

GALLEGATI, M. Making leading indicators more leading: A wavelet-based method for the construction of composite leading indexes. *OECD Journal: Journal of Business Cycle Measurement and Analysis*, 2014(1), pp. 1–22.

GYOMAI, G. AND GUIDETTI, E. OECD System of Composite Leading Indicators. *OECD Publishing*, 2012, 18.

HAMILTON, J. D. A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica*, 1989, 57(2), pp. 357–384.

HIJMANS, R. et al. *Global Administrative Areas, version 2.8* [online]. 2015. Retrieved from: <http://gadm.org>.

LEVANON, G. Evaluating and comparing leading and coincident economic indicators. *Business Economics*, 2010, 45(1), pp. 16–27.

MINTZ, I. *Dating Postwar Business Cycles: Methods and Their Application to Western Germany, 1950–1967.* New York: National Bureau of Economic Research, 1969.

NILSSON, R. AND GYOMAI, G. Cycle extraction: A comparison of the Phase-Average Trend method, the Hodrick-Prescott and Christiano-Fitzgerald filters. *OECD Publishing*, 2011, 23.

OECD. *OECD Cyclical Analysis and Composite Indicators System (CACIS) User's Guide* [online]. 2010. Retrieved from: <https://community.oecd.org/community/cacis/overview>.

OZYILDIRIM, A., SCHAITKIN, B., ZARNOWITZ, V. Business cycles in the euro area defined with coincident economic indicators and predicted with leading economic indicators. *Journal of Forecasting*, 2010, 29(1–2), pp. 6–28.

RAVN, M. AND UHLIG, H. On adjusting the Hodrick-Prescott filter for the frequency of observations. *Review of Economics and Statistics*, 2002, 84(2), pp. 371–376.

SVATOŇ, P. *Composite Leading Indicators: A Contribution to the Analysis of the Czech Business Cycle.* Prague: Ministry of Finance of the Czech Republic, 2011.

THE CONFERENCE BOARD. *Business Cycle Indicators Handbook* [online]. New York, 2001. Retrieved from: <https://www.conference-board.org/publications/publicationdetail.cfm?publicationid=852>.

VRANÁ, L. Alternative to the Construction of Czech Composite Indicators. *International Days of Statistics and Economics,* Slaný: Melandrium, 2013, pp. 1502–1511.

ZARNOWITZ, V. AND OZYILDIRIM, A. Time series decomposition and measurement of business cycles, trends and growth cycles. *Journal of Monetary Economics*, 2006, 53(7), pp. 1717–1739.

ZHOU, P. AND ANG, B. W. Comparing MCDA Aggregation Methods in Constructing Composite Indicators Using the Shannon-Spearman Measure. *Social Indicators Research*, 2009, 94(1), pp. 83–96. DOI: https://doi.org/10.1007/s11205-008-9338-0.

APPENDIX

Table A1 Structure of international CLIs (indicators are sorted according to the strength of their relationship with the reference series)

	country	indicator code	indicator full name	note
AUT	DEU	BRBUFT02	Business tendency surveys (retail trade) > Business situation - Activity > Future tendency > National indicator	
AUT	DEU	BCOBLV02	Business tendency surveys (construction) > Order books > Level > National indicator	
AUT	AUT	PIEAFD01	Producer Prices Index > Economic activities > Manufacture of food products > Total	inverted
AUT	DEU	BREMFT02	Business tendency surveys (retail trade) > Employment > Future tendency > National indicator	
AUT	DEU	BCSPFT02	Business tendency surveys (construction) > Selling prices > Future tendency > National indicator	
AUT	POL	BVCICP02	Business tendency surveys (services) > Confidence Indicators > Composite Indicators > National indicator	
AUT	DEU	BVEMFT02	Business tendency surveys (services) > Employment > Future tendency > National indicator	
AUT	DEU	BRCICP02	Business tendency surveys (retail trade) > Confidence indicators > Composite indicators > National indicator	
AUT	DEU	PITGCD01	Producer Prices Index > Type of goods > Durable consumer goods > Total	inverted
AUT	DEU	BRBUTE02	Business tendency surveys (retail trade) > Business situation - Activity > Tendency > National indicator	
AUT	POL	BSPRTE02	Business tendency surveys (manufacturing) > Production > Tendency > National indicator	
AUT	POL	SPASTT01	Share Prices > All shares/broad > Total > Total	
AUT	AUT	PITGCG01	Producer Prices Index > Type of goods > Consumer goods > Total	inverted
AUT	DEU	BVDETE02	Business tendency surveys (services) > Demand evolution > Tendency > National indicator	
AUT	AUT	PITGND01	Producer Prices Index > Type of goods > Non durable consumer goods > Total	inverted

Source: Own construction

Table A2 Structure of international CLIs (indicators are sorted according to the strength of their relationship with the reference series)

	country	indicator code	indicator full name	note
CZE	POL	SPASTT01	Share Prices > All shares/broad > Total > Total	
	CZE	SPASTT01	Share Prices > All shares/broad > Total > Total	
	POL	BVCICP02	Business tendency surveys (services) > Confidence Indicators > Composite Indicators > National indicator	
	CZE	BSPRTE02	Business tendency surveys (manufacturing) > Production > Tendency > National indicator	
	CZE	BSPRFT02	Business tendency surveys (manufacturing) > Production > Future Tendency > National indicator	
	DEU	SPASTT01	Share Prices > All shares/broad > Total > Total	
	POL	BRCICP02	Business tendency surveys (retail trade) > Confidence indicators > Composite indicators > National indicator	
	CZE	XTIMVA01	International Trade > Imports > Value (goods) > Total	
	DEU	BRCICP02	Business tendency surveys (retail trade) > Confidence indicators > Composite indicators > National indicator	
	DEU	PRMNCG03	Production > Manufacturing > Consumer goods > Non durable goods	
	AUT	BRBUFT02	Business tendency surveys (retail trade) > Business situation - Activity > Future tendency > National indicator	
	DEU	SLMNCN01	Sales > Manufacturing > Consumer goods non durable > Volume	
	AUT	PRMNIG01	Production > Manufacturing > Intermediate goods > Total	
	POL	CSCICP02	Consumer opinion surveys > Confidence indicators > Composite indicators > National indicator	
	POL	BVDEFT02	Business tendency surveys (services) > Demand evolution > Future tendency > National indicator	

Source: Own construction

Table A3 Structure of international CLIs (indicators are sorted according to the strength of their relationship with the reference series)

	country	indicator code	indicator full name	note
DEU	POL	SPASTT01	Share Prices > All shares/broad > Total > Total	
	DEU	PRMNCG03	Production > Manufacturing > Consumer goods > Non durable goods	
	DEU	LRHUADMA	Labour Force Survey - quarterly rates > Harmonised unemployment - monthly rates > Aged 25 and over > Males	
	POL	BVCICP02	Business tendency surveys (services) > Confidence Indicators > Composite Indicators > National indicator	
	POL	CSCICP02	Consumer opinion surveys > Confidence indicators > Composite indicators > National indicator	
	SVK	BRVSLV02	Business tendency surveys (retail trade) > Volume of stocks > Level > National indicator	inverted
	POL	BVDEFT02	Business tendency surveys (services) > Demand evolution > Future tendency > National indicator	
	CZE	SPASTT01	Share Prices > All shares/broad > Total > Total	
	DEU	LFHUADTT	Labour Force Survey - quarterly levels > Harmonised unemployment - monthly levels > Aged 25 and over > All persons	
	DEU	LRHUADFE	Labour Force Survey - quarterly rates > Harmonised unemployment - monthly rates > Aged 25 and over > Females	
	DEU	BRCICP02	Business tendency surveys (retail trade) > Confidence indicators > Composite indicators > National indicator	
	POL	PITGND02	Producer Prices Index > Type of goods > Non durable consumer goods > Domestic	inverted
	DEU	BRBUFT02	Business tendency surveys (retail trade) > Business situation - Activity > Future tendency > National indicator	
	AUT	PIEAFD01	Producer Prices Index > Economic activities > Manufacture of food products > Total	inverted
	POL	BSPRTE02	Business tendency surveys (manufacturing) > Production > Tendency > National indicator	

Source: Own construction

Table A4 Structure of international CLIs (indicators are sorted according to the strength of their relationship with the reference series)

	country	indicator code	indicator full name	note
POL	POL	BCBUTE02	Business tendency surveys (construction) > Business situation - Activity > Tendency > National indicator	
	POL	SPASTT01	Share Prices > All shares/broad > Total > Total	
	SVK	CPGDFD02	Consumer Price Index > Goods > Food > Food (excl restaurants)	inverted
	DEU	BVDETE02	Business tendency surveys (services) > Demand evolution > Tendency > National indicator	
	POL	BSPRTE02	Business tendency surveys (manufacturing) > Production > Tendency > National indicator	
	SVK	PIEAFD02	Producer Prices Index > Economic activities > Manufacture of food products > Domestic	inverted
	AUT	PITGCD02	Producer Prices Index > Type of goods > Durable consumer goods > Domestic	inverted
	DEU	BRBUFT02	Business tendency surveys (retail trade) > Business situation - Activity > Future tendency > National indicator	
	DEU	BRCICP02	Business tendency surveys (retail trade) > Confidence indicators > Composite indicators > National indicator	
	POL	XTIMVA01	International Trade > Imports > Value (goods) > Total	
	CZE	BVCICP02	Business tendency surveys (services) > Confidence Indicators > Composite Indicators > National indicator	
	DEU	BREMFT02	Business tendency surveys (retail trade) > Employment > Future tendency > National indicator	
	POL	BVBUTE02	Business tendency surveys (services) > Business situation - Activity > Tendency > National indicator	
	POL	BVCICP02	Business tendency surveys (services) > Confidence Indicators > Composite Indicators > National indicator	
	AUT	CPALTT01	Consumer Price Index > All items > Total > Total	inverted

Source: Own construction

Table A5 Structure of international CLIs (indicators are sorted according to the strength of their relationship with the reference series)

	country	indicator code	indicator full name	note
SVK	POL	BSEMFT02	Business tendency surveys (manufacturing) > Employment > Future Tendency > National indicator	
	DEU	CSCICP02	Consumer opinion surveys > Confidence indicators > Composite indicators > National indicator	
	POL	BRBUTE02	Business tendency surveys (retail trade) > Business situation - Activity > Tendency > National indicator	
	POL	BSPRFT02	Business tendency surveys (manufacturing) > Production > Future Tendency > National indicator	
	AUT	BSPRFT02	Business tendency surveys (manufacturing) > Production > Future Tendency > National indicator	
	DEU	BSPRTE02	Business tendency surveys (manufacturing) > Production > Tendency > National indicator	
	AUT	BSEMFT02	Business tendency surveys (manufacturing) > Employment > Future Tendency > National indicator	
	DEU	BSBUCT02	Business tendency surveys (manufacturing) > Business situation > Current > National indicator	
	POL	BRBUFT02	Business tendency surveys (retail trade) > Business situation - Activity > Future tendency > National indicator	
	AUT	PRMNCG01	Production > Manufacturing > Consumer goods > Total	
	POL	BVCICP02	Business tendency surveys (services) > Confidence Indicators > Composite Indicators > National indicator	
	DEU	BSFGLV02	Business tendency surveys (manufacturing) > Finished goods stocks > Level > National indicator	inverted
	AUT	XTNTVA01	International Trade > Net trade > Value (goods) > Total	
	DEU	BSOBLV02	Business tendency surveys (manufacturing) > Order books > Level > National indicator	
	CZE	PRINTO01	Production > Industry > Total industry > Total industry excluding construction	

Source: Own construction

Risk Substance of Newly Established Businesses

Gabriela Chmelíková [1] | *Mendel University in Brno, Czech Republic*
Kristina Somerlíková [2] | *Mendel University in Brno, Czech Republic*

Abstract

This paper investigates the risk substance of newly established business in the Czech Republic. As traditional methods for risk measurement in this case come across the lack of reliable data from capital markets, lack of any financial history for start-up companies and low level of diversification of investors, we draw on the theory of business risk and empirically investigate the role that risk fundamentals may play in the overall riskiness of start-up firms in the Czech Republic. Our findings confirm that the return fluctuations of start-up companies are statistically significantly associated with the operating and financial leverage.[3]

Keywords	***JEL code***
Risk measurement, start-up business, degree of operating leverage, degree of financial leverage	*M21, G12*

INTRODUCTION

The capital markets are an important source of data for a number of techniques used to quantify risk. The efficiency of the capital market is then the same for the quality of the information contained in data coming from the capital markets and therefore can restrict the range of techniques that can be used to quantify the discount rates. Capital market efficiency at all levels has become a subject for a number of empirical studies, e.g. Cross (1973), French (1980), Gibbons and Hess (1981), Bonin and Moses (1974), Abeyratana et al. (1996) and Al-Deehani (2003). Their conclusions can be summarised as follows: developed markets tend towards a weak form of market efficiency, while developing markets can be seen as mostly inefficient at all levels. Methods for quantifying risk based purely on technical analysis can therefore only be used in economic systems which have the efficient price mechanisms of capital markets in its weak form. Companies operating in other than developed conditions must therefore rely on data from the developed economies when assessing risk or on other quantification techniques.

The capital markets in the Czech Republic are usually rated as developing capital markets, c.f. FTSE (2017), Hull and McGroarty (2014). This is largely because of their short history compared to that of the capital markets of traditional market economies. This is also true in the scope of transactions done

[1] Department of Regional and Business Economics, Faculty of Regional Development and International Studies, Mendel University in Brno, Czech Republic. Corresponding author: e-mail: gabriela.chmelikova@mendelu.cz, phone: (+420)731624694.

[2] Department of Applied Statistics, Faculty of Regional Development and International Studies, Mendel University in Brno, Czech Republic.

[3] Paper following this article will be published in the next Statistika issue No. 3/2018.

with capital resources, since the share of capital acquired through trading in capital markets (with regards other long-term sources of financing) is smaller in comparison with Anglo-Saxon capital markets. This largely restricts the options of existing approaches in a big way.

The nature of newly established businesses lets imply two important aspects when evaluating suitable risk measurement techniques. The first specific feature of a start-up company is the lack of history. This aspect does not allow for the use of usual methods based on a regression analysis of data on the company`s performance and performance of the market over time. This is a principle which is specific to the CAPM (Capital Asset Pricing Model), but also to its mutation on an accounts basis. This restriction means that alternative approaches should be used by looking for analogical companies or by identifying risk factors using multiple regression.

Another important restriction is the character of companies' capital resources. Analyses of capital structure of start-up companies in the Czech Republic (see Chmelíková and Somerlíková, 2014) have shown that the most prevalent source of financing is of an internal character. The term internal is used to refer to financial resources which the entrepreneur, his family members or employees invested or lent in the company. According to Damodarana (2009), we can assume about the stakeholders that their investment into the start-up company is their only or at least their dominant personal investment. This investor is likely to invest a considerable portion of his property in the company and it is unlikely that this share is part of a well-diversified portfolio. Therefore, we can see a low diversification of these investors` personal capital and the impossibility of using traditional methods based primarily on systematic risk evaluation.

The below mentioned facts resulting from the character of the emerging capital markets as well as the character of start-up companies restrict the use of methods for quantifying risk:

- Lack of reliable data from capital markets,
- Lack of any financial history for start-up companies,
- Low level of diversification of investors.

The above-mentioned limits represent a serious restriction on using common methods for setting discount rates for start-up companies in the transitive economies and can be an important barrier to their development. New born businesses are said to provide the thrust for economic growth, which is actually supported by the statistics for national economies (Horell and Litan, 2010). Fast-growing economies usually have a higher number of start-up companies than stagnating ones. Knowing the correct discount rate will encourage the owners to establish new businesses and thereby significantly increase the economic growth. The aim of this paper is therefore to identify the substance of riskiness of newly established firms operating in the conditions of developing capital markets, which could help in designing a suitable technique for the discount rate estimation.

This paper is organized as follows: After the introduction, in the following section we develop our hypothesis, describe the data including the descriptive statistics and present the empirical study. We conclude with a discussion on main results.

1 BUSINESS RISK

1.1 Hypothesis, data, methods

Business risk is a function of uncertainty connected with the future earnings from doing business (Galasyuk and Galasyuk, 2007). The rate of required return of used capital should then match the fluctuation of earnings on the investment. Traditionally, these earnings are measured via the invested capital profitability indicator, whose future volatility is, according to Brayman (2012), a risk indicator of the intended investment. However, using this for risk quantification conflicts with the character of this variable's calculation. To identify future earnings volatility requires a knowledge of how these future profits develop, which requires admitting a certain degree of inaccuracy set by the forecast of these qualities. The validity of the result of estimating the fluctuation in return based on future earnings analysis is naturally deformed by the inaccuracy of estimating these future earnings. This estimation is based on a high-quality sales forecast, which requires not only an analysis and forecast of the company's market, but also an analysis of the inner potential and rival strength of the company itself. Quantifying these categories allows for a certain amount of subjectivity to solve the problem, however, which limits the exactness of the sales forecast and the subsequent estimation of future return and its fluctuation.

Nevertheless, when it comes to applying this procedure to the quality of risk quantification techniques assessment, this deficiency is suppressed. A common way of measuring the fluctuation of a random variable is dispersion, from which a standard deviation is derived. The random variable is represented here by the expected earnings from investment into own capital. We suggest to measure returns on equity using the indicator *Free Cash Flow to Equity (FCFE)* as the most appropriate profitability indicator (Kislingerová, 2010). A question to what degree is this criterion an objective scale arises here. The answer may be found in comparing the fluctuation figures for this profitability measure with the mortality rate of new companies. The survival time of newly established companies is linked to the risk of a given investment in the new company. One may assume that the lower the percentage of surviving companies in a number of newly-established firms after a specific amount of time has elapsed since their establishment, the higher the risk of the given industry. The survival percentage on a number of newly-established firms determines the probability of decline for companies in individual industries. Therefore, it is a suitable scale for checking that the right accounting productivity fluctuation criterion for risk measurement has been selected.

To verify this conjecture, the following hypothesis may be formulated:

H_1: Probability of decline across individual economic sectors is related to the average fluctuation of the rate of return on equity for companies in these sectors.

For the purposes of testing and eventually supporting the presented hypothesis, its zero alternative is formulated as follows:

H_0: Probability of decline across individual economic sectors is unrelated to the average fluctuation of the rate of return on equity for companies in these sectors.

The probability of decline for individual industries can be identified thanks to the Eurostat database "Business Demography Project" (European Comission – Eurostat, 2014). Since this database publishes the figures for company lifespan until 2007 and the electronic financial statements of Czech companies that made them public are accessible only after 2004 inclusive, only the figures of both variables in 2005, 2006 and 2007 may be used for mutual comparison. This reduction of the time period does not, however, lower the quality of the studied sample, since it captures the phase of a company's lifespan that is the subject of this study (initial, start-up phase).

The analysis of firm specific variables is based on the data published by Bisnode in the corporate database Albertina – Gold Edition (Bisnode Czech Republic, 2012). There were 6 581companies established in 2004 in the Czech Republic that also published their financial statements. In 2004, almost 90 000 economic subjects were established in the Czech Republic. Therefore, the sample of 6 581 presents only a small part of them. A significant reduction in the sample of companies available for testing fluctuation in the profitability also shown by the fact that only 3 507 of them reached their third year. Nevertheless, the size of this sample is sufficient for testing the stated hypothesis. Given the fact that survival is monitored as an average for individual sections according to NACE classification, it is also necessary to express accounting return fluctuation in average values for the given sections of economic activity. As has already been stated, the degree of fluctuation in return of individual companies is characterised by the standard deviation, whose absolute level is influenced by not only the dispersion of the observed quality, but also by the level of the mean value of a given random variable. Because of the mutual comparability of the observed companies and the ability to characterise the average fluctuation of a whole industry, for every company the standard deviation has been relativized by conversion to a coefficient of variation in accordance with the following equation:

Coefficient of variation of FCFE for firm *i*:

$$Coefficient\ of\ variation\ of\ FCFE_i = \frac{\sigma_i}{\mu_i}, \tag{1}$$

where σ_i stands for standard deviation of financial return of a firm *i* in the 4 – years time after inception and μ_i represents mean of this variable for the firm *i*.

The average of coefficients of variation for individual sections, according to NACE classification, was discovered from data on the accounting return for companies established in 2004 that made their financial statements of 2005, 2006 and 2007 public. In contrast to Chmelíková (2014), the weighted average of coefficients of variation for individual sections was used. The weights for particular companies were calculated according to the following formula:

$$w_i = \frac{Total\ Assets_i}{Total\ Assets\ in\ the\ sector}, \tag{2}$$

where *Total Assets*$_i$ stands for total assets of firm *i* in the year of inception and *Total Assets in the sector* represents the sum of *Total assets* of all firms in the respective NACE sector.

The weighted average values of variation coefficients calculated this way are shown in Table 1.

Table 1 Financial fluctuation in return for newly-established companies in the Czech Republic and the probability of decline for new companies.

NACE code	Description	Probability of business´s death within 3 years	Number of business´s births in 2004	Number of newly born in 2004 survived the first 3 years of life	Weighted average coefficient of variation of FCFE
1	Crop and animal production, hunting and related service activities	N/A	76	35	0.257
2	Forestry and logging	N/A	27	13	0.458
3	Fishing and aquaculture	N/A	N/A	N/A	N/A
5	Mining of coal and lignite	33%	4	1	0.357
6	Extraction of crude petroleum and natural gas	50%	1	1	0.502
7	Mining of metal ores	52%	N/A	N/A	N/A
8	Other mining and quarrying	N/A	6	4	0.415
9	Mining support service activities	N/A	4	2	0.123
10	Manufacture of food products	43%	44	21	0.435

(continuation)

NACE code	Description	Probability of business´s death within 3 years	Number of business´s births in 2004	Number of newly born in 2004 survived the first 3 years of life	Weighted average coefficient of variation of FCFE
11	Manufacture of beverages	41%	14	6	0.339
12	Manufacture of tobacco products	40%	N/A	N/A	N/A
13	Manufacture of textiles	54%	20	11	0.537
14	Manufacture of wearing apparel	37%	20	10	0.299
15	Manufacture of leather and related products	N/A	5	4	0.218
16	Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials	38%	75	39	0.319
17	Manufacture of paper and paper products	37%	1	1	0.236
18	Printing and reproduction of recorded media	N/A	39	17	0.421
19	Manufacture of coke and refined petroleum products	0%	N/A	N/A	N/A
20	Manufacture of chemicals and chemical products	27%	14	7	0.216
21	Manufacture of basic pharmaceutical products and pharmaceutical preparations	N/A	1	0	N/A
22	Manufacture of rubber and plastic products	42%	47	31	0.540
23	Manufacture of other non-metallic mineral products	46%	34	20	0.487
24	Manufacture of basic metals	40%	8	8	0.346
25	Manufacture of fabricated metal products, except machinery and equipment	30%	196	107	0.200
26	Manufacture of computer, electronic and optical products	29%	33	20	0.193
27	Manufacture of electrical equipment	N/A	58	30	0.212
28	Manufacture of machinery and equipment n.e.c.	N/A	74	46	0.328

(continuation)

NACE code	Description	Probability of business´s death within 3 years	Number of business´s births in 2004	Number of newly born in 2004 survived the first 3 years of life	Weighted average coefficient of variation of FCFE
29	Manufacture of motor vehicles, trailers and semi-trailers	45%	14	9	0.455
30	Manufacture of other transport equipment	N/A	6	2	0.362
31	Manufacture of furniture	N/A	18	14	0.421
32	Other manufacturing	45%	31	16	0.403
33	Repair and installation of machinery and equipment	x	14	11	0.246
35	Electricity, gas, steam and air conditioning supply	17%	14	10	0,18
36	Water collection, treatment and supply	N/A	4	1	0.201
37	Sewerage	N/A	4	2	0.453
38	Waste collection, treatment and disposal activities; materials recovery	N/A	31	14	0.247
39	Remediation activities and other waste management services	41%	3	1	0.311
41	Construction of buildings	45%	242	119	0.356
42	Civil engineering	N/A	18	12	0.178
43	Specialised construction activities	N/A	271	140	0.147
45	Wholesale and retail trade and repair of motor vehicles and motorcycles	33%	143	98	0.214
46	Wholesale trade, except of motor vehicles and motorcycles	31%	1033	561	0.201
47	Retail trade, except of motor vehicles and motorcycles	48%	580	299	0.423
49	Land transport and transport via pipelines	35%	170	91	0.252
50	Water transport	33%	1	1	0.245

(continuation)

NACE code	Description	Probability of business´s death within 3 years	Number of business´s births in 2004	Number of newly born in 2004 survived the first 3 years of life	Weighted average coefficient of variation of FCFE
51	Air transport	N/A	1	1	0.207
52	Warehousing and support activities for transportation	45%	52	34	0.442
53	Postal and courier activities	44%	2	1	0.343
55	Accommodation	40%	59	37	0.272
56	Food and beverage service activities	48%	200	98	0.421
58	Publishing activities	N/A	51	26	0.252
59	Motion picture, video and television programme production, sound recording and music publishing activities	N/A	N/A	6	0.154
60	Programming and broadcasting activities	N/A	N/A	N/A	N/A
61	Telecommunications	38%	16	9	0.333
62	Computer programming, consultancy and related activities	40%	100	47	0.373
63	Information service activities	37%	24	12	0.233
64	Financial service activities, except insurance and pension funding	51%	24	8	0.526
65	Insurance, reinsurance and pension funding, except compulsory social security	N/A	N/A	N/A	N/A
66	Activities auxiliary to financial services and insurance activities	58%	14	5	0.443
68	Real estate activities	47%	1546	630	0.411
69	Legal and accounting activities	54%	134	75	0.427
70	Activities of head offices; management consultancy activities	N/A	180	132	0.254
71	Architectural and engineering activities; technical testing and analysis	N/A	156	122	0.014

(continuation)

NACE code	Description	Probability of business´s death within 3 years	Number of business´s births in 2004	Number of newly born in 2004 survived the first 3 years of life	Weighted average coefficient of variation of FCFE
72	Scientific research and development	44%	10	4	0.476
73	Advertising and market research	48%	129	74	0.422
74	Other professional, scientific and technical activities	N/A	167	112	0.274
75	Veterinary activities	N/A	5	3	0.254
77	Rental and leasing activities	N/A	38	26	0.453
78	Employment activities	48%	24	15	0.290
79	Travel agency, tour operator and other reservation service and related activities	N/A	34	27	0.187
80	Security and investigation activities	33%	24	17	0,19
81	Services to buildings and landscape activities	N/A	17	10	0.254
82	Office administrative, office support and other business support activities	N/A	27	10	0.132
84	Public administration and defence; compulsory social security	40%	N/A	N/A	N/A
85	Education	N/A	71	32	0.189
86	Human health activities	45%	46	27	N/A
87	Residential care activities	23%	N/A	N/A	N/A
88	Social work activities without accommodation	N/A	1	N/A	N/A
90	Creative, arts and entertainment activities	N/A	7	5	0.125
91	Libraries, archives, museums and other cultural activities	N/A	N/A	N/A	N/A
92	Gambling and betting activities	N/A	15	10	0.098

(continuation)

NACE code	Description	Probability of business´s death within 3 years	Number of business´s births in 2004	Number of newly born in 2004 survived the first 3 years of life	Weighted average coefficient of variation of FCFE
93	Sports activities and amusement and recreation activities	N/A	64	35	0.161
94	Activities of membership organisations	35%	9	N/A	N/A
95	Repair of computers and personal and household goods	39%	16	11	0.302
96	Other personal service activities	43%	20	11	0.404
97	Activities of households as employers of domestic personnel	N/A	N/A	N/A	N/A
98	Undifferentiated goods- and services-producing activities of private households for own use	N/A	N/A	N/A	N/A
99	Activities of extraterritorial organisations and bodies	N/A	N/A	N/A	N/A
	Total		**6681**	**3507**	

Note: N/A – not available data.
Source: Own calculation based on data from Eurostat (Business Demography Project) and Albertina

1.2 Results

The method of regression analysis was used to analyse the relationship between financial fluctuation in return-on-investment and the probability of decline. Individual sets of data were first subjected to normality verification by the Kolmogorov-Smirnov test as well as on the basis of a normal probability plot, and then came a regression analysis of the following two variables:

– weighted average coefficient of variation of *FCFE* companies established in 2004 as independent variable and
– dependant variable probabilities of decline within 3 years.

All variables, including a description of the measures used and their descriptive statistics, are summarized in Table 2.

Table 2 Variable description and summary statistics

Variable	Abbreviation	Mean	SD	Min	Max	N
Dependent Variable						
Probability of Decline	*PoD*	40.9024	8.1511	17.0000	58.0000	41
Independent Variable						
Weighted average variation Coefficient of Free Cash Flow to Equity	*VCoFCEF*	0.3501	0.1063	0.1770	0.5400	41

Source: Eurostat (Business Demography Project) and Albertina

The resultant regression line is $y' = 18.95 + 62.71 * x$, the coefficient of correlation reaches the value $r = 0.67$ and points to a significant dependence of decline on the weighted average of *FCFE* coefficient of variation. The resultant value of the coefficient of correlation is lower in comparison to the version with simple average of FCFE coefficient of variation (see Chmelíková, 2014), however, still points to a significant relationship between the observed variables and so disproves the hypothesis H_0. So, this conclusion supports the hypothesis H_1, and thus also the assumption about the suitability of choosing accounting fluctuation in *FCFE* an objective scale for the determining the risk substance.

The hypothesis about the relationship between the observed variables was statistically tested on the significance level α = 0.001. The summary results of statistical analysis are presented in Table 3.

Table 3 Statistical hypothesis H_0 test on the significance level $\alpha = 0.001$

Independent Variable *VCoFCEF*	**Dependent Variable (Coefficients)**
Intercept	**18.9471** ***
	(0.0000)
VCoFCEF	**62.7078*** **
	(0.0000)
R^2	0.6682
F-test	78.5540
p-value	0.0000
	< 0.001

Note: Standard errors in parentheses ***p<0.001.
Source: Own calculations (processed in software Unistat)

The null hypothesis of the independence assumption is rejected on the basis of statistical significance (p-value is less than the given significance level α = 0.001). We can hence support the base hypothesis: Probability of decline across individual economic sectors is related to the average fluctuation of the rate of return on equity for companies in these sectors, which explain 83% of the decline probability. This finding is in accordance with the theoretical prediction that future volatility of profitability indicator indicates the total riskiness of intended investment (Brayman, 2012). This result supports the idea that fluctuation in return to equity to owners is convenient predictor of future financial distress and hence can serve an objective tool for risk substance identification.

2 RISK SUBSTANCE

2.1 Hypothesis, data, methods

Business risk is partly independent of the pressures of the cost structure with its fixed elements. If the share of fixed costs is high, even a small fall in demand can lead to a large fall in profitability (cf. Toms and Nguyen, 2005). Therefore, it can be said that a higher share of fixed costs leads to a higher business risk. A higher share of fixed costs is usually typical for companies with highly-automated processes, for firms with highly-qualified staff (who need to be paid even in a recession) and for companies who have invested into research and development in the past and whose fixed costs therefore include the depreciation of the R&D.

If the share of fixed costs of the overall costs is high, the company is seen as having high operating leverage. As in physics and finance, leverage means bigger impact using less energy. Here, the high level of leverage means that a relatively small change in the turnover of the company produces a relatively large change in the profitability of all the capital invested and vice-versa.

From the point of view of the owners the operating risk is not the only risk connected with their investment. An extra risk on their investment comes from using debt and with fixed payments to creditors connected to this. The owners facing a certain level of operating risk are exposed to a greater risk on their investment by being involved in the debt. If the firm is exclusively financed from own sources, the risk run by the owners is the portion of the operating risk for the whole company. However, they are properly rewarded for meeting certain conditions for this high risk in the form of higher return on their investment. This effect is known in the jargon of company finance as financial leverage. Financial leverage only has a positive effect on the profitability of own capital in conditions where the return on total assets exceeds interest rates paid for using debt.

According to Toms (2012) the fluctuation of future revenue is closely linked to the level of fixed payments in the company`s cash flow. The higher the level of fixed liabilities (whether in the form of past investment, contracts with suppliers or creditors), the lower the ability of the company to react both to changes in demand (real and nominal) and also to changes in the level of business costs (again real and nominal). The research question of this paper was therefore formulated as to what extend the volatility of returns of start-up companies is caused by the risk fundamentals – operating and financial leverages.

In fact, these connections have become the subject of a number of research papers into the relationship to total risk measures, however, the research of the influence of the two basic components on shareholder risk measured for markets has been limited. Nevertheless, it has shown that the influence is roughly balanced (Mandelker and Rhee, 1984; Li and Henderson, 1991; Toms and Nguyen, 2005).

To research these conjectures, we develop the following null hypotheses (and their alternatives):

- H_{02}: *The return fluctuations of start-up companies are not associated with the operating and financial leverage,*
- H_2: *The return fluctuations of start-up companies are associated with the operating and financial leverage.*

The hypothesis is formulated in line with expectations stemming from the conclusions of Toms and Nguyen (2005), who provide clear evidence that basic risk fundamentals are connected to the increased fluctuations of returns.

The method of multiple regression analysis and statistical hypotheses testing is used to analyse the relationship between return fluctuations of start-up companies and risk fundamentals – financial and operating leverage. Individual sets of data are first subjected to normality verification by the Kolmogorov-Smirnov test as well as on the basis of a normal probability plot. Then comes a multiple regression analysis of the following variables:

- Degree of Operating *(DOL)* and Financial Leverage (*DF*L) by a start-up firm as independent variables (which enables to monitor the intensity of basic risk fundamentals in the newly established companies in the Czech Republic),
- Free Cash Flow to Equity (*FCFE*) variation coefficient[4] by a start-up firm as dependent variable (which enables to describe the level of riskiness in the companies after their inception). We use *FCFE* indicator, as it is the most convenient measure of financial return for the owners, that is long-term sustainable (Brealey et al., 2012).

The degree of operating leverage and the degree of financial leverage are calculated according to following formulas:

[4] The degree of fluctuation in return of individual companies is characterised by the standard deviation, whose absolute level is influenced by not only the dispersion of the observed variable, but also by the level of the mean value of a given random variable. Because of the mutual comparability of the observed companies the standard deviation has been relativized by conversion to a coefficient of variation.

$$DOL = \frac{Earnings\ before\ interest_{(t+1)} / Earnings\ before\ interest_{(t)}}{Sales_{(t+1)} / Sales_{(t)}}, \tag{3}$$

$$DFL = \frac{Net\ profit_{(t+1)} / Net\ profit_{(t)}}{Earnings\ before\ interest_{(t+1)} / Earnings\ before\ interest_{(t)}}, \tag{4}$$

where *(t)* is a year of inception of a firm and (t+1) is one year after inception of a firm.

Coefficient of variation of free cash-flow to equity is calculated according to following formula:

$$Variation\ Coefficient_{\ FCFE_{(t)}\ to\ FCFE_{(t+5)}} = \frac{\sigma_{FCFE_{(t)}\ to\ FCFE_{(t+5)}}}{\mu_{FCFE_{(t)}\ to\ FCFE_{(t+5)}}}, \tag{5}$$

where $\sigma_{FCFE_{(t)}\ to\ FCFE_{(t+5)}}$ stands for standard deviation of financial return of a start-up firm for 5 year period after inception and $\mu_{FCFE_{(t)}\ to\ FCFE_{(t+5)}}$ represents mean of this variable.

On the basis of data from Czech start-up companies the assumption about positive relationship between the operating and financial leverage is verified with the use of statistical hypothesis testing. The software Unistat is used for calculations. The null hypothesis is rejected or accepted on the basis of statistical significance (the significance level $\alpha = 0.05$).

The analysis is based on the data published by Amadeus – the trans-European database compiled by Bureau van Dijk Electronic Publishing. The dataset covers the period from 2008 to 2015 and consists of start-up firms in the Czech Republic. In the research sample we deployed those firms that were born in the period from 2008 to 2011 and survived and published their financial statements for five years after inception.[5] The total number of firms in the sample is 11 371. All variables, including a description of the measures used and their descriptive statistics, are summarized in Table 4.

Table 4 Variable description and summary statistics

Variable	Abbreviation	Mean	SD	Min	Max	N
Dependent Variable						
Free Cash Flow to Equity	*FCEF*	0.2472	0.0560	0.0270	0.5962	11 371
Independent Variable						
Degree of Operating Leverage	*DOL*	2.3066	1.5312	0.0000	10.6051	11 371
Degree of Financial Leverage	*DFL*	2.0739	1.4167	0.0000	9.7860	11 371

Note: Standard errors in parentheses ***p<0.001.
Source: Own calculations (processed in software Unistat)

2.2 Results

The hypothesis H_{02} has been tested in order to determine the relationship between the degree of financial and operating leverages and fluctuations in the returns. The equation $y' = 0.2022 + 0.0068\ x_1 + 0.0141\ x_2$ was calculated to determine the influence of the two variables *DOL* (x_1) and *DFL* (x_2) on the *FCF*E variation

[5] Firms are obliged to publish their financial statements when they reach given limits on assets, turnover or number of employees. The exact rules are given by Law n. 563/1991 Sb. § 20. The sample doesn´t include only those firms that did not meet their legal duties of publishing their financials.

coefficient variable as a percentage. The rising nature of function shows that fluctuation in the earnings is positively dependent on the extent of fixed costs in the operating and financial cost structure. This is in line with expectations stemming from the empirical evidence and theoretical principles of corporate finance (Brealey et al., 2012). The correlation coefficient $r = 0.5111$ shows a relatively high connection. The results of multiple regression analysis have shown a significant close relationship between financial leverage and a less-distinctive risk connection with operating leverage. The combination of both the risk fundamentals analysed therefore explains the fluctuation in free cash flow for owners of more than 26%.

The hypothesis about the relationship between the observed variables was statistically tested on the significance level α = 0.001. The summary results of statistical analysis are presented in Table 5.

Table 5 Statistical hypothesis H_{02} test on the significance level α = 0.001

Independent Variable	Dependent Variable (Coefficients)
Intercept	**0.2022***
	(0.0000)
DOL	**0.0068***
	(0.0000)
DFL	**0.0141***
	(0.0000)
R^2	0.2613
F-test	2009.9104
p-value	0.0000
	< 0.001

Note: Standard errors in parentheses ***p<0.001.
Source: Own calculations (processed in software Unistat)

The null hypothesis of the independence assumption is rejected on the basis of statistical significance (p-value is less than the given significance level α = 0.001). We can hence support the base hypothesis: The return fluctuations of start-up companies are associated with the operating and financial leverage, which explain 26% of the movements. This finding is in accordance with the majority of findings from empirical studies devoting to influence of risk fundamentals to the total riskiness (e.g. Mandelker and Rhee, 1984; Li and Henderson, 1991; Toms and Nguyen, 2005). This finding is very useful when looking for appropriate risk indicator for start-up companies as it may help to overcome the limits set by lack of reliable data from capital markets, lack of any financial history for start-up companies and low level of diversification of investors.

CONCLUSIONS

The main challenge for researchers and policy makers with regard to small enterprises is to support institutional frameworks that enable to unlock the potential of start-up companies in the economy. The academics may help to overcome the barriers in the decision making process of potential investors by exploring the techniques for riskiness evaluation.

In our study we therefore investigate the relationship between risk fundamentals and total riskiness of newly born firms in the Czech Republic. For description of the overall riskiness of Czech start-ups we use variation coefficient of return to owners, which occurred to be statistically positively associated with probability of decline and hence a good measure of riskiness. Generally, our empirical results support our hypothesis that the riskiness is strongly determined by the burdening by fixed costs.

What is very important and will allow a new approach to quantifying risk is identification of the dependence of the fluctuation of free cash flow on the combined level of risk – operating and financial leverage. This could have been expected intuitively since both forms of leverage are among the primary determinants of company risk. The factors for the fluctuation in future earnings can actually be divided into two groups. On the one hand, there are factors which affect the level of future profits, such as the level of demand and the development of input prices. On the other hand, there is the company`s ability to adapt to these changes. The ability of the company to adapt to exogenous changes is then determined by the amount it is burdened by fixed payments (operating and also financial from the owners` point of view). The business risk is therefore partly dependent on the burden of the cost structure with its fixed elements. If there is a high level of fixed costs, even a small fall in demand can cause of large drop in the return-on-investment.

By verifying the dependence of risk on the burdening of the cost structure with fixed elements, it is possible to suggest that a model should be constructed to quantify the discount rate for start-up companies in the conditions of an economy with emerging capital markets. Since new companies play an important role in the national economy, this makes it a useful tool enriching the theory which can be used in practice in real-life decision-making.

ACKOWLEDGMENT

This work was supported by the IGA Mendelu under Grant No. FRRMS – IGA – 2018/006 "Role mikrofinančního sektoru v rozvoji rurálních oblastí".

References

ABEYRATNA, G., LONIE, A. A., POWER, D. M., SINCLAIR, C. D. The stock market reaction to dividend announcements: a UK study of complex market signals. *Journal of Economic Studies*, 1996, 23(1), pp. 32–52.

AL-DEEHANI AND MANEGOLD, T. The Association Between Market-Determined and Accounting-Determined Measures of Systematic Risk: Some Further Evidence. J*ournal of Financial and Quantitative Analysis*, 1975(10), pp. 213–284.

BISNODE CZECH REPUBLIC. *Databáze Albertina – Gold Edition. Bisnode* [DVD]. Prague: Bisnode, 2012.

BONIN, J. M. AND MOSES, E. A. Seasonal Variations in Prices of Individual Dow Jones Industrial Stocks. *Journal of Financial and Quantitative Analysis*, 1974, 9(6), pp. 963–991.

BRAYMAN, S. Defining and Measuring Risk Capacity. *Financial Services Review*, 2012, 21(2), pp. 131–48.

BREALEY, R. A., MYERS, S. C., ALLEN, F., MOHANTY, P. *Principles of corporate finance*. Tata McGraw-Hill Education, 2012.

CHMELÍKOVÁ, G. AND SOMERLÍKOVÁ, K. Capital structure in start-up firms in the conditions of the Czech economy. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis*, 2014, 62(2), pp. 363–372.

CHMELÍKOVÁ, G. *Riziko v hodnocení investic do začínajících společností v České republice*. Mendel University in Brno, 2014, 154 p.

CROSS, F. The Behavior of Stock Prices on Friday and Monday. *Financial Analysts Journal*, 1973, 29(6), pp. 67–69.

DAMODARAN, A. *Valuing Young, Start-up and Growth Companies: Estimation Issues and Valuation Changes*. Stern School of Business, New York Universtiy, Working paper, 2009.

EUROPEAN COMISSION − EUROSTAT. *Structural Business Statistics* [online]. Brussels: European Comission. [cit. 28.3.2014]. <http://epp.eurostat.ec.europa.eu/portal/page/portal/european_business/special_sbs_topics/business_demography>.

FTSE. *FTSE Global Equity Index Series Country Classification* [online]. 2017. [cit. 28.10.2017]. <https://www.ftserussell.com/sites/default/files/ftse-country-classification-process---sept-2017_0.pdf>.

GALASYUK, V. AND GALASYUK, V. Consideration of Economic Risks in a Valuation Practice: Journey from the Kingdom of Tradition to the Kingdom of Common Sense. *Working paper, Ukrainian Society of Financial Analysts*, 2007.

GIBBONS, M. R. AND HESS, P. Day of the week effects and asset returns. *Journal of business*, 1981, pp. 579–596.

HORELL, M. AND LITAN, R. *After Inception: How Enduring is Job Creation by Startups?* [online]. 2010. <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1657121>.

HULL, M. AND MCGROARTY, F. Do emerging markets become more efficient as they develop? Long memory persistence in equity indices. *Emerging Markets Review*, 2014, 18, pp. 45–61.

KISLINGEROVÁ, E. Manažerské finance. 3rd Ed. Prague: CH Beck, 2010, xxxviii, 811 p. ISBN 978-80-7400-194-9

LI, R.-J. AND HENDERSON, G. Combined Leverage and Stock Risk. *Quarterly Journal of Business and Economics*, 1991, Vol. 30, No. 1, pp.18–40.

MANDELKER, G. N. AND RHEE, S. G. The Impact of the Degrees of Operating and Financial Leverage on Systematic Risk of Common Stock. *Journal of Financial and Quantitative Analysis,* 1984, 19(1), pp. 45–57.

TOMS, S. *Accounting based risk measurement: An alternative to CAPM derived discount factors*. University of York, The York Management School, Working Paper No. 68, 2012.

TOMS, S. S. AND NGUYEN, D. T. *The association between Accounting and Market-Based Risk Measures*. Working Paper, Department of Management Studies, University of York, 2005.

Control Charts for Processes with an Inherent Between-Sample Variation

Eva Jarošová [1] | *ŠKODA AUTO University, Mladá Boleslav, Czech Republic*

Abstract

A number of processes to which statistical control is applied are subject to various effects that cause random changes in the mean value. The removal of these fluctuations is either technologically impossible or economically disadvantageous under current conditions. The frequent occurrence of signals in the Shewhart chart due to these fluctuations is then undesirable and therefore the conventional control limits need to be extended. Several approaches to the design of the control charts with extended limits are presented in the paper and applied on the data from a real production process. The methods assume samples of size greater than 1. The performance of the charts is examined using the operating characteristic and average run length. The study reveals that in many cases, reducing the risk of false alarms is insufficient.

Keywords

Statistical process control, modified chart, acceptance chart, variance component chart, average run length

JEL code

C44, C83, L15

INTRODUCTION

Control charts, the main tool of statistical process control (SPC), have been used for more than 80 years. The idea behind control charts is to separate the variation due to assignable causes from the random variation that is inherent to a process. Apart from the Shewhart chart introduced in 1931, the CUSUM chart based on cumulative sums or the EWMA chart using exponentially weighted moving averages belong to the best known ones. All these charts are based on the assumption that unless some special causes exist in a process, its parameters are constant.

The extensive study of real production processes in Germany performed by Kaiser et Nowack (2000; cf. Michálek, 2001) revealed that only 2% of processes met the assumption of constant parameters. Within the Six Sigma approach, the goal is no longer to maintain a constant mean value; it is allowed to move around the target as long as the process output conforms to specification. This applies to processes in which the variation within a sample taken from the process is very small as compared with the allowable variation given by the specification limits. When samples are taken from a process, the differences between their averages are greater than would correspond to the within-sample variation on which the conventional Shewhart control limits are based, which results in frequent alarms. Bringing such

[1] Na Karmeli 1457, 293 01 Mladá Boleslav, Czech Republic. E-mail: eva.jarosova@savs.cz, phone: (+420)730803162.

processes to the state when the process mean is constant could be technically or economically impossible and, therefore, common changes of the process mean are considered a part of the inherent process variation.

To allow for the process mean's changes that are still acceptable, the control limits have to be extended. Several methods have been introduced in the literature but the information about the properties of the resulting charts is missing. The aim of the paper is to examine and discuss the effectiveness of some selected control charts. The attention is paid to $\overline{X}$ -charts. As with the conventional $\overline{X}$ chart, the within-sample variation is monitored using the *R*-chart.

1 OVERVIEW OF CONTROL CHARTS WITH EXTENDED LIMITS

Limiting to cases when samples of size greater than 1 are taken from a process, two main approaches can be distinguished. The first approach uses specification limits *USL* and *LSL*, the other is based on the inherent process variability.

The charts based on the specification limits were introduced and discussed long ago (Rissik, 1943; Hill, 1956; Freund, 1957). The centre line of the conventional control chart is replaced by bounds μ_U and μ_L for the true process mean and the usual 3-sigma limits are drawn outwards from the interval (μ_L, μ_U). The resulting charts are called modified or acceptance control charts and they differ in how the interval limits are determined. These charts were presented by many authors without any criticism (Duncan, 1986; Montgomery, 2009; Mitra, 2008; Wadsworth et al., 2002). They are also included in the standard ISO 7870-3:2012. On the other hand, Bissell (1994) and Wheeler (2004) remark that this approach is contrary to the philosophy of continuous improvement because there is no incentive for reducing variability.

The charts based on the inherent process variation appear to be less referred to in the literature. None of them is mentioned in the books listed above, with the exception of Bissell (1994), who introduces some of the methods from Section 4. Methods of constructing control limits can be divided into two groups.

The first group includes methods in which the standard deviation representing the inherent variation is estimated using sample averages. Cryer et Ryan (1990) advocate the use of the overall standard error, Wheeler et Chambers (1986), Woodall et Thomas (1995), Laubscher (1996) and Bissell (1994) use moving ranges (or their squares) of sample averages.

The methods in the second group are based on the ANOVA model with random effects and variance components, which represent the within-sample and between-sample variability. The variance component chart (Laubscher, 1996; Woodall et Thomas, 1995; Wetherill et Brown, 1991) employs 3-sigma limits; the standard deviation of sample averages is derived using the ANOVA model. Dietrich et Schulze (2010) suggest an approach that is similar to the one using the specification limits, however, the bounds for the mean are based on the between-sample variance component.

2 PROCESS CAPABILITY

As mentioned above, the extended control limits are generally used in situations where the within-sample variation is considerably smaller than the allowable range *USL – LSL*. Such processes are called highly capable. Process capability reflects the ability of a process where no assignable causes are present to function in such a manner that its output, represented by a quality characteristic distribution, lies almost completely within specification limits *USL* and *LSL*. The concept of process capability was introduced in the '80s (Sullivan, 1984; Kane, 1986). The most common indices C_p and C_{pk} are given by the formulas (see e.g. Kotz et Johnson, 2002):

$$C_p = \frac{USL - LSL}{6\sigma} \qquad C_{pk} = \min(C_{pU}, C_{pL}), \tag{1}$$

where

$$C_{pU} = \frac{USL - \mu}{3\sigma} \qquad C_{pL} = \frac{\mu - LSL}{3\sigma} \;. \tag{2}$$

It is assumed that the process output is normally distributed with constant μ and σ over time, where σ is the measure of the within-sample variability. The assumption of the constant parameters is verified by control charts. While the C_p index measures only the process ability to meet the specification limits and its construction assumes μ as the midpoint of these limits (usually a target), C_{pk} accounts for the real process location. The difference between C_{pk} and C_p represents the potential improvement to be attained by centering the process. The generally accepted minimum value for C_p is 1.33. In the Six Sigma methodology, C_p of 2 is considered the aim of a process improvement; even if the mean of such process shifts by 1.5σ from the midpoint, no serious problems arise since the expected fraction nonconforming is as low as 3.4 ppm (C_{pk} equals 1.5 in this case). Referring to the shift of 1.5 σ relates to the performance of the conventional Shewhart chart – the shifts smaller than 1.5σ may be detected quite late by this chart.

It should be noted that the overall performance of a process with an inherent between-sample variation is evaluated using the performance indices P_p and P_{pk} recommended by the Automotive Industry Action Group (AIAG, 2005). These indices take into account the total process variation. However, the construction of the charts in Section 4.2 is based on the short-time process behaviour and therefore the capability index C_{pk} is considered when the charts are designed.

3 PERFORMANCE OF CONTROL CHARTS

The performance of control charts is evaluated using an operating characteristic (OC). The OC curve describes the relationship between the probability β of not detecting a shift from the reference value μ_0 to $\mu = \mu_0 + k\sigma$ on the first subsequent sample. Considering the conventional $\overline{X}$-chart (control chart for averages) we can write:

$$\beta = P(LCL \le \bar{x} \le UCL \mid \mu = \mu_0 + k\sigma) \,, \tag{3}$$

where the magnitude of the shift is expressed in k-multiples of σ. Usually the normal distribution $\mathrm{N}(\mu, \sigma^2/n)$ of sample averages with known parameters is assumed. When evaluating the performance of charts with extended control limits, the magnitude of a shift will be expressed in k-multiples of total standard deviation σ_x and the distribution $\mathrm{N}(\mu, \sigma_{\bar{x}}^2)$ of sample averages will be considered.

In SPC the average run length (ARL) is widely used. It is the expected value of the number of samples taken until the first point exceeds a control limit. If the values of a plotted characteristic can be considered independent, run lengths have the geometric distribution $\mathrm{G}(1-\beta)$ and

$$ARL = \frac{1}{1-\beta} \;. \tag{4}$$

In a certain control chart, ARL depends on the shift magnitude $k\sigma$ ($k\sigma_x$). The average run length $ARL(0)$ for $k = 0$ is an important characteristic; it should be as large as possible since frequent false alarms may lead to overcontrol, which results in a larger variation of the process output, or at least they discourage operators. $ARL(0)$ of the conventional $\overline{X}$-chart is 370.4, which means that the false alarm (type I error) can be expected after 370 samples on average. Conversely, ARL for a given $k > 0$ should be as small as possible so that the shift of $k\sigma$ ($k\sigma_x$) can be detected quickly.

4 CONTROL CHARTS BASED ON SPECIFICATION LIMITS

The bounds μ_U and μ_L for the true process mean are based on the specified proportion of units exceeding limits *USL* and *LSL*. Formulas for calculation of control limits assume that the distribution from which the sample comes is normal with the current value of μ and variance σ^2, i.e. only the within-sample variation is taken into account.

4.1 Modified and acceptance control charts

The aim of the modified control chart is to determine whether the process mean is within interval (μ_L, μ_U) such that the fraction nonconforming does not exceed the chosen value p_A. The bounds for the process mean are given by the formulas:

$$\mu_U = USL - u_{1-p_A}\sigma \qquad \mu_L = LSL + u_{1-p_A}\sigma\,, \tag{5}$$

where σ denotes the within-sample standard deviation, which is often estimated using the average of sample ranges, $\hat{\sigma} = \overline{R}/d_2$. Values of d_2 can be found in ISO 7870-2:2013 or any book dealing with Shewhart control charts. The control limits are drawn outwards from the interval (μ_L , μ_U) and are positioned at:

$$UCL = USL - \left(u_{1-p_A} - \frac{u_{1-\alpha}}{\sqrt{n}}\right)\hat{\sigma} \qquad LCL = LSL + \left(u_{1-p_A} - \frac{u_{1-\alpha}}{\sqrt{n}}\right)\hat{\sigma}\ , \tag{6}$$

where α is the type I error risk.

As Hill (1956, p. 16) points out, the control limits determined by (6) "accept a sample mean nearer to the tolerance when there is less information, than when there is more information", which is an undesirable feature. He suggests that the other possibility should be used; the bounds for the mean are:

$$\mu_U = USL - u_{1-p_R}\sigma \qquad \mu_L = LSL + u_{1-p_R}\sigma\,, \tag{7}$$

and for process fraction nonconforming p_R to be rejected with probability 1 – β the control limits are:

$$UCL = USL - u_{1-p_R}\hat{\sigma} - \frac{u_{1-\beta}}{\sqrt{n}}\hat{\sigma} \qquad LCL = LSL + u_{1-p_R}\hat{\sigma} + \frac{u_{1-\beta}}{\sqrt{n}}\hat{\sigma}\,. \tag{8}$$

In this case the control limits lie within the interval (μ_L, μ_U) and they are nearer to the specification limits for larger sample sizes. The latter chart is sometimes classified as a variant of the acceptance chart (Montgomery, 2013; Mitra, 2008).

The acceptance control chart (Freund, 1957) is based on both risks α and β related to p_A and p_R and therefore the sample size must meet the condition:

$$n = \frac{u_{1-\alpha} + u_{1-\beta}}{u_{1-p_A} - u_{1-p_R}}\,, \tag{9}$$

which follows from equating either the upper or the lower control limits in (6) and (8).

4.2 Choice of parameters

In the following considerations only a predetermined sample size n is assumed.

The choice of p_A, p_R, α and β or directly percentiles u_{1-p_A}, u_{1-p_R}, $u_{1-\alpha}$ and $u_{1-\beta}$ will affect the chart performance. The values $u_{1-\alpha} = 3$ and $u_{1-\beta} = 1.65$ corresponding to the risks of 0.00135 and 0.05 are mostly used and will be applied here, too. The choice of u_{1-p_A}, u_{1-p_R} requires some attention.

So that the modified limits (6) are wider than the conventional 3-sigma limits, the following inequality must apply:

$$USL - LSL > 2u_{1-p_A}. \tag{10}$$

The choice of p_A can be based on the value of process capability index C_{pk}. Using relations (1) and (2) and the minimum acceptable value 1.33 of C_{pk}, we get $u_{1-p_A} = 4$ and $USL - LSL > 8\sigma$.

Some authors use $u_{1-p_A} = 3$ or $u_{1-p_A} = 4.5$ (Jarošová et Noskievičová, 2015, p. 81).

As for u_{1-p_R}, the control limits (8) are wider than the conventional 3-sigma limits if the condition

$$USL - LSL > 2\left(u_{1-p_R} + \frac{4.65}{\sqrt{n}}\right)\sigma \tag{11}$$

is met. For $u_{1-p_R} = 2.33$ corresponding to the fraction nonconforming of 0.01 (see e.g. Montgomery, 2013, p. 442) and $n = 5$ we get $USL - LSL > 9\sigma$, approximately.

It should be emphasized that the requirement of wider control limits itself does not guarantee the expected properties of a control chart, namely a sufficiently high value of $ARL(0)$ and a reasonably low value of ARL for a shift that should be detected. ARL depends on the variability of sample averages, which is affected both by within-sample and by between-sample variation.

5 CONTROL CHARTS BASED ON THE INHERENT VARIABILITY

The between-sample variation as a part of the inherent variation of a process is taken into account when constructing control limits. In most cases 3-sigma limits are used like in the Shewhart chart, i.e.:

$$UCL = \bar{\bar{x}} + 3\sigma_{\bar{x}} \qquad LCL = \bar{\bar{x}} - 3\sigma_{\bar{x}}, \tag{12}$$

where $\bar{\bar{x}}$ denotes the total average and $\sigma_{\bar{x}}$ the standard deviation of sample averages. Two main approaches to estimate $\sigma_{\bar{x}}$ exist: the first approach is based on sample averages, the second approach uses the ANOVA model and variance components.

Table 1 Estimation of $\sigma_{\bar{x}}$ using sample averages

	Standard error estimate	References
Overall standard error	$\frac{s_{\bar{x}}}{c_4} = \frac{1}{c_4}\sqrt{\frac{\sum_{j=1}^{m}(\bar{x}_j - \bar{\bar{x}})^2}{m-1}}$	Cryer et Ryan (1990)
Average moving range	$\frac{\overline{MR}}{d_2} = \frac{1}{d_2}\frac{\sum_{j=2}^{m}\lvert \bar{x}_j - \bar{x}_{j-1}\rvert}{m-1}$	Wheeler et Chambers (1986), Woodall et Thomas (1995)
Median moving range	$\frac{MeR}{d_4} = \frac{1}{d_4} Me\lvert \bar{x}_j - \bar{x}_{j-1}\rvert$	Laubscher (1996)
Square root of MSSD	$\frac{1}{c_4}\sqrt{\frac{1}{2}MSSD} = \frac{1}{c_4}\sqrt{\frac{1}{2}\frac{\sum_{j=2}^{m}(\bar{x}_j - \bar{x}_{j-1})^2}{m-1}}$	Bissell (1994)

Source: Own construction

5.1 Charts based on sample averages

Several methods of estimating $\sigma_{\bar{x}}$ together with references are listed in Table 1, where m is the number of samples, $\bar{x}_j$ (j = 1, 2, ..., m) are sample averages and $\bar{\bar{x}}$ is the total average. Unbiasing constants c_4, d_2, and d_4 can be found for example in Wheeler (2004, p. 416), c_4 and d_2 are also available in ISO 7870-2:2013. Constant c_4 depends on the number of samples m (differently from its use in the conventional Shewhart charts, where the sample size is cardinal), constants d_2 a d_4 relate to the use of differences between two adjacent observations and therefore correspond to the "sample size" of 2. Consequently, $d_2 = 1.128$ and $d_4 = 0.9539$ are always used in these calculations. Woodall et Montgomery (2000) examined the bias of various estimates when a shift in the mean is present and concluded that the estimates based on moving ranges are preferable. Moreover, the use of the median can reduce or possibly eliminate the bias.

5.2 Charts based on variance components

Variance component chart (Laubscher, 1996; Woodall et Thomas, 1995; Wetherill et Brown, 1991) is based on the model:

$$x_{ij} = \mu_0 + a_j + \varepsilon_{ij}, \tag{13}$$

where μ_0 denotes the grand process mean, $a_j \sim \mathrm{N}(0, \sigma_A^2)$ is the random effect of sample j, and $\varepsilon_{ij} \sim \mathrm{N}(0, \sigma^2)$ represents the within-sample variation. Under the assumption that a_j and ε_{ij} are independent, we can write:

$$\sigma_x^2 = \sigma_A^2 + \sigma^2, \tag{14}$$

and

$$\sigma_{\bar{x}}^2 = \sigma_A^2 + \frac{\sigma^2}{n}. \tag{15}$$

Control limits are given by (12).

ANOVA is used to estimate variance components σ_A^2 and σ^2:

$$\hat{\sigma}_A^2 = \frac{MSA - MSE}{n} \qquad \hat{\sigma}^2 = MSE, \tag{16}$$

where

$$MSA = \frac{n\sum_{j=1}^{m}(\bar{x}_j - \bar{\bar{x}})^2}{(m-1)} \qquad MSE = \frac{\sum_{j=1}^{m}\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2}{m(n-1)}, \tag{17}$$

measure the between-sample and within-sample variability, respectively.

Another approach is recommended by Dietrich et Schulze (2010). It is also based on the random-effect model (13) but the control limits are constructed differently: the bounds for the process mean are distant by ±Δ from the centre line and the common 3-sigma limits are drawn outwards from them:

$$UCL = \bar{\bar{x}} + \Delta + \frac{3\hat{\sigma}}{\sqrt{n}} \qquad LCL = \bar{\bar{x}} - \Delta - \frac{3\hat{\sigma}}{\sqrt{n}}. \tag{18}$$

The authors suggest to choose $\Delta = 1.5\,\hat{\sigma}_A$, where $\hat{\sigma}_A^2$ and $\hat{\sigma}^2$ are given by (16).

6 EXAMPLE

The process in which steel frames are moulded has specification limits USL = 35.1 mm and LSL = 34.9 mm. Samples of size 5 are taken from the process and the $\overline{X}$-chart and R-chart with centre lines determined by $\overline{\overline{x}}$ = 35.0645 and $\overline{R}$ = 0.008 are drawn (Figure 1). The estimated within-sample standard deviation $\hat{\sigma} = \overline{R}/d_2 = 0.0035$ results in $\hat{C}_p = 9.46$, indicating a highly capable process. Although more than half of the points lie outside the conventional control limits UCL_S and LCL_S, most of them are not considered to signal an assignable cause and therefore the control limits should be extended. All the methods described above were used and the resulting limits together with their distance are shown in Table 2. Three columns on the right contain the values of standard deviations used in the calculations. The variance components from Section 5.2 are displayed in the ANOVA table (Table 3).

Table 2 Extended control limits

	LCL	*UCL*	*UCL – LCL*	$\hat{\sigma}_{\bar{x}}$	$\hat{\sigma}$	$\hat{\sigma}_A$
Cryer	35.0386	35.0905	0.0519	0.0086	-	-
Wheeler	35.0451	35.0840	0.0389	0.0065	-	-
Laubscher	35.0485	35.0806	0.0321	0.0053	-	-
Bissell	35.0440	35.0850	0.0410	0.0068	-	-
Woodall	35.0389	35.0902	0.0513	0.0086	0.0036	0.0084
Dietrich	35.0471	35.0820	0.0348	-	0.0036	0.0084
Modif.	34.9094	35.0906	0.1813	-	0.0035	-
Accept.	34.9108	35.0892	0.1784	-	0.0035	-

Source: Own construction

Table 3 ANOVA, Variance component analysis

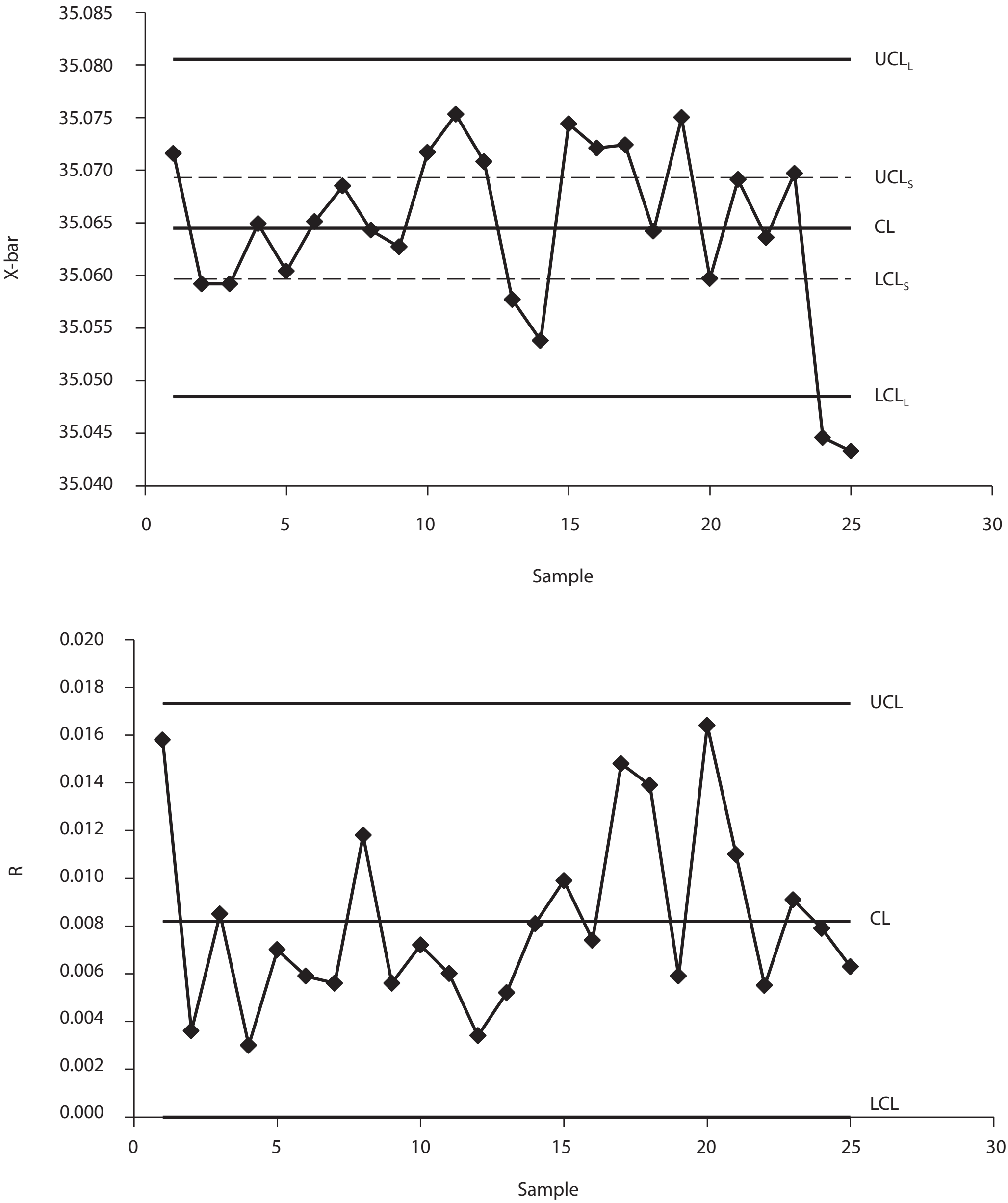
Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value	Var. Comp.
Between groups	0.008785	24	0.000366	28.46	0.0000	0.000071
Within groups	0.001286	100	0.000013			0.000013
Total (Corr.)	0.010071	124				

Source: Statgraphics

Different estimates of $\hat{\sigma}_{\bar{x}}$ obtained by various methods are due to the apparent shift before the last two samples (Figure 1). It should be noted that after the retrospective analysis, such points are usually omitted and the control limits revised, in which case the differences between the various constructions would be much smaller. The narrowest pair of the extended control limits (UCL_L and LCL_L), i.e. the one obtained using the median moving range according to Laubscher (1996), is drawn in Figure 1.

The distance of the modified and acceptance limits is several times larger, thus confirming the criticism of Bissell (1994) and Wheeler (2004).

Figure 1 $\overline{X}$ and R control chart



Source: Own construction

7 COMPARISON OF SELECTED CHARTS

To compare the performance of control charts, only the case with known parameters σ_A and σ is considered. The performance of the charts based on different estimates of $\sigma_{\bar{x}}^2$ is not examined – it would require carrying out some simulations.

To calculate the OC curves, the sample size of 5 and the within-sample $\sigma = 1$ were chosen. Several scenarios defined by the different ratio of σ_A/σ were applied. Mean's shifts were expressed as multiples of σ_x, where $\sigma_x = \sqrt{\sigma_A^2 + \sigma^2}$. In addition, percentiles $u_{1-p_A} = 4$, $u_{1-p_R} = 2.33$ and two values of C_p, namely $C_p = 1.67$ and $C_p = 2.67$ were used to determine the modified and acceptance control limits. The control limits according to Woodall (Eq. 12) and Dietrich et Schulze (Eq. 18) for the chosen σ depend only on the ratio σ_A/σ. Values of *ARL* for the selected charts are given in Table 4 to 6. Comparing charts, we focus on the $ARL(0)$ and $ARL(1.5\sigma_x)$, when the process mean shifts by $1.5\sigma_x$ from the reference value.

Table 4 *ARL* for the modified control chart

	$ARL(k\sigma_x)$							
	$USL - LSL = 10\,\sigma$ $(C_p = 1.67)$				$USL - LSL = 16\,\sigma$ $(C_p = 2.67)$			
k	$\sigma_A = 0.5\sigma$	$\sigma_A = 1\sigma$	$\sigma_A = 1.5\sigma$	$\sigma_A = 2\sigma$	$\sigma_A = 0.5\sigma$	$\sigma_A = 1\sigma$	$\sigma_A = 1.5\sigma$	$\sigma_A = 2\sigma$
0	2 075.8	30.7	7.4	3.9	5.6E+14	9.2E+05	1554.4	109.3
0.5	253.5	14.2	5.1	3.1	2.0E+12	8.6E+04	432.6	48.9
1	29.3	5.0	2.7	2.0	6.6E+09	5.9E+03	84.1	15.4
1.5	6.2	2.4	1.7	1.4	4.3E+07	6.1E+02	21.7	6.0
2	2.3	1.5	1.3	1.2	5.5E+05	9.2E+01	7.5	3.0

Source: Own construction

Table 5 *ARL* for the acceptance control chart

	$ARL(k\sigma_x)$							
	$USL - LSL = 10\,\sigma$ $(C_p = 1.67)$				$USL - LSL = 16\,\sigma$ $(C_p = 2.67)$			
k	$\sigma_A = 0.5\sigma$	$\sigma_A = 1\sigma$	$\sigma_A = 1.5\sigma$	$\sigma_A = 2\sigma$	$\sigma_A = 0.5\sigma$	$\sigma_A = 1\sigma$	$\sigma_A = 1.5\sigma$	$\sigma_A = 2\sigma$
0	256.0	13.0	4.6	2.9	5.3E+12	1.5E+05	619.5	62.4
0.5	49.6	7.2	3.5	2.4	2.9E+10	1.8E+04	197.1	30.5
1	9.0	3.1	2.1	1.7	1.6E+08	1.5E+03	44.1	10.6
1.5	2.9	1.8	1.4	1.3	1.7E+06	2.0E+02	13.0	4.5
2	1.5	1.3	1.2	1.1	3.5E+04	3.7E+01	5.1	2.4

Source: Own construction

The performance of both control charts based on specification limits is similar. For the process with $C_p = 1.67$, $ARL(0)$ is acceptably high only for $\sigma_A < \sigma$. Values of σ_A comparable with σ or higher result in frequent false alarms. On the other hand, $ARL(1.5\,\sigma_x)$ for the process with $C_p = 2.67$ and $\sigma_A < 2\sigma$ is unacceptably high, which means that such shifts may be detected quite late.

Due to the use of 3-sigma limits, the varcomp chart by Woodall keeps $ARL(0)$ at the same level regardless of σ_A, but *ARL* increases with σ_A (Table 6). Differently from the Shewhart chart, $ARL(1.5\,\sigma_x)$ is much longer for σ_A equal to or greater than σ. The chart by Dietrich et Schulze reveals relatively fast shifts of $1.5\,\sigma_x$ and greater regardless of σ_A, but for σ_A equal to or greater than σ, $ARL(0)$ is too small and hence the risk of false alarm is high.

Table 6 *ARL* of the control charts based on variance components

	ARL($k\sigma_x$)							
	Woodall				*Dietrich, Schulze*			
k	$\sigma_A = 0.5\sigma$	$\sigma_A = 1\sigma$	$\sigma_A = 1.5\sigma$	$\sigma_A = 2\sigma$	$\sigma_A = 0.5\sigma$	$\sigma_A = 1\sigma$	$\sigma_A = 1.5\sigma$	$\sigma_A = 2\sigma$
0	370.4	370.4	370.4	370.4	549.3	105.4	46.0	29.3
0.5	65.8	106.3	127.4	137.9	89.3	38.1	22.3	16.2
1	11.0	22.9	31.0	35.5	13.6	10.4	7.9	6.5
1.5	3.2	7.0	9.8	11.6	3.7	3.9	3.5	3.2
2	1.6	3.0	4.1	4.8	1.7	2.0	2.0	1.9

Source: Own construction

CONCLUSION

It appears that the modified and acceptance control charts generally do not perform well when the process mean fluctuates randomly. This is likely why Wheeler claims that modified limits "can never work as well as the data-based three-sigma limits" (Wheeler, 2004, p. 21) or "It will only encourage alternating periods of benign neglect and intense panic." (Wheeler, 2004, p. 346). Although the performance of these charts is influenced by the extent of the between-sample variation, this is not taken into account when the bounds for the mean are chosen.

The approach by Dietrich et Schulze (2010) is similar to the modified chart, but the bounds for the mean are derived from the between-sample variation. However, the recommended choice of 1.5 σ_A from the centre line results in small values of $ARL(0)$ for larger σ_A.

The varcomp chart (and similarly it can be said about the other charts based on sample averages) detects a mean's shift slower than the previous chart for σ_A equal to or greater than σ, but regardless of σ_A, it retains the desired value of $ARL(0)$.

The properties of the control charts were examined under the assumption of the normal distribution of sample averages with known parameters. As with the conventional Shewhart chart, the values of *ARL* will be influenced by departures from normality and by the fact that the parameters are usually estimated. The frequently chosen sample size of 5 was considered here. With exception of $ARL(0)$ with the varcomp chart, both $ARL(0)$ and *ARL* depend on the sample size. For most control charts, the increasing sample size leads, as with the Shewhart chart, to the narrower control limits and hence to lower values of *ARL*. The opposite is true with the acceptance chart; the control limits become wider and the values of *ARL* are higher when the sample size increases.

Based on the study, the charts using the specification limits and the chart according to Dietrich and Schulze (2010) are not advisable for use primarily because of the possibility of frequent false signals. The use of the two former charts could be taken into account in the statistical control of processes with a trend (Bissell, 1994; Jarošová and Noskievičová, 2015).

References

AIAG – CHRYSLER, FORD, GENERAL MOTORS. *QS-9000 – Statistical Process Control.* 2nd Ed. 2005.

BISSELL, D. *Statistical Methods for SPC and TQM*. London: Chapman & Hall, 1994.

CRYER, J. D. AND RYAN, T. P. The estimation of sigma for an X chart: $\overline{MR}/d_2$ or S/c_4? *Journal of Quality Technology*, 1990, Vol. 22, No. 3, pp. 187–192.

DIETRICH, E. AND SCHULZE, A. *Statistical Procedures for Machine and Process Qualification*. Cincinnati: Hanser, 2010.
DUNCAN, A. J. *Quality control and industrial statistics*. 5th Ed. Homewood: Irwin, 1986.
FREUND, R. A. Acceptance Control Charts. *Industrial Quality Control*, 1957, Vol. 14, No. 4, pp. 13–23.
HILL, D. Modified Control Limits. *Applied Statistics*, 1956, Vol. 5, No. 1, pp. 12–19.
ISO. *ISO 7870-2:2013 Control charts – Part 2: Shewhart control charts*. Geneva: International organization for standardization (ISO), 2013.
JAROŠOVÁ, E. AND NOSKIEVIČOVÁ, D. *Pokročilejší metody statistické regulace procesu*. Prague: Grada Publishing, 2015.
KAISER, B. AND NOWACK, H. M. W. Only an apparent lack of stability. New perspectives for process evaluation and control charting. *Partner Info Quality*, Q-DAS Publication, August 2000, No. 4.
KANE, V. E. Process Capability Indices. *Journal of Quality Technology*, 1986, Vol. 18, No. 1, pp. 41–52.
KOTZ, S. AND JOHNSON, N. L. Process capability indices: a Review, 1992–2000. *Journal of Quality Technology*, 2002, Vol. 34, No. 1, pp. 2–19.
LAUBSCHER, N. F. A Variance Components Model for Statistical Process Control. *South African Statist. J.*, 1996, 30, pp. 27–47.
MICHÁLEK, J. Nový pohled na Shewhartovy regulační diagramy. *Automa*, 2001, No. 7–8, pp. 10–12.
MITRA, A. *Fundamentals of Quality Control and Improvement*. Hoboken: John Wiley & Sons, 2008.
MONTGOMERY, D. C. *Introduction to Statistical Quality Control*. Hoboken: John Wiley & Sons, 2013.
RISSIK, H. Quality Control in Production Engineering. *Aircraft Engineering and Aerospace Technology*, 1943, Vol. 15, No. 2, pp. 55–58.
SULLIVAN, L. P. Targeting Variability – a New Approach to Quality. *Quality Progress*, 1984, Vol. 17, No. 7, pp. 15–21.
WADSWORTH, H. M., STEPHENS, K. S., GODFREY, A. B. *Modern Methods for Quality Control and Improvement*. Hoboken: John Wiley & Sons, 2002.
WETHERILL, G. B. AND BROWN, D. W. *Statistical Process Control: Theory and Practice.* New York: Chapman & Hall, 1991.
WHEELER, D. J. *Advanced Topics in Statistical Process Control (The Power of Shewhart' charts).* 2nd Ed. Knoxville, TN: SPC Press, Inc., 2004.
WHEELER, D. J. AND CHAMBERS, D. S. *Understanding statistical process control.* 2nd Ed. Knoxville, TN: SPC Press, Inc., 1992.
WOODALL, W. H. AND THOMAS, E. V. Statistical process control with several components of common cause variability. *IIE Transactions*, 1995, Vol. 27, No. 6, pp. 757–764.
WOODALL, W. H. AND MONTGOMERY, D. C. Using ranges to estimate variability. *Quality Engineering*, 2000, Vol. 13, No. 2, pp. 211–217.

Savings of the Inspection Cost in Acceptance Sampling

Jindřich Klůfa [1] | *University of Economics, Prague, Czech Republic*

Abstract

This paper refers to the rectifying sampling inspection plans with given lot tolerance percent defective (denoted LTPD). The LTPD sampling plans minimizing mean inspection cost per lot of process average quality, when the remainder of rejected lots is inspected, were originally designed by Dodge and Romig for inspection by attributes (each inspected item is classified as either good or defective). The corresponding rectifying plans for inspection by variables were created by author of this paper. Comparison of these two types of the LTPD plans from economical point of view is presented herein. Using the LTPD plans by variables we can reach fundamental savings of the inspection cost. In this paper we analysed the situations in which the rectifying LTPD plans by variables are more economical than the corresponding attribute sampling plans. A criterion for deciding if inspection by variables is to be used instead of inspection by attributes is suggested and calculated for input parameters of acceptance sampling.

Keywords	***JEL code***
Acceptance sampling, rectifying LTPD plans, inspection by variables, single quality characteristic, one specification limit	*C44, L15, C83*

INTRODUCTION

The rectifying LTPD single sampling plans for inspection by attributes are acceptance sampling plans (n, c) which minimize the mean number of items inspected per lot of process average quality:

$$I_s = N-(N-n)\cdot L(\overline{p};n,c), \tag{1}$$

under the condition:

$$L(p_t;n,c)=\beta, \tag{2}$$

where N is the number of items in the lot (the given parameter), $\overline{p}$ is the process average fraction defective (the given parameter), p_t is the lot tolerance fraction defective (the given parameter, $P_t = 100\ p_t$ is the lot tolerance per cent defective, denoted LTPD), n is the number of items in the sample ($n<N$, the search parameter), c is the acceptance number (the search parameter).

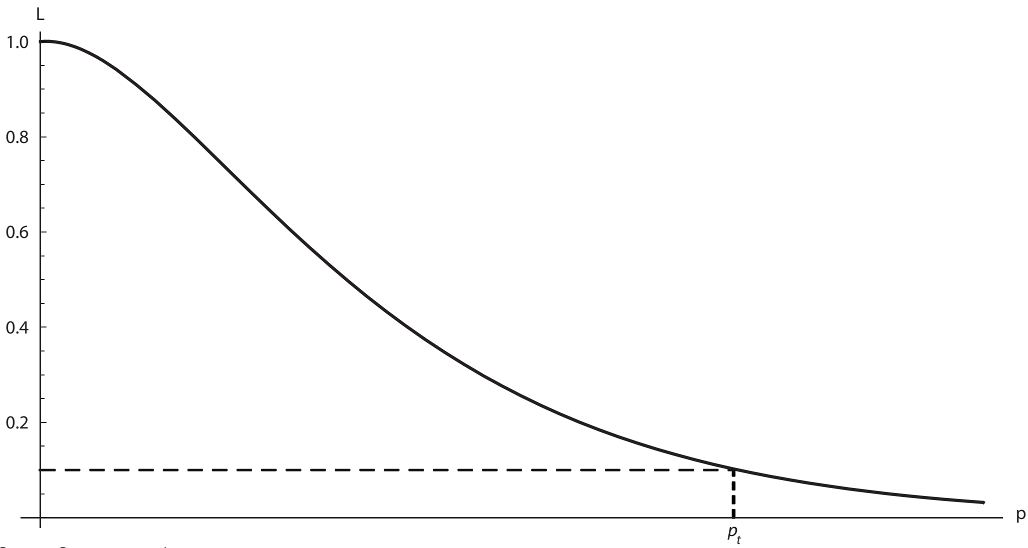
The inspection procedure: The lot is rejected when the number of defective items in the sample is greater than c (see e.g. Hald, 1981).

[1] Department of Mathematics, Faculty of Informatics and Statististics, University of Economics, W. Churchill Sq. 4, 130 67 Prague 3, Czech Republic. E-mail: klufa@vse.cz.

The function $L = L(p;n,c)$ is the operating characteristic. For given acceptance plan (n,c) the $L(p;n,c)$ is probability of accepting a submitted lot with fraction defective p – see Figure 1.

Formula (2) protects the consumer against the acceptance of a bad lot: the probability of accepting a submitted lot of tolerance quality p_t (consumer's risk) shall be β (see Figure 1). The LTPD plans for inspection by attributes are extensively tabulated in Dodge and Romig (1998), value $\beta = 0.1$ is used for consumer's risk in this book.

Figure 1 Typical graph of the operating characteristic $L = L(p)$



Source: Own construction

The rectifying LTPD plans for inspection by variables with the same protection of the consumer were created in Klůfa (1994), using for calculation of the operating characteristic L normal distribution as an approximation of the non-central t distribution. Exact LTPD plans for inspection by variables, using non-central t distribution for calculation of the operating characteristic L, were calculated in Klůfa (2010). Similar problems are solved in Chen and Chou (2001), Wang (2016), Kaspříková and Klůfa (2015), Chen and Chou (2013), Yazdi, Fallah, Shishebori, Mostafaeipour (2016), Aslam et al. (2015), Balamurali et al. (2014), Yen et al. (2014), Wang and Lo (2016), Chen (2016), Yazdi and Fallahnezhad (2017).

The dependence savings of the inspection cost (using the LTPD plan for inspection by variables instead of the corresponding LTPD plan for inspection by attributes) on input parameters of acceptance sampling is analysed in this paper. Moreover, a criterion for deciding if inspection by variables is supposed to be used instead of inspection by attributes, is suggested in present paper.

This paper follows the paper Klůfa (2015a) and the paper Klůfa (2015b) in which the combined inspection (the sample is inspected by variables, remainder of rejected lot is inspected only by attributes) is considered instead of inspection by variables which is in present paper.

1 LTPD PLANS FOR INSPECTION BY VARIABLES

In paper Klůfa (1994) the problem to find LTPD plans for inspection by variables was solved under the following assumptions:

Measurements of a single quality characteristic X are independent, identically distributed normal random variables with unknown parameters μ and σ^2. For the quality characteristic X is given either an upper specification limit U (the item is defective if its measurement exceeds U), or a lower specification limit L (the item is defective if its measurement is smaller than L). It is further assumed that the unknown parameter σ is estimated from the sample standard deviation s.

The inspection procedure is as follows. Accept the lot if:

$$\frac{U-\overline{x}}{s} \geq k\,, \quad \text{or} \quad \frac{\overline{x}-L}{s} \geq k, \tag{3}$$

where:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i\,, \quad s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2}\,. \tag{4}$$

Like Dodge and Romig we shall look for the acceptance plan (n, k) minimizing the mean number of items inspected per lot of process average quality:

$$I_m = N - \left(N-n\right) \cdot L\left(\overline{p}; n, k\right)\,, \tag{5}$$

under the condition $L(p_t;n;k) = \beta$. This condition is the same one as used for protection the consumer Dodge and Romig.

The problem of finding of the LTPD plans for inspection by variables was solved in Klůfa (1994), using for calculation of the operating characteristic L normal distribution as an approximation of the non-central t distribution (approximation of the non-central t distribution by normal distribution is based on the approximation of the distribution s/σ by normal distribution with expected value 1 and dispersion $1/(2n-2)$, see Johnson and Welch, 1940). Exact calculation of the LTPD plans for inspection by variables when the non-central t distribution is used for calculation of the operating characteristic was explained in Klůfa (2010). Now, we shall study economical aspects of these plans.

2 ECONOMICAL ASPECTS

As a measure of economic efficiency of the LTPD single sampling plans for inspection by variables we shall use parameter E defined by relation:

$$E = \frac{I_m}{I_s} 100, \tag{6}$$

where $I_m = N - \left(N-n\right) \cdot L\left(\overline{p}; n, k\right)$ is mean number of items inspected per lot of process average quality for inspection by variables and $I_s = N - (N-n) \cdot L(\overline{p}; n, c)$ is mean number of items inspected per lot of process average quality for inspection by attributes. The parameter $E < 100$ because the sample size in acceptance sampling plans for inspection by variables is always less than the sample size in acceptance sampling plans for inspection by attributes (see e.g. Cowden, 1957). On the other hand, the cost of inspection of one item by variables c_m^* is usually greater than the cost of inspection of the same item by attributes c_s^*, i.e. usually is:

$$c_m = \frac{c_m^*}{c_s^*} > 1\,. \tag{7}$$

For the economical comparison of these plans the parameter c_m, i.e. the ratio of the cost of inspection of one item by variables to the cost of inspection of this item by attributes, must be determined in every real situation (e.g. according to the time of inspection, the cost of the inspection devices etc.). According to (6) and (7):

$$E \cdot c_m = \frac{I_m c_m^*}{I_s c_s^*} 100, \tag{8}$$

where $I_m c_m^*$ is the mean cost of inspection by variables and $I_m c_s^*$ is the mean cost of inspection by attributes (there are no restrictive assumptions on the cost function). Therefore, if c_m is determined and $E \,.\, c_m < 100$ then the LTPD plans for inspection by variables are more economical than the corresponding Dodge-Romig LTPD plans for inspection by attributes. Difference:

$$s = 100 - E \cdot c_m = (1 - \frac{I_m}{I_s} c_m) \cdot 100 \tag{9}$$

then represents the percentage of *savings of the inspection cost* when sampling plan for inspection by variables is used instead of the corresponding plan for inspection by attributes. If:

$$s > 0,$$

then the LTPD plans for inspection by variables are more economical than the corresponding Dodge-Romig LTPD plans for inspection by attributes, if:

$$s < 0,$$

then the LTPD attribute sampling plans are more economical than the LTPD plans for inspection by variables.

Example 1. We have chosen for acceptance sampling the lot tolerance fraction defective $p_t = 0.01$ (i.e. the LTPD is 1%). Let the lot size $N = 4\,000$, the process average fraction defective $\overline{p} = 0.002$ and $c_m = 1.4$ (the cost of inspection of one item by variables is higher by 40% than the cost of inspection of one item by attributes). We shall look for the LTPD plan for inspection by variables. Furthermore, we shall compare this plan and the corresponding LTPD plan for inspection by attributes from economical point of view.

For given parameters $p_t = 0.01$, $N = 4\,000$, $\overline{p} = 0.002$ we shall compute the LTPD plan for inspection by variables – see Klůfa (2010):

$$n = 183,\ k = 2.5233,$$

and $E = 27$. The corresponding LTPD plan for inspection by attributes we find in Dodge and Romig (1998). For these parameters we have:

$$n = 510,\ c = 2,$$

(the sample size for inspection by attributes is greater than the sample size for inspection by variables). For $c_m = 1.4$ the economical parameter s is:

$$s = 100 - 37.8 = 62.2.$$

From this result it follows that under the same protection of consumer the LTPD plan for inspection by variables (183, 2.5233) is more economical than the corresponding Dodge-Romig LTPD attribute sampling plan (510, 2). Since $s = 62.2$, using the LTPD plan for inspection by variables instead of the corresponding plan for inspection by attributes, approximately **62% saving of the inspection cost** (see Table 1) can be expected.

The percentage of savings of the inspection cost when sampling plan for inspection by variables is used instead of the corresponding plan for inspection by attributes, s depends on acceptance sampling parameters p_t, N, $\overline{p}$ and c_m, i.e. s is a function of these parameters:

$$s = s(p_t, N, \overline{p}, c_m). \qquad (10)$$

Values of this function for some parameters p_t, N, $\overline{p}$ and c_m are in Table 1. From Table 1 and from the results of numerical investigations it follows that under the same protection of consumer the LTPD plans for inspection by variables are in many situations *more economical* (saving of the inspection cost is 70% in any cases) than the corresponding Dodge-Romig attribute sampling plans.

Table 1 The percentage of savings s for $p_t = 0.01$, $c_m = 1.4$

$\overline{p}$\N	100	500	1 000	4 000	10 000	50 000	100 000
0.00025	52	64	68	76	80	78	80
0.00050	43	57	64	71	73	75	75
0.00075	36	51	61	69	71	78	73
0.00100	30	47	59	69	69	82	78
0.00125	24	43	57	65	68	71	72
0.00150	20	38	55	64	66	69	69
0.00175	16	36	54	62	66	69	69
0.00200	12	31	51	62	68	71	71
0.00225	8	29	44	59	64	66	68
0.00250	5	24	41	59	62	66	68
0.00275	2	20	38	58	62	68	69
0.00300	–1	17	36	58	62	69	72
0.00325	–4	13	33	54	59	64	65
0.00350	–5	10	29	52	59	64	64
0.00375	–8	6	26	50	58	64	65
0.00400	–9	2	23	48	59	65	66
0.00425	–11	–1	19	47	55	61	65
0.00450	–13	–5	16	45	54	61	66
0.00475	–15	–8	12	43	54	61	68
0.00500	–16	–12	9	41	54	64	71

Source: Own construction

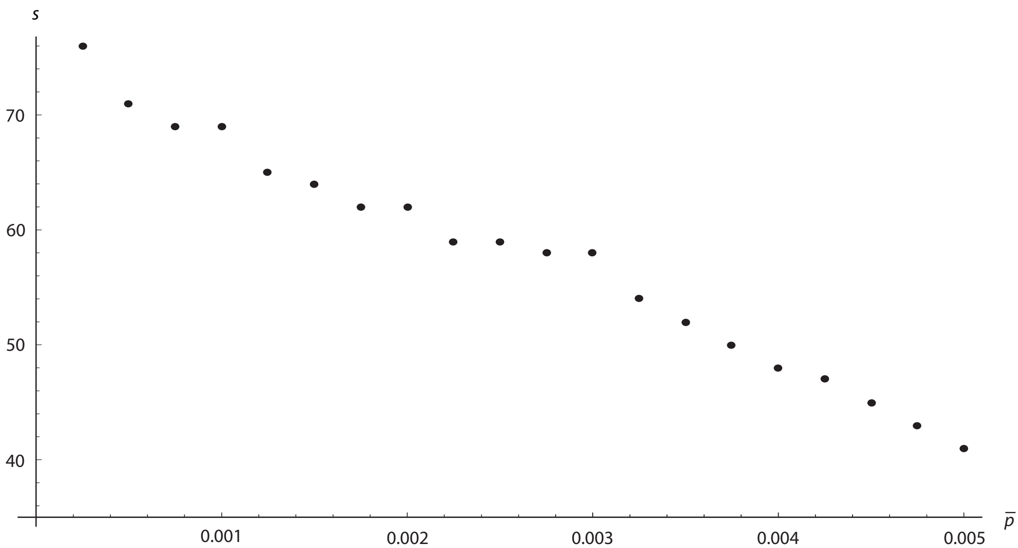
Now, we shall study the dependence savings of the inspection cost (using the LTPD plan for inspection by variables instead of the corresponding LTPD plan for inspection by attributes) on input parameters of acceptance sampling N, $\overline{p}$ and c_m.

Dependence of the percentage of savings s on the lot size N:
In the first step we shall study the dependence savings of the inspection cost (using the LTPD plan for inspection by variables instead of the corresponding LTPD plan for inspection by attributes) on the lot size N. Let p_t, $\overline{p}$, c_m be given parameters. For these given parameters the function s in (10) is a function

of one variable N. This function has an increasing trend in N (it is confirmed by numerical investigations – see also Table 1). Therefore, using the LTPD plan for inspection by variables instead of the corresponding plan for inspection by attributes, *the savings of the inspection cost increases when lot size N increases.*

Dependence of the percentage of savings s on $\overline{p}$:
In the second step we shall study the dependence savings of the inspection cost (using the LTPD plan for inspection by variables instead of the corresponding LTPD plan for inspection by attributes) on the process average fraction defective $\overline{p}$. Let p_t, N, c_m be given parameters. For these given parameters the function s in (10) is a function of one variable $\overline{p}$. This function has mostly decreasing trend in $\overline{p}$ (it is confirmed by numerical investigations – see also Table 1 and Figure 2). Therefore, using the LTPD plan for inspection by variables instead of the corresponding plan for inspection by attributes, *the savings of the inspection cost mostly increases when the process average fraction defective $\overline{p}$ decreases.*

Figure 2 Dependence of savings s (in %) on $\overline{p}$ for $p_t = 0.01$, $N = 4\,000$, $c_m = 1.4$



Source: Own construction

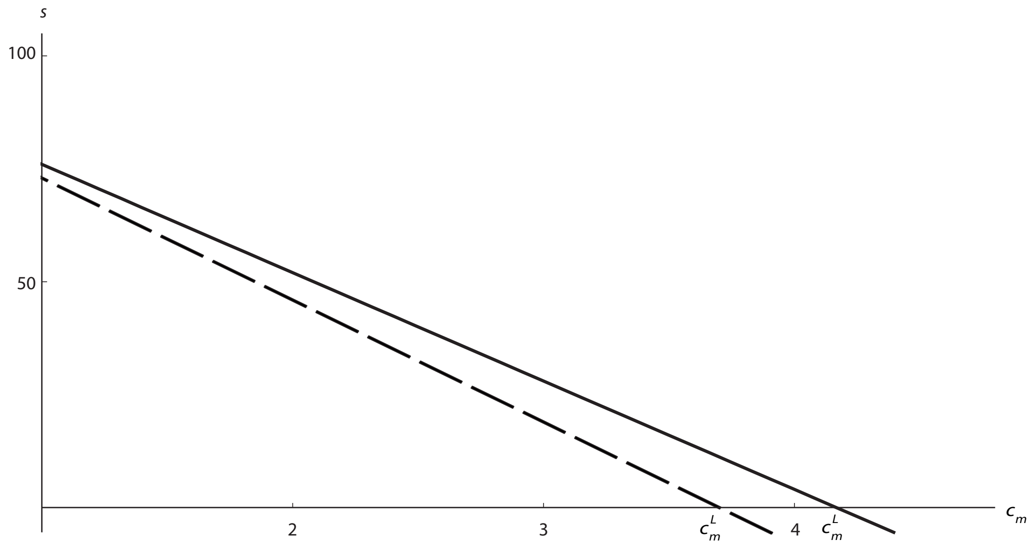
Dependence of the percentage of savings s on the relative cost parameter c_m:
Now, we shall study dependence of the percentage savings of the inspection cost (using the LTPD plan for inspection by variables instead of the corresponding LTPD plan for inspection by attributes) on c_m. Naturally, we can expect that when the ratio of the cost of inspection of one item by variables to the cost of inspection of the same item by attributes c_m increases, then savings of the inspection cost s decreases.

Let p_t, N, $\overline{p}$ be given parameters. Function (10) for given p_t, N, $\overline{p}$ is a function of one variable c_m. Since E in (6) for given p_t, N, $\overline{p}$ is a constant function of c_m (E does not depend on c_m) the function:

$$s = 100 - E \cdot c_m \tag{11}$$

is a linear function of c_m. Due to $E > 0$ this function is decreasing (see Figure 3, the parameter E is the slope of the line). It means that *when c_m increases, then savings of the inspection cost s linearly decreases.*

Figure 3 Dependence of savings s (in %) on c_m for N = 4 000, $\overline{p}$ = 0.002 and p_t = 0.01 (dashed line), p_t = 0.02 (full line)



Source: Own construction

For some value of c_m (denoted c_m^L) is the saving of the inspection cost s = 0 (see Figure 3), i.e. mean inspection cost per lot of process average quality for inspection by variables is equal to mean inspection cost per lot of process average quality for inspection by attributes. From the equation s = 0 (see (11)) we have:

$$c_m^L = \frac{100}{E} = \frac{I_s}{I_m}. \tag{12}$$

The parameter c_m^L defined by Formula (12) can be used for deciding if inspection by variables is considered in place of inspection by attributes. If:

$$c_m < c_m^L, \tag{13}$$

then $s > 0$ (see Figure 3), i.e. the LTPD plans for inspection by variables are more economical than the corresponding Dodge-Romig LTPD attribute sampling plans. On the other hand, if:

$$c_m > c_m^L, \tag{14}$$

then s < 0 (see Figure 3), i.e. inspection by attributes is better than inspection by variables.

Example 2. We have chosen for acceptance sampling the lot tolerance fraction defective p_t = 0.01 (i.e. the LTPD is 1%). Let the lot size N = 4 000 and the process average fraction defective $\overline{p}$ = 0.002. We shall find the values of the relative cost parameter c_m for which the LTPD plan for inspection by variables is more economical than the corresponding Dodge-Romig LTPD attribute sampling plan.

We shall determine deciding point c_m^L (a limit value of parameter c_m) according to (12). For given parameters p_t, N, $\overline{p}$ we have E = 27 (see Example 1). Therefore, the deciding point c_m^L = 3.7. The LTPD plan for inspection by variables is more economical than the corresponding Dodge-Romig LTPD attribute sampling plan when the ratio of cost of inspection of one item by variables to cost of inspection of this item by attributes:

$c_m < 3.7$.

In this situation we can recommend inspection by variables (usually is $c_m^* < 3.7c_s^*$, where c_s^* is the cost of inspection of one item by attributes, c_m^* is the cost of inspection of the same item by variables).

The relative cost parameter c_m is not known in practice. Therefore, for deciding if the LTPD plan for inspection by variables is more economical than the corresponding Dodge-Romig LTPD plan for inspection by attributes we shall calculate the parameter c_m^L. According to c_m^L you can get an idea of whether the inspection by variables is better than the inspection by attributes (see Example 2). If c_m^L is high, then inspection by variables is usually better than inspection by attributes and using the LTPD plan by variables can bring significant savings of the inspection cost. In this situation, it makes sense to determine the relative cost parameter c_m and to find s.

The parameter c_m^L in Formula (12) is a function of three parameters p_t, N, $\overline{p}$, i.e. $c_m^L = c_m^L(p_t, N, \overline{p})$ Values of this function for some parameters $p_t, N, \overline{p}$ are in Table 2 and Table 3.

Table 2 Values of parameter c_m^L for $p_t = 0.01$

$\overline{p}$\N	100	500	1 000	4 000	10 000	50 000	100 000
0.00025	2.9	3.8	4.3	5.9	7.1	6.3	7.1
0.00050	2.4	3.2	3.8	4.8	5.3	5.6	5.6
0.00075	2.2	2.9	3.6	4.5	4.8	6.3	5.3
0.00100	2.0	2.6	3.4	4.5	4.5	7.7	6.3
0.00125	1.9	2.4	3.2	4.0	4.3	4.8	5.0
0.00150	1.8	2.3	3.1	3.8	4.2	4.5	4.5
0.00175	1.7	2.2	3.0	3.7	4.2	4.5	4.5
0.00200	1.6	2.0	2.9	3.7	4.3	4.8	4.8
0.00225	1.5	2.0	2.5	3.4	3.8	4.2	4.3
0.00250	1.5	1.9	2.4	3.4	3.7	4.2	4.3
0.00275	1.4	1.8	2.3	3.3	3.7	4.3	4.5
0.00300	1.4	1.7	2.2	3.3	3.7	4.5	5.0
0.00325	1.4	1.6	2.1	3.0	3.4	3.8	4.0
0.00350	1.3	1.6	2.0	2.9	3.4	3.8	3.8
0.00375	1.3	1.5	1.9	2.8	3.3	3.8	4.0
0.00400	1.3	1.4	1.8	2.7	3.4	4.0	4.2
0.00425	1.3	1.4	1.7	2.6	3.1	3.6	4.0
0.00450	1.2	1.3	1.7	2.6	3.0	3.6	4.2
0.00475	1.2	1.3	1.6	2.4	3.0	3.6	4.3
0.00500	1.2	1.3	1.5	2.4	3.0	3.8	4.8

Source: Own construction

Now, we shall study for given lot tolerance fraction defective p_t the dependence c_m^L on input parameters of acceptance sampling.

Dependence of the limit value c_m^L on the lot size N:
Let p_t and $\overline{p}$ be given parameters. For given parameters p_t and $\overline{p}$ the function c_m^L in (12) is a function of one variable N, which has increasing trend in N (it is confirmed by numerical investigations – see also Table 2 and Table 3). Therefore, *when lot size N increases, then the limit value c_m^L increases* (using the LTPD plan for inspection by variables instead of the corresponding plan for inspection by attributes can be efficient).

Table 3 Values of parameter c_m^L for $p_t = 0.02$

$\overline{p}$\N	100	500	1 000	4 000	10 000	50 000	100 000
0.0005	2.5	3.1	3.7	4.0	4.5	5.0	4.8
0.0010	2.1	2.8	3.7	3.7	3.7	3.8	3.8
0.0015	1.9	2.6	3.7	3.8	3.6	3.7	4.3
0.0020	1.7	2.6	3.7	4.2	3.8	4.3	5.9
0.0025	1.6	2.4	2.7	3.1	3.3	3.4	3.7
0.0030	1.5	2.4	2.6	3.0	3.1	3.2	3.3
0.0035	1.4	2.3	2.6	2.9	3.1	3.2	3.3
0.0040	1.4	2.2	2.5	2.9	3.2	3.4	3.6
0.0045	1.3	1.9	2.2	2.8	2.9	3.1	3.1
0.0050	1.3	1.8	2.2	2.6	2.9	3.1	3.1
0.0055	1.2	1.7	2.1	2.6	2.9	3.2	3.4
0.0060	1.2	1.7	2.0	2.6	3.0	3.6	4.0
0.0065	1.2	1.6	2.0	2.5	2.7	2.9	2.9
0.0070	1.2	1.5	1.9	2.4	2.6	2.9	2.9
0.0075	1.1	1.5	1.8	2.4	2.6	2.9	3.0
0.0080	1.1	1.4	1.8	2.4	2.7	3.1	3.3
0.0085	1.1	1.4	1.7	2.3	2.5	2.8	2.8
0.0090	1.1	1.4	1.6	2.2	2.4	2.7	2.8
0.0095	1.1	1.3	1.6	2.2	2.4	2.8	2.9
0.0100	1.1	1.3	1.5	2.2	2.4	2.9	3.2

Source: Own construction

Dependence c_m^L on the process average fraction defective $\overline{p}$:
Let p_t and N be given parameters. For given parameters p_t and N the function c_m^L in (12) is a function of one variable $\overline{p}$, which has mostly a decreasing trend in $\overline{p}$ (it is confirmed by numerical investigations – see also Table 2 and Table 3). Therefore, *when the process average fraction defective $\overline{p}$ increases, then the limit value c_m^L decreases.*

CONCLUSION

Using the LTPD plans for inspection by variables instead of the corresponding Dodge-Romig LTPD attribute sampling plans we can achieve significant savings of the inspection cost (under the same protection of consumer). For chosen value of the lot tolerance percent defective LTPD the savings of the inspection cost depends on input acceptance sampling parameters N (the lot size), $\overline{p}$ (the process average fraction defective) and the relative cost parameter c_m (the ratio of the cost of inspection of one item by variables to the cost of inspection of the same item by attributes). The results of present paper suggest that the savings of the inspection cost increases when lot size N increases and the process average fraction defective $\overline{p}$ decreases. Naturally, the saving of the inspection cost is greater when c_m is close to one (usually is $c_m > 1$, for $c_m \leq 1$ the LTPD plans for inspection by variables are evidently most economical).

The limit value of parameter c_m (denoted c_m^L) was suggested in this paper as a criterion for deciding if inspection by variables is considered instead of inspection by attributes. When $c_m < c_m^L$ then the LTPD plans for inspection by variables are more economical than the corresponding Dodge-Romig LTPD attribute sampling plans, i.e. inspection by variables is efficient especially for high values of c_m^L. Values of parameter c_m^L depend (for chosen value of the LTPD) on the lot size and the process average fraction defective. Similarly, the limit value c_m^L increases when the lot size N is increasing and when the process average fraction defective $\overline{p}$ decreases.

Further savings of the inspection cost can be achieved by using the LTPD plans for inspection by variables and attributes (all items from the sample are inspected by variables, remainder of rejected lot is inspected only by attributes) – see Klůfa (2015b). However, the application of these LTPD plans for inspection by variables and attributes is a little more complicated. For determination of the LTPD plan for inspection by variables and attributes we first need to estimate the relative cost parameter c_m.

ACKNOWLEDGMENT

This paper was processed with contribution of long term institutional support of research activities by Faculty of Informatics and Statistics, University of Economics, Prague, Czech Republic (IP400040).

References

ASLAM, M., AZAM, M., JUN, C.-H. A new lot inspection procedure based on exponentially weighted moving average. *International Journal of Systems Science,* 2015, 46, pp. 1392–1400. DOI: 10.1080/00207721.2013.822128.

BALAMURALI, S., AZAM, M., ASLAM, M. Attribute-Variable Inspection Policy for Lots using Resampling Based on EWMA. *Communications in Statistics – Simulation and Computation,* 2014, 45(8), pp. 3014–3035. DOI: 10.1080/03610918.2014.938827.

CHEN, C. H. AND CHOU, C. Y. Economic design of Dodge-Romig lot tolerance per cent defective single sampling plans for variables under Taguchi's quality loss function. *Total Quality Management,* 2001, 12, pp. 5–11. DOI: 10.1080/09544120020010057.

CHEN, C. H. AND CHOU, C. Y. Optimum process mean setting based on variable sampling plans with specified consumer's risk. *Journal of industrial and Production Engineering,* 2013, 30, pp. 473–479. DOI: 10.1080/21681015.2013.851124.

CHEN, C. H. Optimum process mean setting with specified average outgoing quality limit protection for variable single sampling plan. *Journal of Statistics & Manegement Systems,* 2016, 19(3), pp. 499–508. DOI: 10.1080/09720510.2015.1131036.

COWDEN, D. J. *Statistical methods in quality control.* New Jersey: Prentice-Hall, 1957.

DODGE, H. F. AND ROMIG, H. G. *Sampling Inspection Tables: Single and Double Sampling.* New York: John Wiley, 1998.

HALD, A. *Statistical Theory of Sampling Inspection by Attributes.* London: Academic Press, 1981.

JOHNSON, N. L. AND WELCH B. L. Applications of the Non-central t distribution. *Biometrika*, 1940, 31, pp. 362–389.

KASPŘÍKOVÁ, N. AND KLŮFA, J. AOQL Sampling Plans for Inspection by Variables and Attributes Versus the Plans for Inspection by Attributes. *Quality Technology and Quantitative Management,* 2015, 12, pp. 133–142. DOI: 10.1080/16843703.2015.11673372.

KLŮFA, J. Acceptance sampling by variables when the remainder of rejected lots is inspected. *Statistical Papers,* 1994, 35, pp. 337–349. DOI: 10.1007/BF02926427.

KLŮFA, J. Exact calculation of the Dodge-Romig LTPD single sampling plans for inspection by variables. *Statistical Papers,* 2010, 51, pp. 297–305. DOI: 10.1007/s00362-008-0160-1.

KLŮFA, J. Sampling Inspection by Variables Versus Sampling Inspection by Attributes. In: *The 9th International Days of Statistics and Economics,* 2015a, pp. 804–810. ISBN 978-80-87990-06-3.

KLŮFA, J. Economic aspects of the LTPD single sampling inspection plans. *Agricultural Economics – Czech*, 2015b, 61(7), pp. 326–331. DOI: 10.17221/186/2014-AGRICECON.

WANG, F. K. AND LO, S.-C. Single Mixed Sampling Plan Based on Yield Index for Linear Profiles. *Quality and Reliability Engineering International,* 2016, 32(4), pp. 1535–1543. DOI: 10.1002/qre.1892.

WANG, F. K. A Single Sampling Plan Based on Exponentially Weighted Moving Average Model for Linear Profiles. *Quality and Reliability Engineering International,* 2016, 32(5), pp. 1795–1805. DOI: 10.1002/qre.1914.

YAZDI, A. A., FALLAH NEZHAD, M. S., SHISHEBORI, D., MOSTAFAEIPOUR, A. Development of an Optimal Design for Conforming Run Length Sampling Methods in the Presence of Inspection Errors. *Journal of Testing and Evaluation,* 2016, 44(5), pp. 1885–1891. DOI: 10.1520/JTE20140478.

YAZDI, A. A. AND FALLAHNEZHAD, M. S. Comparison between count of cumulative conforming sampling plans and Dodge-Romig single sampling plan. *Communications in Statistics – Theory and Methods,* 2017, 46, pp. 189–199. DOI: 10.1080/03610926.2014.988263.

YEN, C.-H., ASLAM, M., JUN, C.-H. A lot inspection sampling plan based on EWMA yield index. *The International Journal of Advanced Manufacturing Technology,* 2014, 75, pp. 861–868. DOI: 10.1007/s00170-014-6174-z.

Trend and Seasonality in Fatal Road Accidents in the U.S. States in 2006–2016

Jiří Procházka [1] | *University of Economics, Prague, Czech Republic*
Milan Bašta [2] | *University of Economics, Prague, Czech Republic*
Matej Čamaj [3] | *University of Economics, Prague, Czech Republic*
Samuel Flimmel [4] | *University of Economics, Prague, Czech Republic*
Milan Jantoš [5] | *University of Economics, Prague, Czech Republic*

Abstract

Understanding the dynamics of the daily number of fatal road traffic accidents is important for local authorities, police departments, healthcare facilities and insurance companies, enabling them to design preventive measures, provide appropriate emergency service and care and reliably estimate traffic accident insurance costs. In the present study, using the Fatality Analysis Reporting System provided by the U.S. National Highway Traffic Safety Administration, we construct a daily time series of the number of accidents for each state of the United States. We model the trend as well as yearly and weekly seasonality present in the time series and provide respective trend and seasonality statistics. Differences in accident rates and yearly seasonality between states were detected, clustering analysis being applied to identify clusters of states with similar yearly seasonality, weekly seasonal patterns for different states proving to be about the same.

Keywords	***JEL code***
Road traffic accidents, yearly seasonality, long seasonal period, generalized linear model, model selection, cluster analysis	*C22, R41*

INTRODUCTION

The main aim of this paper is to examine the daily number of motor vehicle accidents on the roads of the Unites States that involve at least one fatality. Special focus will be given on the characteristics

[1] Department of Statistics and Probability, Faculty of Informatics and Statistics, University of Economics, Prague, W. Churchill Sq. 4, 130 67 Prague 3, Czech Republic.

[2] Department of Statistics and Probability, Faculty of Informatics and Statistics, University of Economics, Prague, W. Churchill Sq. 4, 130 67 Prague 3, Czech Republic. Corresponding author: e-mail: milan.basta@vse.cz.

[3] Department of Statistics and Probability, Faculty of Informatics and Statistics, University of Economics, Prague, W. Churchill Sq. 4, 130 67 Prague 3, Czech Republic.

[4] Department of Statistics and Probability, Faculty of Informatics and Statistics, University of Economics, Prague, W. Churchill Sq. 4, 130 67 Prague 3, Czech Republic.

[5] Department of Statistics and Probability, Faculty of Informatics and Statistics, University of Economics, Prague, W. Churchill Sq. 4, 130 67 Prague 3, Czech Republic.

of the trend and seasonal components, the yearly seasonal component in particular. The findings of the study can be useful for local authorities, police departments and hospitals as well as for insurance companies.

Levine et al. (1995) investigated changes in the daily number of motor vehicle accidents for the City and County of Honolulu in 1990. Their results suggest that more accidents occur during Fridays and Saturdays and during minor holidays. They also identified weather conditions as a relevant factor influencing the number of accidents. Nofal and Saeed (1997) examined monthly variations of the number of road accidents in Riyadh city from 1989 through 1993. Among others, they observed seasonal variations in the number of accidents, the accidents being maximal during the summer season. Edwards (1996) identified an increasing pattern in the number of accidents throughout the calendar year for England and Wales in the period from 1980 to 1990, the first quarter of the year having the lowest level and the last quarter the highest level of accidents. Jones et al. (2008) studied variations in mortality and morbidity from road traffic accidents in England and Wales from 1995 through 2000. Using a geographical approach and district-level data with various explanatory variables (population numbers and characteristics, traffic exposure, road length, curvature and junction density, land use, elevation, hilliness, etc.), they identified risk factors that predicted variations in mortality and morbidity.

In our analysis, we construct daily time series of the number of motor vehicle road accidents involving at least one fatality for each state of the United States from 2006 to 2016. We model the trend and seasonal components of the time series for each state separately and provide summary statistics. We also explore geographical associations.

The data used in the analysis are presented in Section 1. The model for the number of accidents is introduced in Section 2. The results of the analysis are provided in Section 3. The last section concludes.

1 DATA

Fatality Analysis Reporting System (FARS) data provided free of charge by the U.S. National Highway Traffic Safety Administration (NHTSA)[6] have been used. The FARS database offers detailed information on each motor vehicle road accident in the United States which involves at least one fatality. FARS stores each accident as an individual data record with a unique identification number, revealing the details of the accident such as the date and time, geographic location, number of fatalities, weather conditions, etc.

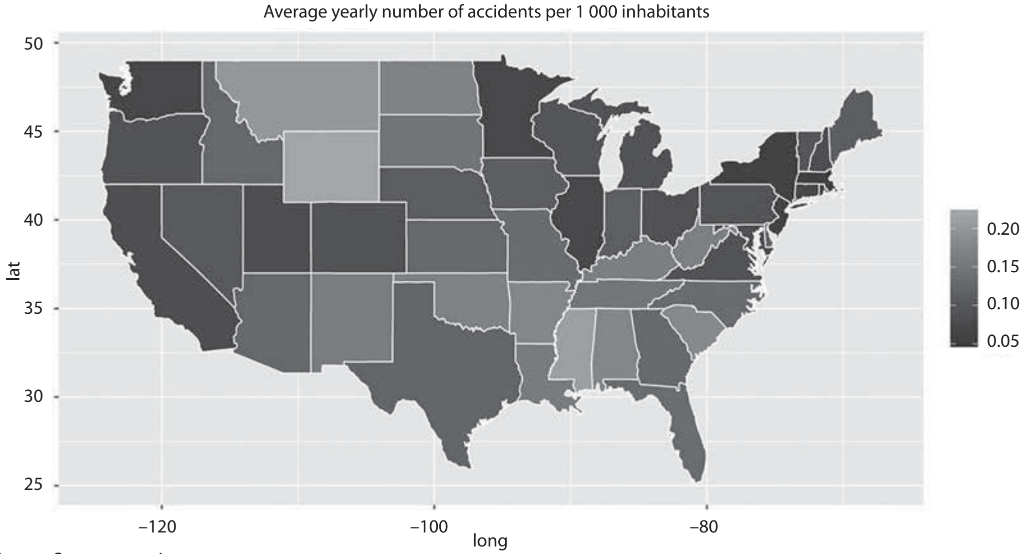
We have removed duplicate records using the identification number of each accident. Yearly, monthly and daily averages and corresponding rates per 1 000 population (in parentheses) for the entire United States[7] are as follows: 32 664.64 (0.1058), 2 722.97 (0.0088) and 89.49 (0.0003).

Then we constructed a daily time series of number of accidents for each U.S. state from the beginning of 2006 to the end of 2016, leap days having been removed to simplify the analysis of yearly seasonality (see below). As a result, we have obtained a time series consisting of eleven times 365 (i.e. 4015) observations for each of the 50 states of the U.S.A. and the District of Columbia.[8]

Average annual accident rates per 1 000 population for each state[9] in the 2006–2016 period are presented in Figure 1. The average rate per state is 0.12, the lowest (0.050) and highest (0.21) ones occurring in the District of Columbia and in Mississippi, respectively. The Pacific coast and north-eastern states generally report lower rates per 1 000 population compared to the rest of the United States.

6 https://www-fars.nhtsa.dot.gov.
7 The number of inhabitants was obtained from the 2010 United States Census at: https://www.census.gov.
8 Although the federal District of Columbia (Washington, D.C.) is not a state, it is considered as such (i.e. "the 51st state") in the present analysis.
9 Data on the population of each state were obtained from the 2010 United States Census, available at: https://www.census.gov.

Figure 1 Average yearly number of accidents per 1 000 population, excluding results for Hawaii (0.078) and Alaska (0.088)



Source: Own construction

2 MODEL FOR THE NUMBER OF ACCIDENTS

Let $\{X_t\}$ be a time series of the number of daily road accidents of length $N = 4\ 015$. Since the number of accidents is a non-negative integer, it can be assumed – providing that the number of accidents is not large enough – that X_t (i.e. the number of accidents at a specific time t) is non-Gaussian. Thus, it may be useful to consider some distribution for X_t which relaxes the normality assumption. Specifically, we can assume that X_t has a density belonging to the exponential family of distributions (Nelder and Baker, 1972; McCullagh and Nelder, 1989):

$$f\left(X_t;W_t,k_t\right)=\exp\left(\frac{X_tW_t-\mathrm{a}\left(W_t\right)}{k_t}\right)\mathrm{c}\left(X_t,k_t\right), \tag{1}$$

where W_t is the canonical parameter, k_t the dispersion parameter and a(.) and c(.) denote some functions. The expected value of a random variable from the exponential family of distributions is equal to the first derivative of $\mathrm{a}(W_t)$ with respect to W_t, while the variance is equal to k_t times the second derivative of $\mathrm{a}(W_t)$ with respect to W_t. Further, the second derivative of $\mathrm{a}(W_t)$ with respect to W_t expressed as a function of the expected value is called the variance function and captures the relationship between the variance of the random variable and its mean.

To be more specific, let us assume that X_t is a Poisson random variable with parameter λ_t. Such a distribution is a special case of an exponential family distribution with $W_t = \log\lambda_t$, $\mathrm{a}(W_t) = e^{W_t}$ and $k_t = 1$. Consequently, we get the following results: the expected value of X_t is given as $\mu_t = E(X_t) = \mathrm{a}^{(1)}(W_t) = \lambda_t$, its variance as $D(X_t) = k_t\mathrm{a}^{(2)}(W_t) = \lambda_t$ and the variance function as $V(\mu_t) = \mu_t$.

We further assume that $\{X_t: t = 1, \ldots, N\}$ is a sequence of N *independent* Poisson random variables with parameters λ_t $(t = 1, \ldots, N)$, the means μ_t $(t = 1, \ldots, N)$ of the variables being given as:

$$\log\left(\mu_t\right)=T_t+S_t,\ \ t=1,\ldots,N, \tag{2}$$

$$\mu_t=\exp\left(T_t+S_t\right)=\exp\left(T_t\right)\exp(S_t),\ \ t=1,\ldots,N. \tag{3}$$

The model of Formula 2 can be considered as a generalized linear model (GLM), see McCullagh and Nelder (1989). In generalized linear models a monotonic function of the expected value, called a link function, rather than the expected value itself is modeled as a linear combination of regressors. This linear combination is called a linear predictor.

A possible choice (but not the only one) of the link function is the *canonical* link function $g(\mu_t)$ which satisfies $g^{(1)}(\mu_t) = \frac{1}{V(\mu_t)}$. For the case of Poisson distribution this implies that the canonical link function is a logarithmic function. Such a canonical link function is also used on the left-hand side of Formula 2.

The linear predictor on the right-hand side has two parts: a trend component of the linear predictor (T_t) and a seasonal component of the linear predictor (S_t). The two parts will be specified in Sections 2.1 and 2.2. The trend and seasonal component in the mean daily number of accidents are given as $\{\exp(T_t)\}$ and $\{\exp(S_t)\}$, the model for the mean daily number of accidents being multiplicative (see Formula 3).

McCullagh and Nelder (1989) or Dobson and Barnett (2008) present further details on generalized linear models which also include the estimation of the models.

In the generalized linear model we assumed that the Poisson random variables $\{X_t: t = 1,\ldots, N\}$ are independent. This assumption was checked during our analysis and was found to be reasonably satisfied (see Section 3 for details).

If the assumption of the independence of the N Poisson random variables $\{X_t: t = 1,\ldots, N\}$ was not satisfied, the generalized linear model could be extended to capture the dependence among the variables by assuming Poisson generalized ARMA models (GARMA) which can be formulated as:

$$\log(\mu_t) = R_t + \sum_{j=1}^{p}\phi_j\left[\log\left(X'_{t-j}\right) - R_{t-j}\right] + \sum_{j=1}^{q}\theta_j\left[\log\left(\frac{X'_{t-j}}{\mu_{t-j}}\right)\right], \tag{4}$$

where μ_t is the expected value of $\{X_t\}$ conditional on all the information available at time t, $R_t = T_t + S_t$, and ϕ_j, for $j = 1, 2, ..., p$, and θ_j, for $j = 1, 2, ..., q$ are parameters and X'_t is a modified time series defined as $X'_t = \max(X_t, c)$, where $c \in (0, 1)$ (Dunsmuir and Scott, 2015; De Andrade, 2016).

2.1 Trend and yearly seasonal component representation

Four different models for $\{T_t\}$ will be examined: no trend (i.e. intercept only), linear, quadratic and cubic deterministic trend.

$\{S_t\}$ will be decomposed into two parts $\{S_{1t}\}$ and $\{S_{2t}\}$, the former representing yearly and the latter weekly seasonality. Specifically, we write:

$$S_t = S_{1,t} + S_{2,t}, \qquad t = 1,\ldots,N. \tag{5}$$

Yearly seasonality has a long seasonal period, namely $L = 365$. On the other hand, the period of weekly seasonality cannot be considered as a long seasonal period since it is equal to 7.

Yearly seasonality can be modeled using cubic splines. Specifically, let us assume the following cubic spline function (Ramsay and Silverman, 2002; or Ramsay and Silverman, 2005) defined in the interval $[0, L]$, where $L = 365$,

$$C(t_*) = \beta_0 + \beta_1 t_* + \beta_2 t_*^2 + \beta_3 t_*^3 + \sum_{k=1}^{K}\theta_k\left(t_* - \xi_k\right)_+^3,\ 0 \le t_* \le L, \tag{6}$$

where $0 \le t_* \le L$ is a real-valued variable, $(.)_+$ denotes the positive part of the expression in brackets, β_0, β_1, β_2, β_3 and θ_k, for $k = 1, 2, ..., K$, are parameters and ξ_k, for $k = 1, 2, \ldots, K$, called the *knots*, are integer- or real-valued constants which satisfy $0 \leq \xi_1 < \xi_2 < \cdots < \xi_K < L$. $C(t_*)$ is a piecewise cubic polynomial and is continuous in the interval $[0, L]$. At the knots, the first and second derivatives of $C(t_*)$ exist, whereas the third ones do not.

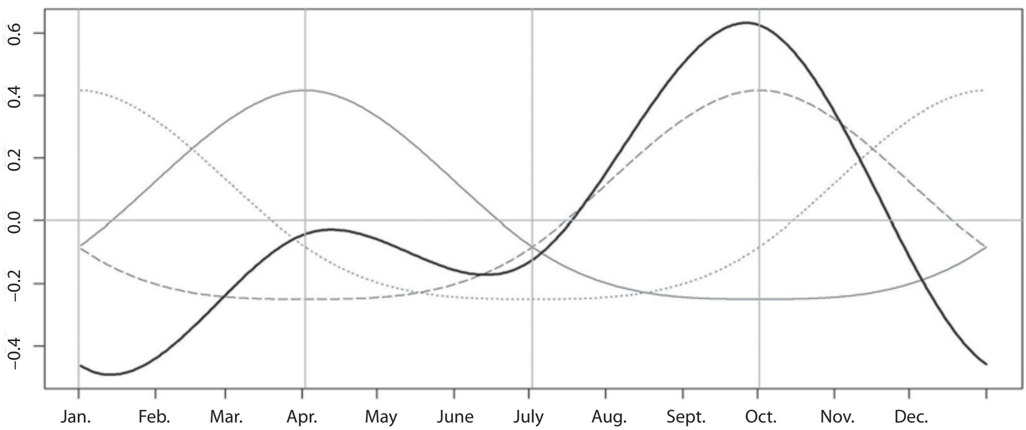
Further, the following constraints are applied to the function $C(t_*)$:

$$\begin{aligned} &C(0) = C(L), \\ &C(0)^{(1)+} = C(L)^{(1)-}, \\ &C(0)^{(2)+} = C(L)^{(2)-}, \\ &C(0)^{(3)+} = C(L)^{(3)-}, \\ &\int_0^L C(l)\,dl = 0, \end{aligned} \tag{7}$$

where $C(0)^{(1)+}$, $C(0)^{(2)+}$, $C(0)^{(3)+}$, are the first, second and third right derivatives at the point 0, and $C(L)^{(1)-}$, $C(L)^{(2)-}$, $C(L)^{(3)-}$, the three left ones at the point L. The five constraints of Formula 7 effectively reduce the number of the cubic spline parameters by 5, ensuring that both ends of the function "connect smoothly" and that the seasonal deviations sum up to zero.

Even though we do not provide an explicit formula for the constrained cubic spline since it would be too complex, it is important to emphasize that the constrained cubic spline can be written as a *linear combination* of K – 1 basis functions. An illustrative example is presented in Figure 2.

Figure 2 An example of a constrained cubic spline function (solid black curve) with four knots in the interval [0, 365]. The position of the knots is denoted by four gray vertical lines. The spline function is obtained as a linear combination of three basis functions denoted by the gray solid, gray dotted and gray dashed curve. The weights of the linear combination are chosen as 1 (solid), –0.5 (dotted) and 2 (dashed).



Source: Own construction

If the constrained cubic spline function is periodically extended with a period equal to L, we obtain a *periodic cubic spline* function. If the periodic cubic spline function is sampled at discrete values 1, 2, …, N, it can serve as a representation of the yearly seasonal component $\{S_{1,t}\}$.

Generally, a large number of knots K leads to a less-biased and high-variance estimate of the yearly seasonal pattern, whereas a low number of knots results in a highly biased and low-variance estimate. The time positioning of knots along the calendar year, though subjective to some extent, is also crucial. More densely positioned knots in a given period of the year lead to a less-biased but more variable estimate of the cycle in that period. The positioning of the knots in our paper is described in the next paragraph and the selection of the K value is given in Section 2.3.

In total, eleven different models for the yearly seasonal cycle will be examined in our analysis: no yearly seasonal cycle and ten models based on periodic cubic splines differing in the number of parameters (K – 1): 3, 5, 7, 9, 11, 13, 15, 17, 20 and 30. The knots will be placed equidistantly throughout the year so that the distance between two neighboring knots is constant, the first knot being placed at the beginning of the calendar year. The equidistant placement of the knots is a common choice which often works well (see e.g. Ramsay and Silverman, 2002; or Ramsay and Silverman, 2005). An alternative to the equidistant placement of the knots is to place more knots in those regions where the estimated function exhibits the most complex variations (see e.g. Ramsay and Silverman, 2005) – this approach will not be used in our analysis.

2.2 Weekly seasonal component representation

Weekly seasonality will be modeled as follows:

$$S_{2,t} = \sum_{m=1}^{6} \psi_m Z_{m,t}, \qquad t = 1,\ldots,N, \tag{8}$$

where ψ_m, for m = 1, 2, ..., 6, are parameters and $\{Z_{m,t}$: t = 1, ..., $N\}$, for m = 1, 2, ..., 6, are effect coding variables.

In total, two different models for weekly seasonality will be considered: a model with no weekly seasonality and the model of Formula 8.

2.3 Best model selection

88 different models will be examined for each time series, differing in the number of parameters used for the deterministic trend (four alternatives), yearly (eleven alternatives) and weekly seasonality (two alternatives) in the linear predictor. The model with the lowest value of Akaike information criterion (AIC) among these 88 models will be selected as the best, AIC being defined as:

$$AIC = 2P - 2\hat{l}, \tag{9}$$

where P is the number of parameters to be estimated and $\hat{l}$ is the natural logarithm of the maximized likelihood function.

The best model selected by AIC for each state was further checked whether it conforms to the assumptions of the GLM approach (see Section 3).

R software (R Core Team, 2017) has been employed in the analysis. Namely, the glm() function from the R stats package has been used to perform Poisson regression. The part of the model matrix corresponding to yearly seasonality has been created making use of the pbs() function from the pbs R package (Wang, 2013).

3 RESULTS

As explained above, 88 different models have been considered for each of the 51 time series, Table 1 displaying the frequencies of the models that were selected as the most appropriate.

Let $\{\widehat{X_t}$: t = 1, ..., $N\}$ be the fitted values from the best model and

$$\left\{ \frac{X_t - \widehat{X_t}}{\sqrt{\widehat{X_t}}} : t = 1,\ldots,N \right\} \tag{10}$$

the corresponding Pearson residuals. Based on Cameron and Trivedi (2013, Sec. 7.3.2) we have used the sample autocorrelation function of Pearson residuals from the best model as well as the related significance tests (test for autocorrelation at an individual lag based on a test statistic following normal distribution and portmanteau test based on the Box-Pierce test statistic, see Cameron and Trivedi, 2013, Sec. 7.3.2) to assess the assumption of independence of $\{X_t: t = 1,\ldots, N\}$. This assumption seems to be reasonably satisfied for all the 51 time series.[10]

Table 1 Frequencies of the models selected for the 51 time series (WS stands for "weekly seasonality")

			Number of parameters of $\{T_t\}$				Row sums
			1	2	3	4	
Number of parameters of $\{S_{1t}\}$	0	no WS	0	0	0	1	1
		WS	0	0	1	0	1
	3	no WS	0	0	0	0	0
		WS	0	2	8	3	13
	5	no WS	0	0	1	0	1
		WS	0	2	9	7	18
	7	no WS	0	0	0	0	0
		WS	0	0	5	2	7
	9	no WS	0	0	0	0	0
		WS	0	0	1	2	3
	11	no WS	0	0	0	0	0
		WS	0	0	1	1	2
	13	no WS	0	0	0	0	0
		WS	0	0	1	0	1
	15	no WS	0	0	0	0	0
		WS	0	0	0	1	1
	17	no WS	0	0	0	0	0
		WS	0	0	0	1	1
	20	no WS	0	0	0	0	0
		WS	0	0	1	0	1
	30	no WS	0	0	0	0	0
		WS	0	0	0	1	1
Column sums			0	4	28	19	

Source: Own construction

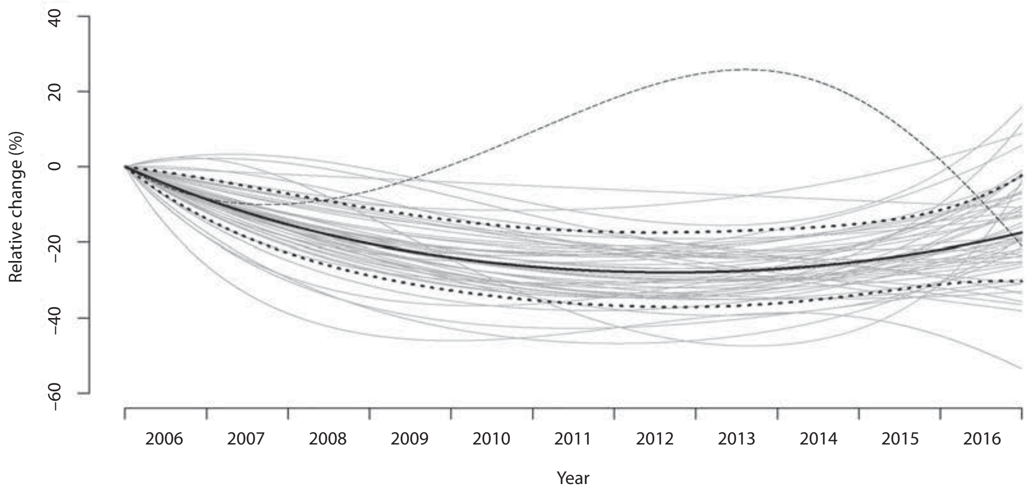
3.1 Examining the trend component

It follows from Table 1 that a quadratic (three-parameter) trend is often the best choice for $\{T_t\}$. If $\{\widehat{T_t}\}$ is an estimate of $\{T_t\}$, then the estimated trend in the mean daily number of accidents is $\{\exp \widehat{T_t}\}$. In Figure 3, the relative change of $\{\exp \widehat{T_t}\}$ (in percentage terms) is indicated for each state with respect to the beginning of the year 2006. There has been a decrease in the number of accidents in most states since 2006 – the least accidents occurring around the year 2012 –, followed by a slight increase in most states.

The relative change of the level of $\{\exp \widehat{T_t}\}$ from the beginning of 2006 to the end of 2016 is shown in percentage points in Figure 4, the average relative change being –16.4 per cent, the minimum and maximum reaching –53.6 and 16.0 per cent in South Dakota and New Hampshire, respectively. There is no clear geographical pattern in the relative change of $\{\exp \widehat{T_t}\}$.

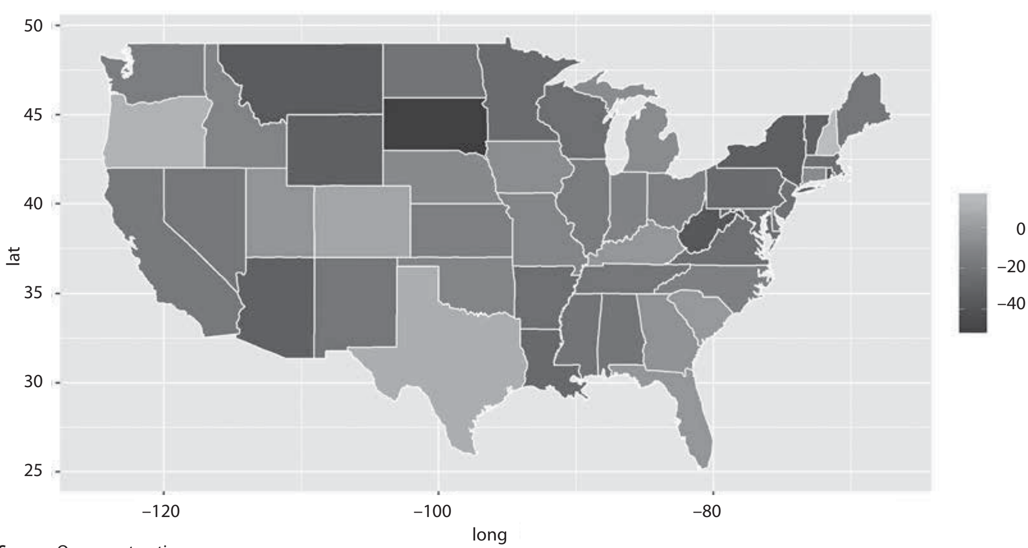
[10] To check the results, we have also performed GARMA modeling, using the glarma() function from the glarma R package (Dunsmuir and Scott, 2015), with $p = 1$, $q = 2$ and $c = 0.01$ (default value). The estimated shapes of seasonal cycles were highly similar to those obtained from GLM modeling.

Figure 3 Relative change (in percentage terms) of $\{\exp \widehat{T_t}\}$ with respect to the beginning of the year 2006 for the 51 states. North Dakota is depicted in thin dashed black, the other states in gray. The geometric mean of the corresponding growth rates (translated into a relative change) is represented by the solid black curve, the two dotted black curves denoting the distance of one geometric standard deviation from the geometric mean.



Source: Own construction

Figure 4 Relative change (in percentage terms) of $\{\exp \widehat{T_t}\}$ from the beginning of 2006 to the end of 2016, excluding results for Hawaii (–29.7) and Alaska (–2.4)



Source: Own construction

3.2 Examining the yearly seasonal component

Table 1 shows that yearly seasonality is present in most of the time series, the most common number of parameters for $\{S_{1,t}\}$ being five. The only two states that do not exhibit yearly seasonality are Hawaii and the District of Columbia. If $\{\widehat{S_{1,t}}\}$ is the estimate of $\{S_{1,t}\}$, then the estimated yearly multiplicative seasonal component in the mean daily number of accidents is $\{\exp \widehat{S_{1,t}}\}$. There is a relatively wide diversity among the 51 states in the temporal variability of $\{\exp \widehat{S_{1,t}}\}$ along the year, some of them showing minor and others considerable variability of the seasonal component.[11]

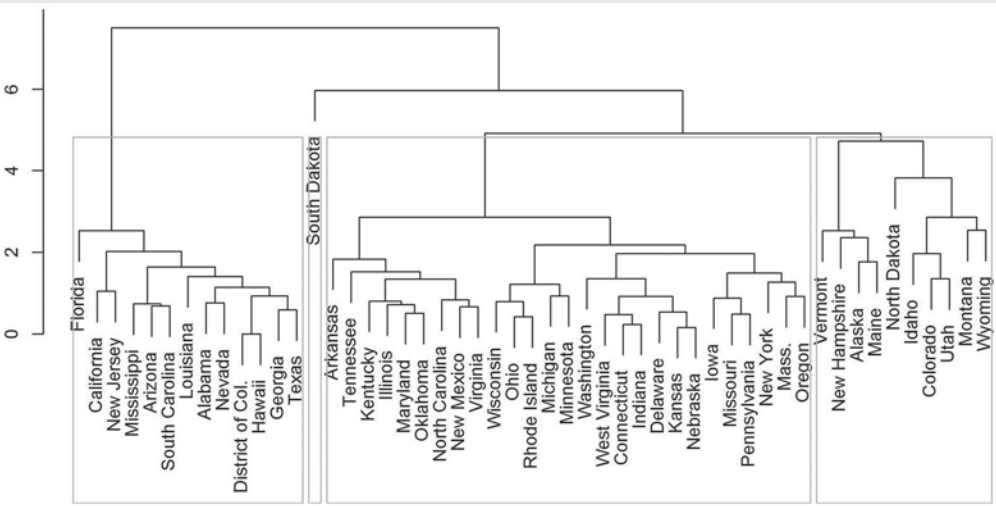
We conduct cluster analysis and form four groups of states with similar yearly seasonalities. Specifically, agglomerative hierarchical clustering with the Euclidean distance and complete linkage (see Everitt and Hothorn, 2011) is used,[12] with each state represented by 365 values[13] $\{\widehat{S_{1,t}} : t = 1, ..., 365\}$. The clustering dendrogram is presented in Figure 5, with the four clusters denoted by gray rectangles.[14] The clusters are also shown in Figures 6 and 7, yearly seasonal cycles $\{\exp \widehat{S_{1,t}}\}$ being illustrated separately for the respective clusters in the latter figure.

It is obvious that the clusters are closely related to the geographical location of each state. The first cluster comprises the Sun Belt southern states such as Florida, Texas and California. Yearly seasonality is not too variable in these states. The geometric mean of geometric standard deviations of $\{\exp \widehat{S_{1,t}}\}$ is 1.05 in the whole cluster, the geometric standard deviation for a single state being defined as:

$$\sigma_g = \exp\left(\sqrt{\frac{1}{365}\sum_{m=1}^{365}\widehat{S_{1,m}}^{\,2}}\right), \tag{11}$$

where $\widehat{S_{1,m}}$, for $m = 1, ..., 365$, is the estimated yearly seasonal component of the linear predictor for the mth day of the year.

Figure 5 Hierarchical clustering dendrogram. The four clusters are denoted by the gray rectangles.



Source: Own construction

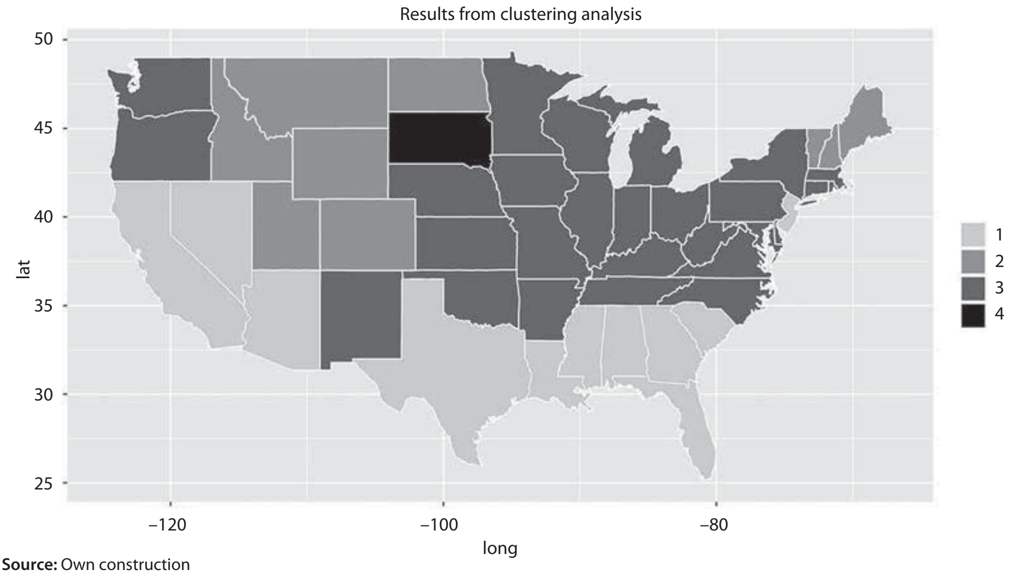
[11] Kirkwood (1979) argues that the geometric mean and geometric standard deviation are reasonable measures of location and spread for the variables which are subject to multiplicative rather than additive variations. Consequently, we use the geometric mean and geometric standard deviation as the measures of location and spread of multiplicative seasonal patterns $\{\exp \widehat{S_{1,t}}\}$ throughout the text.

[12] We use dist(), hclust() and cutree() functions from the R base package to perform the analysis in R.

[13] Since the Euclidean distance is applied, it seems more sensible to do the clustering on $\{\widehat{S_{1,t}}\}$ rather than $\{\exp \widehat{S_{1,t}}\}$.

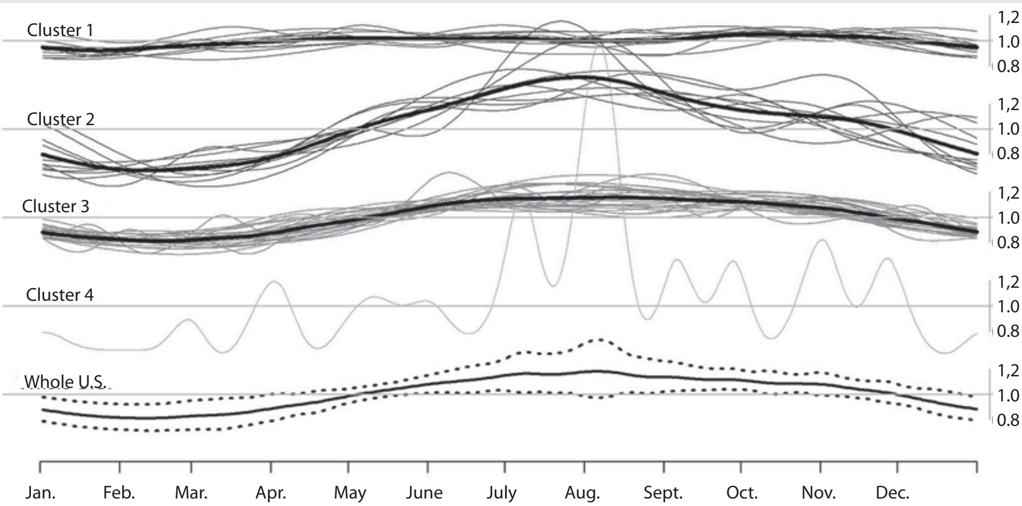
[14] Increasing the number of clusters above five does not lead to a pronounced decrease in the within-cluster sum of squares, while decreasing the number of clusters below three results in a marked increase in the within-cluster sum of squares. Thus, the choice of three, four or five clusters seems to be reasonable. In this analysis, we have opted for four clusters.

Figure 6 Members of the four clusters from hierarchical clustering, except Hawaii (part of the 1st cluster) and Alaska (2nd cluster)



Source: Own construction

Figure 7 Estimated multiplicative yearly seasonal components. The bottom plot (the entire U.S.A.) represents the geometric mean of the 51 seasonal components (thick solid black curve) and the distance of one geometric standard deviation from the geometric mean (two thick dotted curves). Individual seasonal components are presented in the upper plots being grouped into four separate clusters. The geometric mean of each cluster is shown by the thick solid black curve, except for cluster 4 which consists of only one state (South Dakota). Each cluster has its own y axis on the right-hand side of the plot.



Source: Own construction

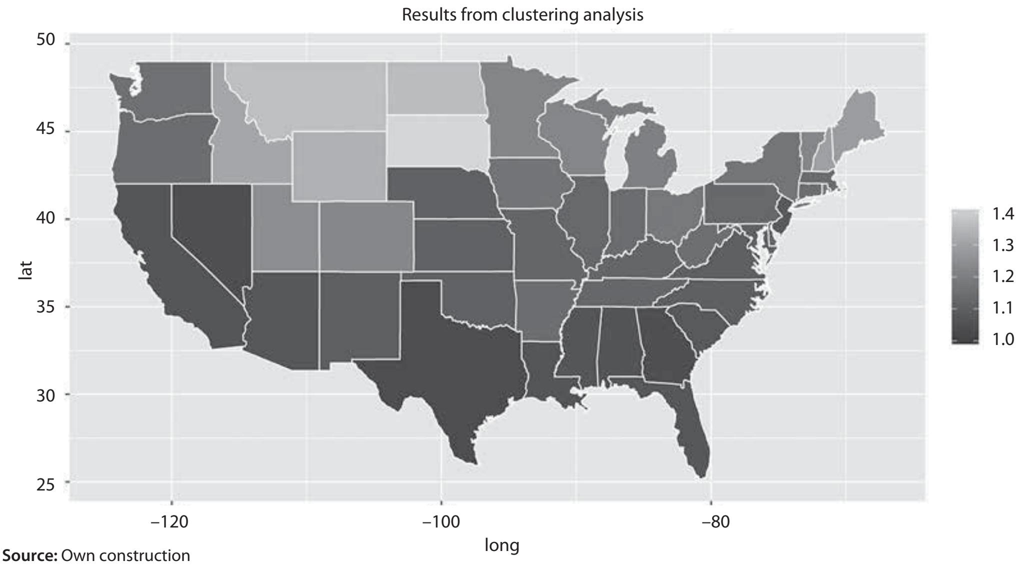
The second cluster consists of the Mountain West states, e.g. Idaho, Montana and Wyoming. The yearly seasonal cycle exhibits significant variations, the geometric mean of geometric standard deviations of $\{\exp \widehat{S_{1,t}}\}$ being 1.30.

The remaining states, except for South Dakota, form the third cluster. They report medium yearly seasonal variability and the geometric mean of geometric standard deviations of $\{\exp \widehat{S_{1,t}}\}$ is 1.14.

South Dakota is the only member of the fourth cluster. It displays extreme yearly seasonal variations, the geometric standard deviation of $\{\exp \widehat{S_{1,t}}\}$ being 1.42.

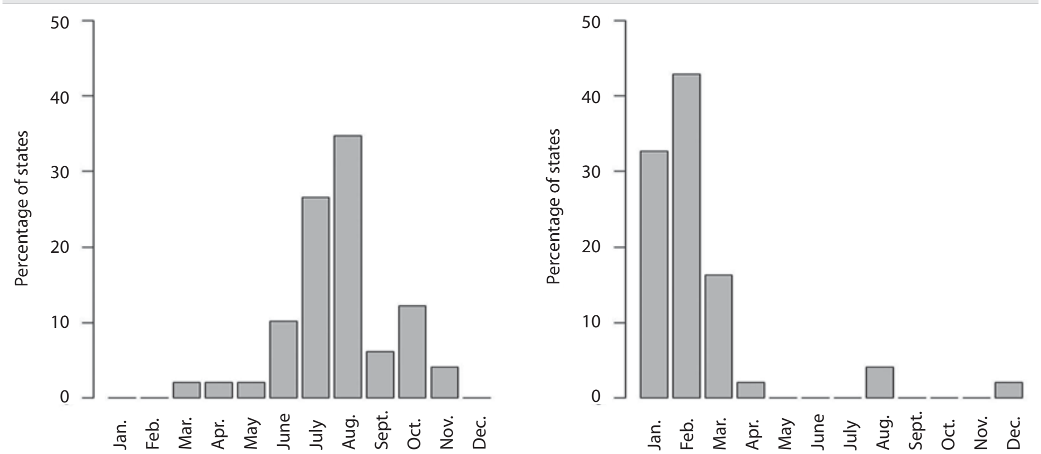
σ_g is plotted for individual states in Figure 8. A clear geographical pattern emerges, related to the above clusters.

Figure 8 Geometric standard deviation of the multiplicative yearly seasonal cycle for each state, except Alaska (1.345) and Hawaii (1). The light and dark colors indicate large and small standard deviations, respectively.



Source: Own construction

Figure 9 Distribution of the maxima (left plot) and minima (right plot) of the yearly multiplicative seasonal component among the states along the year. Percentage of states having a maximum or minimum in a given month is presented on the vertical axis. Hawaii and the District of Columbia are excluded from the plot since they do not exhibit yearly seasonality.



Source: Own construction

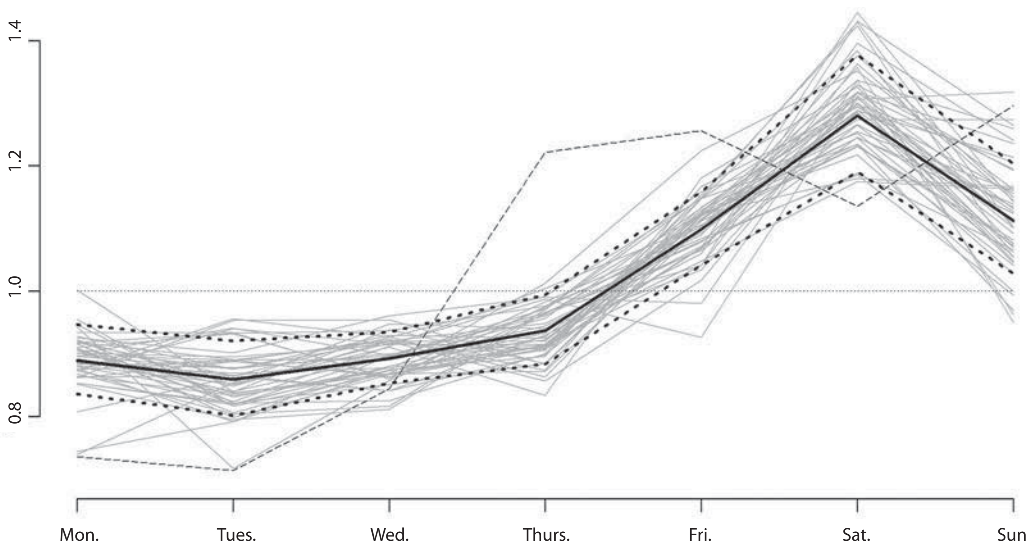
In general, accidents tend to occur more frequently during summer and autumn months as shown in Figures 7 and 9. The latter figure illustrates the distribution of the maxima (left plot) and minima (right plot) of the yearly multiplicative seasonal cycle among the 51 states along the year, the maxima occurring mostly in the summer (July and August) and the minima during winter months (January, February, March).

3.3 Examining the weekly seasonal component

It follows from Table 1 that weekly seasonality is often present in the time series. The only two states that do not exhibit weekly seasonality are Alaska and the District of Columbia. If $\{\widehat{S_{2,t}}\}$ is an estimate of $\{S_{2,t}\}$, then the estimated multiplicative weekly seasonal component in the mean daily number of accidents is $\{\exp \widehat{S_{2,t}}\}$.

Multiplicative weekly seasonal components for the 51 states plotted in Figure 10 are largely similar, except for Alaska and the District of Columbia (see the thin dotted gray line in the figure), which do not show a weekly seasonal pattern, and Rhode Island (thin dashed gray curve), whose pattern slightly differs from that of the other states. The typical pattern has a minimum on Tuesdays and a maximum on Saturdays.

Figure 10 Multiplicative weekly seasonal component for the 51 states. Alaska and the District of Columbia are represented by the thin dotted gray line, Rhode Island by the thin dashed gray curve and the other states by the gray curves. The solid black curve is the geometric mean of individual curves, the two dotted black curves showing the distance of one geometric standard deviation from the geometric mean.



Source: Own construction

DISCUSSION AND CONCLUSIONS

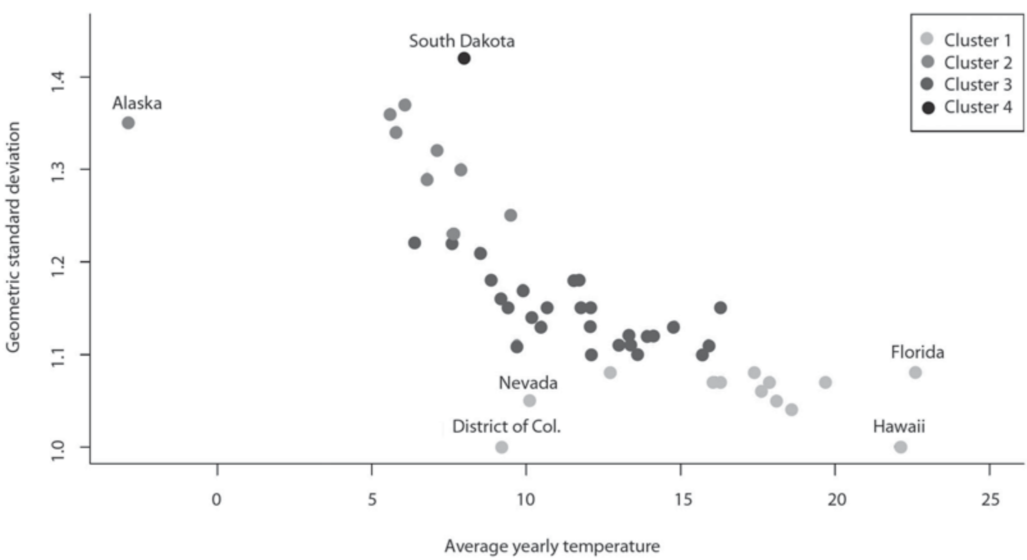
We have constructed daily time series of the number of motor vehicle road accidents with at least one fatality for the individual states of the U.S.A. and the District of Columbia in the period from the beginning of 2006 to the end of 2016. Poisson generalized linear models were used to examine the trend component as well as yearly and weekly seasonal cycles in the time series. Given the long period analyzed, the yearly seasonal component was represented by periodic cubic splines.

Summary statistics on annual averages, the trend component, and yearly and weekly seasonality are presented.

Despite having no intention of exploring causal mechanisms by which the fluctuations of the number of traffic accidents occur, we have some ideas to think about in this context. Specifically, the differences in annual accident rates per 1 000 population between the U.S. states (see Section 1) can be explained by different numbers of motor vehicles and diverse travel motivations and behavior, the above factors resulting in different distances covered by car in a year per 1 000 population. Also, the road network range and quality as well as decisions of the local authorities regarding transport can help to explain the differences (see Peden et al., 2004).

A similar pattern of trend components in the states observed in Section 3.1 suggests that the long-run dynamics of the number of accidents may be governed by some common factors (such as the improved vehicle safety features, the U.S. economic performance and gasoline prices). Longthorne et al. (2010) argue that the decline in the number of accidents is largely due to a decrease in the number of young drivers' car crashes, implying that it might be caused by rising unemployment among the youth. It is obvious that yearly seasonal patterns are related to the geographical location of the states (Section 3.2). Moreover, Figure 11 reveals the relationship between the geometric standard deviation of the multiplicative yearly seasonal pattern $\{\exp \widehat{S_{1,t}}\}$ and the average annual temperature of each state.[15] The states with higher average temperatures tend to have less variable yearly seasonal patterns, whereas those with lower average temperatures report higher variability. This may indicate lower winter traffic volumes (compared to summer ones) in the states with low annual average temperatures since some local roadways are not safe enough to drive on. Summer driving, on the other hand, seems to be more comfortable, which is reflected in higher traffic density levels and accident rates.

Figure 11 Geometric standard deviation of the yearly seasonal cycle plotted against the average annual temperature (in degrees Celsius) for the 51 states



Source: Own construction

[15] Average annual temperatures were obtained from the National Centers for Environmental Information website at: https://www.ncdc.noaa.gov.

The similarity between weekly seasonal patterns (Section 3.3) is explicable by comparable lifestyles in different states, people traveling more by car to see relatives and friends or go on trips, particularly on weekends.

In further research, we will focus on a causal explanation for different accident rates.

ACKNOWLEDGMENTS

This paper is supported by the grant F4/67/2016 (Modelování sezónních časových řad s velkou délkou sezónnosti) which has been provided by Interní grantová agentura Vysoké školy ekonomické v Praze (Internal grant agency of University of Economics, Prague).

We thank Karel Helman, PhD, (Private University College of Economic Studies) for kind assistance in proofreading the text.

We thank the anonymous reviewers whose comments helped improve the content of the paper.

References

CAMERON, A. C. AND TRIVEDI, P. K. *Regression analysis of count data*. Cambridge university press, 2013.

DE ANDRADE, B. S., ANDRADE, M. G., EHLERS, R. S. Bayesian GARMA models for count data. *Communications in Statistics: Case Studies, Data Analysis and Applications*, 2016, 1(4), pp. 192–205.

DOBSON, A. AND BARNETT, A. *An Introduction To Generalized Linear Models.* 3rd Ed. Chapman & Hall/CRC, 2008.

DUNSMUIR, W. AND SCOTT, D. The glarma package for observation driven time series regression of counts. *Journal of Statistical Software*, 2015, 67(7), pp. 1–36.

EDWARDS, J. B. Weather-related road accidents in England and Wales: a spatial analysis. *Journal of Transport Geography*, 1996, 4(3), pp. 201–212.

EVERITT, B. AND HOTHORN, T. *An introduction to applied multivariate analysis with R*. Springer Science & Business Media, 2011.

JONES, A. P., HAYNES, R., KENNEDY, V., HARVEY, I. M., JEWELL, T., LEA, D. Geographical variations in mortality and morbidity from road traffic accidents in England and Wales. *Health & place*, 2008, 14(3), pp. 519–535.

KIRKWOOD, T. Geometric means and measures of dispersion. *Biometrics*, 1979, 35, pp. 908–909.

LONGTHORNE, A., SUMBRAMANIAN, R., CHOU-LIN, C. *An Analysis of the Significant Decline in Motor Vehicle Traffic Crashes in 2008* [online]. Technical report No. DOT HS 811 346, U.S. Department of Transportation, National Highway Traffic Safety Administration, 2010. <https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/811346>.

LEVINE, N., KIM, K. E., NITZ, L. H. Daily fluctuations in Honolulu motor vehicle accidents. *Accident Analysis & Prevention*, 1995, 27(6), pp. 785–796.

MCCULLAGH, P. AND NELDER, J. A. *Generalized linear models*. Chapman and Hall/CRC, 1989.

NELDER, J. A. AND BAKER, R. J. *Generalized linear models*. John Wiley & Sons, Inc., 1972.

NOFAL, F. H. AND SAEED, A. A. W. Seasonal variation and weather effects on road traffic accidents in Riyadh city. *Public health*, 1997, 111(1), pp. 51–55.

PEDEN, M. et al., eds. *World report on road traffic injury prevention*. Geneva: World Health Organization, 2004.

R CORE TEAM. *R: A Language and Environment for Statistical Computing* [online]. Vienna, Austria: R Foundation for Statistical Computing, 2017. <https://www.r-project.org>.

RAMSAY, J. O. AND SILVERMAN, B. W. *Applied functional data analysis: methods and case studies*. New York: Springer, 2002.

RAMSAY, J. O. AND SILVERMAN, B. W. *Functional Data Analysis.* 2nd Ed. Springer, 2005.

WANG, S. *pbs: Periodic B Splines*. R package version 1.1, 2013.

120th Anniversary of Founding the Land Statistical Office of the Bohemian Kingdom

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

Abstract

In the Habsburg monarchy demographic and economic statistics was concentrated mainly in Vienna offices. The agricultural statistics was the only one that was consequently developed in Bohemia (already from year 1770) by Patriotic-Economic Society, that in the years 1856–1858 created its own miniature statistical office. That became after long negotiations the base of the Land Statistical Office that started its operation 120 years ago (year 1898). The Czechoslovak State Statistical Office like today's Czech Statistical Office are its direct followers. Founders of statistical service in Bohemia were Karel Kořistka and Dobroslav Krejčí. Similar statistical office was set up in Moravia in the year 1899. A remarkable achievement here was an approval of the Land Act, which sets the reporting obligation for land statistics actions.

Keywords	***JEL code***
History of statistics, Patriotic-Economic Society, Land Statistical Office, Karel Kořistka, Dobroslav Krejčí	*B16, B23, B31*

INTRODUCTION

Establishment of the Land Statistical Office in Bohemia (LSO) 120 years ago was a significant milestone in the development of official statistics in our country. This issue has already been dealt at the occasion of various anniversary events in the past, e. g. by Krejčí (1925), Závodský (1997b), etc. The circumstances of origin of LSO are also described in several publications focused on the history of statistical service at the Czechoslovak territory – e. g. Podzimek (1974), 70 let (1989), etc. Our contribution is to bring a somewhat new look at the events which happened 120 years ago, to describe in more detail the development leading to the establishment of LSO and correct some mistaken proclamations that appears in existing literature.

1 STATE STATISTICS IN AUSTRIA–HUNGARY

First, let us briefly describe the situation of state statistics in the Habsburg monarchy in the last decade of the 19th century. Both parts of the empire had their own, completely independent statistical service,

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

organized differently in many aspects. In the Austrian part of the monarchy, all most important branches of statistics (with the exception of agriculture) were concentrated in Vienna. It was mainly Imperial-Royal (I. r.) Central Statistical Commission (K. k. statistische Zentralkommission), classified under the Ministry of Religious Affairs and Education.

The Central Statistical Commission fulfilled (since the year 1884) the function of the body with acting and executive powers. The significant modernization of this office and its position has been accomplished on scientific foundations thanks to Karl Theodor von Inama–Sternegg (1843–1908), who had been at the forefront of Austrian statistical services for nearly a quarter of a century (1881–1905).[3] The Central Statistical Commission published many source works (mainly *Österreichische Statistik* and yearbook *Östrereichisches statistisches Handbuch*) and also a monthly published journal *Statistische Monatschrift.*[4] The Commission did not have any territorial authorities or subsidiaries in the individual countries of the monarchy.

Central statistics was decentralized to a large extend in Austrian part of monarchy, many ministries had their own statistical offices or bureaus. In particular, the Ministry of Commerce's statistical service, which had the competence of statistics of industry, trade, foreign trade and "inter-trade" (i.e. trade between the Austrian and Hungarian parts of the monarchy), could compete with the Central Statistical Commission. This Labour Office also established the Labour Statistical Office that was dealing with the social conditions of workers in 1898.

The General Intelligence Obligation in the Habsburg Monarchy was not enacted, only in four cases it was laid down by special legal standards. This refers to the Population Census (law from 29th March 1869, RGBl. Nr. 67/1869) that took place in the years 1869, 1880 and then within a ten-year period. Then, it was census of agricultural and business trades (law from 25th March 1902, RGBl. Nr. 56/1902 – happening only once in the year 1902), statistics of foreign trade (law from 26th June 1890, RGBl. Nr. 132/1890) and statistics of inter-trade (imperial order of 21st September 1899, RGBl. Nr. 176/1899). In all other cases, Vienna's offices could delegate the administration to regional offices of district governments and municipal or town councils.

2 THE BEGINNINGS OF OFFICIAL STATISTICS IN BOHEMIA

In this section, we briefly describe the emergence and gradual development of agricultural statistics in Bohemia. Based on the decree by Maria Theresia from July 1767 I. r. Patriotic-Economic Society (K. k. patriotisch–ökonomische Gesellschaft im Königreich Böhmen) (PES)[5] was founded and started to act in 1770. The aim of the Society was to provide a comprehensive care for the development and modernization of agricultural production in the country, including vocational training at all levels, education, etc. PES was financially supported, but also controlled by Vienna offices and gubernium.

PES issued, in addition to various popular-scientific publications, calendars etc., also scientific journals, the list of which is shown References. Already in the 1st volume of *Abhandlungen* (1797, pp. 53–98), the former secretary of PES, Franz Fuss (1745–1805), published a very successful statistical survey of agriculture in Bohemia, mostly in the form of tables and balance sheets comparing the calculated need for grain, meat, wood, etc. with their actual production. Since the very beginning PES cooperated with the Prague gubernium during various censuses – of cattle, bee colonies, fruit trees, and even maples (during the Napoleonic wars they were used to produce a sugar substitute).

[3] He was called to Vienna from Prague, where he lectured at the University on political economy. In 1899 he was elected president of the International Institute of Statistics (ISI).
[4] *Geschichte* (1979, pp. 57–70).
[5] This definitive name of the company was only since year 1788.

Publication of F. Fuss had not found any follower for a long time. After nearly twenty years of interruption, in 1825 PES began to issue a new scientific journal *Neue Schriften* (until 1847 – 10 volumes per 2 issues were issued). The journal regularly published extensive statistical tables of meteorological observations as well as their analysis – various (even multi-annual) averages of temperatures, precipitation, atmospheric pressure, and average wind direction, calculated according to the complicated formula (e. g. in the year 1828, an average for the whole Bohemia was 53º 52', which meant roughly southwest).[6] Interesting is also the frequent demonstration of different dependencies between the variables, e. g. the dependence of the average temperature on the altitude of the meteorological station, etc. Gradually, the tables of yields of different crops per unit area, tables of average prices of individual crops in different cities, and other analyses and accompanying comments had been expanded and improved.

Since 1850 PES had issued a weekly journal *Centralblatt* (see References) with regular statistical section. An increased interest in actual data about agriculture in Bohemia in a new changed political environment after 1848 raised the question of establishing a statistical department within PES. Antonín J. Oppelt (1789–1864), a landowner and one of the most active PES representatives, was a person who at the management meeting held on 2nd and 9th January 1853,[7] proposed the establishment of a special section on agricultural and forestry statistics. The proposal was unanimously adopted, and a three-member drafting committee was set up. Its member, prof. G. N. Schnabel, failed to prepare the promised program of activities, the proposal of prof. Peter Mischler was found hardly feasible.

The principles of how the PES should organize agricultural and forestry statistics was finally proposed by a member of the Committee of the Society Antonín E. Komers (1814–1893) in the year 1855. Negotiations continued with the support of Governor´s Office and the plenary meeting of PES on 20th April 1856 decided to establish the Central Committee on Statistics of Agriculture and Forestry (Central–Comité für die Land– und Forstwirthschaftliche Statistik Böhmens), which had its first meeting on 30th November 1856.[8] The Central Committee had originally 6 members, including the chairman. One of them was professor of statistics at Prague University Eberhard A. Jonák (1820–1879). He was the author of the action plan of the Central Committee and of the Statistical Chair (miniature office). The plan was approved in the year 1857. The following year the Statistical Chair started its activity under the leadership of prof. Jonák. At the beginning it only employed 2 persons and the number of the staff grew only moderately.

The establishment of the decision-making body (the Central Committee) and the Statistical Chair as an executive body took place at the recommendation on the 2nd International Statistical Congress in Paris (1855) according to the example of Quètelet from Brussel.[9] For the data collection, a system of unpaid delegates (i.e., reporters, two of each political district) and sub-delegates (for smaller territorial units) – economic officials, landowners, foresters, pastors, etc., was gradually developed. Activities of the Statistical Chair were financed mainly from the land budgets, partly by Governor´s Office.

The result of the Statistical Chair's activities, in particular the summary tables, were presented by Governor´s Office of Vienna's offices and usually published in the journals of PES (*Centralblatt* and *Verhandlungen* – see References). These were mainly the results of harvesting of individual crops and the structure of soil. The summary table report about crop yield results according to a new modern pattern was, at the request of the Viennese Department of Agriculture, processed annually, starting with data for the year 1868.

---

6 *Neue Schriften,* Bd. II, H. 1, p. 183.

7 *Verhandlungen,* 1853, Nr. 9, pp. 68–69. Former authors report incorrectly a year 1842 *(Zpráva o činnosti 1867*, p. 6), or 1852 – Krejčí (1905, p. XIII) etc.

8 This date is mistakenly referred to as "the beginning of organized statistics in Bohemia" – Podzimek (1974, p. 28 and elsewhere). The Industrial Unity in Bohemia set up its statistical chair already in year 1841, see Závodský (1992, p. 98).

9 See also *Zpráva o činnosti 1867*, p. 12.

In addition to news in journals of PES the Central Committee with its Statistical Chair started to issue in the year 1861 a thirteen-volume sources work *Tafeln zur Statistik der Land– und Forstwirthschaft des Königreiches Böhmen* (12 books of about 160 pages devoted to individual regions had been published in the years 1861–1872, last one in year 1881). For individual municipalities and summary for the districts and the whole region the data about the soil, its distribution according to property conditions, agricultural use, population numbers and cattle states can be found there. Remarkable is a range of calculated proportional numbers. When publishing this work, it was considered to indicate the individual volumes by characterizing the geographic conditions of the respective region. The Central Committee turned to a leading specialist in this field in Bohemia, Karel Kořistka (1825–1906), professor of Prague Polytechnic (predecessor of today's Czech Technical University – CTU).

Figure 1 Karel Kořistka in year 1881[10]

The consequences of this step were unexpected – all volumes were finally released without planned geographic introduction, but prof. Kořistka was pulled into the statistical work which became his main occupation for the next four decades. In addition to his work on Technique, he processed various expertise reports, e. g. for the reorganization of high technical education in the monarchy etc. In the year 1864 Kořistka became a member of the Central Committee and after the resignation of prof. Jonák (at the end of 1867) on the insistence of the entire committee, he accepted the head position of the Statistical Chair. This Chair under his leadership, despite its modest staffing and financial security, carried out agile activities, which were not limited only to the issues of agriculture and forestry. Since the year 1867 the Central Committee published annually both Czech and German reports on its activities.[11] They included an overview of activities in the previous year and table surveys of agricultural production in Bohemia as well as verbal analyses of the examined facts.

Statistical Chair was located in the basement of PES at the Wenceslas square No. 799/II (opposite the today's hotel Jalta). Beside prof. Kořistka the recognised expert and author of many reports and analysis, was JUDr. Josef Bernat (1834–1907), who worked for Statistical Chair in years 1866–1904.

Unlike the vast majority of former (and later) representatives of statistics in Bohemia and whole monarchy, Kořistka was not a graduated lawyer, but mathematician and surveyor. He has become (among others) a pioneer in using modern statistical charts. Already in *Zpráva o činnosti 1868* he published six colourful cartograms. He made similar charts also for some other issues of the *Zprávy*... showing not only the yields of different crops by district, but also the distribution of damage due to natural disasters, the intensity of afforestation of individual districts, the ratio of the number of beef cattle to 100 inhabitants etc.

Starting with *Zprávy* for the year 1872, Kořistka published in some annual issues the graphical charts showing long-term time series, most often the price development of various agricultural commodities.

[10] Photo by J. N. Langhans, Bibliothèque nationale de France [online]. (http://gallica.bnf.fr/ark:/12148/btv1b8450251f/f1.item).

[11] The reports were always published in Czech and German versions; in References we named the Czech volumes only.

Kořistka recommended to show price development for several different products in one graph that on vertical axis is a logarithmic scale better reflecting the relatively faster rise in price of cheaper commodities (e. g. potatoes) against a relatively slower rise in price of more expensive commodities (e. g. wheat).[12]

Figure 2 House No. 799/II on Wenceslas square (today's state)[13]

[12] *Zprávy kanceláře… for year 1872*, Iss. II, pp. LXII-LXIV.
[13] Photo by V. Puci.

Let us return to the question of collecting crop data through a network of delegates (see above). The amount of the harvest was estimated on the basis of an estimate of sowing area (here it was possible to rely on solid data from the cadastral maps) and an estimate of the yield of the crop in the given year. Estimates of yields of individual crops were then confronted with results from other municipalities and districts. To put such control on a solid foundation, Kořistka divided the entire territory of Bohemia into 11 so-called natural landscapes, that always included several judicial districts with similar conditions for agriculture (terrain altitude and shape, soil conditions, climate and the most frequently cultivated crops). This division of Bohemia allowed, among other things, for one or two districts to take into account the results of all the districts of the same "landscape". The division that Kořistka described in *Zpráva o činnosti* for the year 1871 (pp. VII–XIX) became an example for similar division of other monarchy countries. In Bohemia it has been for over half a century and was revised based on changes of the situation only in the year 1923.[14]

When the PES with all its components was officially dissolved in the spring of 1872 after 102 years of Governor´s Office for political reasons, the Statistical Chair was preserved and continued its activities, the Central Committee on Statistics of agriculture and forestry was shut down. The Vienna government established The Agricultural Council for Bohemian Kingdom the following year and a Statistical Chair was attached to it. The Agricultural Council was then divided into the Czech and German departments with the Centre for Joint Matters in 1891. The office was subordinated to them as the Statistical Chair of the Agricultural Council for Bohemian Kingdom.

In the 50s of 19th century it began, based on the initiative of the famous Belgian statistician A. Quètelet, to develop international cooperation of statisticians. Prof. Jonák took part already in the 2nd International Statistical Congress in Paris (1855). On the 3rd Congress in Vienna (1857) prof. Kořistka worked in the graphic display section of statistics and attracted the attention to his proposal on unification of large city plans (other participants were: prof. Jonák and L. Brdiczka, predecessor of J. Bernat as Secretary of the Statistical Chair). Kořistka participated in meetings of other congresses – in the Hague (1869 – referred to agricultural statistics in Bohemia), in St. Petersburg (1872) and in Budapest (1876 – here he successfully demonstrated forest statistics in Bohemia and its own methodology of statistical graphs). He also presented the success of Bohemian land statistics as a member of the International Statistical Institute (ISI – since year 1889), especially at sessions held in Christiania (todays Oslo – 1899) and in Budapest (1901).

3 FOUNDING OF LAND STATISTICAL OFFICE IN BOHEMIA

To understand lengthy negotiations about the origin of Land Statistical Office, let us briefly remind some facts about government and administration in Bohemia in last quarter of the 19th century. The decisive power was still in hands of Vienna's offices and their extended hand – Governor´s Office in Prague. Beside them, in Bohemia there existed (like in other countries of the monarchy) the Landtag with very limited powers. It consisted of 242 deputies, of which 236 were elected in three curies; there was no general or equal voting right.[15] The election period was six years (unless the Landtag was dissolved earlier). The administrative and executive body of the Landtag consisted of the 9-member Provincial Committee. The Landtag met in the former Thun Palace at Malá Strana (Lesser Town) in Prague, today's seat of the Chamber of Deputies, the appearance of the conference hall has not changed much since then – see Figure 3.

---

[14] See Závodský (1997a, pp. 101–102).

[15] E.g. during the election in 1895, 7.68% of the Czech population had the right to vote. See *Zprávy zemského* ... Vol. I, Iss. I, Prague, 1899, p. XLI.

Figure 3 Conference Hall of the Landtag in Bohemia (perhaps in the year 1898)[16]

The Provincial Committee occasionally organized various statistical surveys, sometimes directly requested by the Landtag. The survey related work was entrusted to their own non-specialized officials. Given that the Agricultural Council's Statistical Chair was partly financed from the Land budget, the Landtag sometimes asked the chair to carry out some statistical surveys (e. g. about the finances of the districts and municipalities in Bohemia in year 1883 etc.). Already around the year 1870 the calls for the establishment of the Land Statistical Office, were recorded, or better to say, the opinions requiring that the statistical chair of PES (later the Agricultural Council) should be taken over by the Provincial Committee appeared.

This issue had been discussed in the end by the four Landtags for two decades. It was for the first time that the Landtag approved the takeover of the Statistical Chair directly by the provincial administration in year 1878. It was only three years when it was decided to postpone the matter. The same proposal, which was presented and thoroughly reasoned at the meeting on 11th January 1886 by the Czech national economist and deputy prof. Albín Bráf,[17] was adopted again by the Landtag, but not implemented.

[17] The Landtag of the Kingdom of Bohemia 1883–1889, 3rd meeting, 19th session. In: *Společná česko-slovenská digitální parlamentní knihovna – stenoprotokoly.*

The situation slowly improved in the 90s when Karel Adámek (1840–1918) took the initiative. He was an extremely agile National Liberal Party deputy.[18] The proposal for the establishment of Land Statistical Office (similar to an earlier proposal by prof. Bráf) was submitted on behalf of the Group of Deputies on 21st March 1892, again on 10th April 1893. At the meeting held on 2nd May 1893, he made a substantiated statement of his draft, emphasized the need for proper statistics for the preparation of administrative reforms and for other decisions of the assembly.[19] Adámek's proposal was joined by member of Landtag – Masaryk already in the year 1892, who originally planned to present a similar proposal.[20]

Figure 4 Karel Adámek (perhaps in the year 1895)[22]

On 2nd May 1893, the Landtag discussed the proposal of Adámek on the 1st reading and forwarded it to the Budget Committee. The same procedure took place one year later, and only on 12th January 1895, after the third and last reading the matter started to move.

The recommendation of the Conference on Land Statistics was a drive. At the beginning of the year 1894 the Landtags in Silesia and Moravia suggested, that the representatives of the individual provincial authorities of the Austrian part of the monarchy should convene, together with the Central Statistical Commission in Vienna, a conference, in order to discuss the unification of methodology in statistical surveys in different countries, as well as the possibilities of co-operation of land statistics with the Central Statistical Commission. In November 1894, the gathering took place in Vienna chaired by the head of the Central Statistics Commission Inama-Sternegg, inaugurating his twenty-year activity at the Conference on Land Statistics (Konferenz für Landesstatistik).[21]

This conference, where knight Kořistka[23] regularly took part as a deputy of Bohemia, used to hold its meetings in Vienna in the beginning, (since year 1904 it changed places and was held in various countries of Austrian part of the empire (e.g. in 1907 the 9th meeting took place in Brno)). Already at the meeting in year 1895 the participants agreed on some unified tables, according to which statistical surveys should be held in all countries since year 1897. In the following years, the conference recommended a gradual increase in the scale of land self-government statistics according to the agreed single consensus.

[18] Karel Adámek, landowner and autodidact, was a politician National Liberal Party (Mladočeši) and a journalist, member of Landtag continuously in years 1881–1913. From a long series of his publications (often popularly learned) we can name a 98-page brochure *O statistice (About Statistics)* – Adámek (1899), where the author explains the importance of statistics for science and, in particular, state and provincial administrations, describes a long way, leading eventually to the formation of the Land Statistical Office, explains the conditions of its operation etc.

[19] See *Společná česko-slovenská digitální parlamentní knihovna – stenoprotokoly.*

[20] An undated handwritten draft of Masaryk printed in publication by Herben (1926), p. 176a. Masaryk spoke about the need for the Land Statistical Office in his speech (on another matter) on 7th April 1892. T. G. Masaryk was a member of Landtag since December 1891 until September 1893.

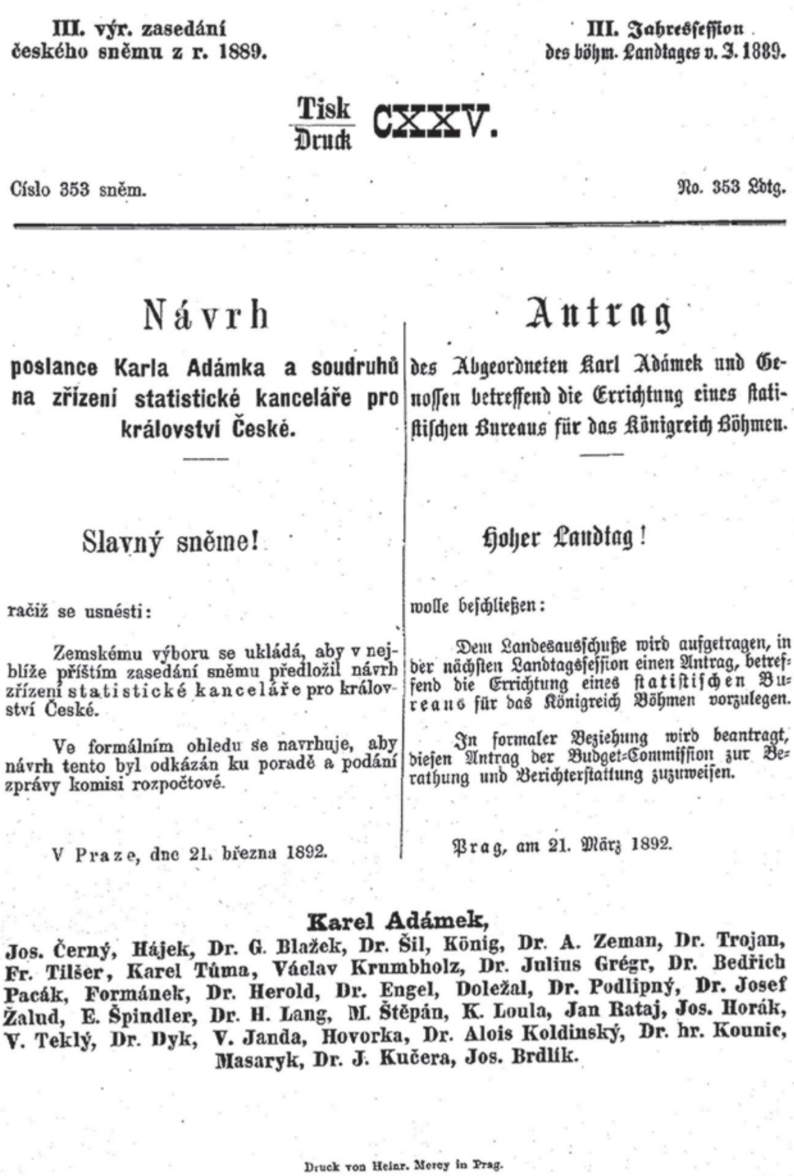
[21] See Berthold (1909).

[22] *Zlatá Praha*, 1895, Vol. XII, Iss. 28, p. 328.

[23] Prof. Kořistka was nobilitated by the emperor in year 1879.

The resolution of the conference was an important impetus for the establishment of statistical offices in Bohemia, Moravia and Silesia. In the years 1900–1918, thanks to Vienna's Central Statistical Commission, the total of 14 volumes of statistical yearbook of land self-governments (*Statistisches Jahrbuch der autonomen Landesverwaltung in den im Reichsrathe vertretenen Königreichen und Ländern*) were published. Statistical offices in the Czech lands were among the most active contributors.

Figure 5 Adámek's proposal from 21st March, 1892[24]



III. výr. zasedání
českého sněmu z r. 1889.

III. Jahressession
des böhm. Landtages v. J. 1889.

Tisk / Druck CXXV.

Číslo 353 sněm.

No. 353 Ldtg.

Návrh

poslance Karla Adámka a soudruhů na zřízení statistické kanceláře pro království České.

Slavný sněme!

račiž se usnésti:

Zemskému výboru se ukládá, aby v nejblíže příštím zasedání sněmu předložil návrh zřízení statistické kanceláře pro království České.

Ve formálním ohledu se navrhuje, aby návrh tento byl odkázán ku poradě a podání zprávy komisi rozpočtové.

V Praze, dne 21. března 1892.

Antrag

des Abgeordneten Karl Adámek und Genossen betreffend die Errichtung eines statistischen Bureaus für das Königreich Böhmen.

Hoher Landtag!

wolle beschließen:

Dem Landesausschuße wird aufgetragen, in der nächsten Landtagssession einen Antrag, betreffend die Errichtung eines statistischen Bureaus für das Königreich Böhmen vorzulegen.

In formaler Beziehung wird beantragt, diesen Antrag der Budget-Commission zur Berathung und Berichterstattung zuzuweisen.

Prag, am 21. März 1892.

Karel Adámek,
Jos. Černý, Hájek, Dr. G. Blažek, Dr. Šil, König, Dr. A. Zeman, Dr. Trojan, Fr. Tilšer, Karel Tůma, Václav Krumbholz, Dr. Julius Grégr, Dr. Bedřich Pacák, Formánek, Dr. Herold, Dr. Engel, Doležal, Dr. Podlipný, Dr. Josef Žalud, E. Špindler, Dr. H. Lang, M. Štěpán, K. Loula, Jan Rataj, Jos. Horák, V. Teklý, Dr. Dyk, V. Janda, Hovorka, Dr. Alois Koldinský, Dr. hr. Kounic, Masaryk, Dr. J. Kučera, Jos. Brdlík.

Druck von Heinr. Mercy in Prag.

24 IIIrd Annual Meeting of the Landtag (1891–1892), press CXXV.

Kořistka submitted to the Provincial Committee in December 1895 a report on the recommendations of the first two Viennese sessions as well as the rationale for the proposal on establishment of the Land Statistical Office. On the 14th February 1896 the Landtag[25] discussed Adámek's proposal and based on the recommendation of the Trade Commission to the Landtag it was approved at the second reading. The Provincial Committee was required to submit an outline of the organizational statute of the Land Statistical Office. Kořistka then drafted a detailed proposal of the statute and submitted it to the provincial committee. The provincial committee submitted it to the Landtag on 6th February 1897. Following the approval by the Budget Committee, the proposal was discussed at the meeting of the Landtag on 6th March 1897. Adámek again briefly justified the need to set up an LSO, which will be capable to conduct proper data collection and, in particular, a qualitative data analysis, he presented the success of similar offices in Styria and Galicia and expressed the belief that in the future the LSO's competencies and program of activities will expand. Establishment of Land Statistical Office, its statute and funding were finally approved by the Landtag on 6th March 1897.

Land Statistical Office consisted (according to the proven Quèletlet's pattern) of the Land Statistical Commission and the Land Statistical Bureau. Statistical Commission was an advisory body (to the Provincial Committee) and decisive on the Land Statistical Bureau (as the executive body). In a 10-member Statistical Commission, headed by an elected representative of the Land Committee by the earl Vojtěch Schönborn,[26] was made by deputies of Governor´s Office, Agricultural Council, member of Provincial Committee Adámek, economists from both Prague universities (prof. Bráf and prof. Ulbrich), Head of the Land Statistical Bureau Kořistka etc. The Commission usually met once or twice a year.

The staff of the Statistical Bureau, taken from the Agricultural Council, was joined by a young lawyer JUDr. Dobroslav Krejčí (1869–1936) and three other officers, i.e. 8 employees plus chief in total The Bureau was seated in the House of Agricultural Council at Wenceslas square, that had been partly rebuild at that time. Statement of the publication of the Czech Statistical Office *Statistics: From history to present* (2006, p. 10) "This was the first centralisation of all statistical units that had functioned within the different ministries and other institutions so far."[27] is a complete nonsense.[28]

The Land Statistical Office was given according to § 1 of its statute the quest "to cultivate statistics in all matters governed by the laws of the land in force." D. Krejčí said: "It was limited in its competencies to mere leftovers after the state statistics."[29] It cannot obtain other competencies due to the law. At the beginning of 20th century it was considered the extension of the authority of the provincial self-government, but with regard to the old emperor's mentality and conflicting relations in the Imperial Council, this has never been the case.

In Section 2 of the Status of LSO there is a limited scope of the office activities exhaustively named. LSO should deal with statistics of:

1. agriculture and forestry,
2. public buildings carried out by the cost of the country (roads, railways, water structures, etc.),
3. medical and charitable institutes, subsidized by land resources,
4. schools, artistic and scientific institutions, held in full or in part by land resources,
5. land, district and municipal finances,
6. land banks, saving banks and saving institutions,

---

25 It was again a new assembly, elected in November 1895.
26 He was the head of the Land Statistical Commission throughout its functioning.
27 Similarly, in the Czech version *Statistika: Od historie po současnost* (2006, p. 10). There are many mistakes in the quoted text. In the same paragraph, for example, the mother tongue is mistaken.
28 Even an ordinary high school student knows that there were no ministries in Prague during the period of Austro-Hungarian Empire.
29 Krejčí (1920, p. 84).

7. election to Landtag,
8. police matter under the jurisdiction of the country,
9. military matters (if they were under the authority of the provincial authorities).

Provincial Committee as a superior body of LSO could request other statistical surveys "in the interest of the country". The statute did not exclude statistical work for the state authorities, subject to the approval of the Provincial Committee and the reimbursement of costs by the contracting authority.

Figure 6 Statute Land Statistical Office approved by the Landtag on 6th March 1897 (p. 1)[30]

Příloha 2/2. XXI. sezení 1896/97. 33 XXI. Sitzung 1896/97. Beilage 2/2.

Statut

statistitkého zemského úřadu

pro

království České.

I. Úkol a složení.

§ 1.

Statistický úřad zemský má za úkol, aby pěstoval statistiku ve všech záležitostech, které samosprávě zemské platnými zákony jsou přikázány. Podřízen jest výboru zemskému.

§ 2.

K záležitostem těmto náležejí tudíž:

1. Veškeré zemědělství (polní a lesní hospodářství, pozemkový majetek, hospodářská výroba, hospodářský úvěr a spolčování).
2. Veřejné stavby podnikané nákladem zemským (stavby vodní, silnice, mosty a železnice).
3. Z peněz zemských nadané ústavy dobročinné (nemocnice a ústavy humanitní, nadace pro ně, opatrování chudých).
4. Z peněz zemských zcela neb z části vydržované ústavy vyučovací, dále ústavy pro umění a vědu, musea, knihovny atd.
5. Daně a finance zemské, okresní a obecní (přirážky).
6. Zemské úvěrnictví (zemská banka, hypotheční banka, záložny).
7. Volby do sněmu zemského.
8. Policejní záležitosti (postrk, stravovny, donucovací pracovny a polepšovny).
9. Vojenské záležitosti (ubytování, vojenské nadace).

Mimo to může výbor zemský statistickému úřadu zemskému uložiti také jiná statistická

Statut

des statistischen Landesamtes

für das

Königreich Böhmen.

I. Aufgabe und Gliederung.

§ 1.

Das statistische Landesamt hat die Aufgabe, die Statistik in allen jenen Angelegenheiten zu pflegen, welche der autonomen Verwaltung des Landes durch die bestehenden Gesetze zugewiesen sind. Dasselbe untersteht dem Landesausschusse.

§ 2.

Zu diesen Angelegenheiten gehören demnach:

1. Die gesammte Landeskultur (land- und forstwirthschaftlicher Grundbesitz, landwirthschaftliche Produktion, landwirthschaftliches Kredit- nnd Vereinswesen).
2. Die öffentlichen Bauten aus Landesmitteln (Wasserbauten, Strassen, Brücken und Eisenbahnen).
3. Die aus Landesmitteln dotirten Wohlthätigkeitsanstalten (Kranken- und Humanitätsanstalten, Stiftungen für dieselben, Armenpflege).
4. Die aus Landesmitteln ganz oder zum Theile erhalteuen Unterrichtsanstalten, dann Anstalten für Kunst und Wissenschaft, Museen, Bibliotheken u. s. w.
5. Die Besteuerung nnd die Finanzen des Landes, der Bezirke nnd Gemeinden (Zuschläge).
6. Das Landeskreditwesen (Landesbank, Hypothekenbank, Vorschußkassen).
7. Landtagswahlen.
8. Polizeiangelegenheiten (Schubwesen, Naturalverpflegsstationen, Zwangsarbeits- und Besserungsanstalten).
9: Militärangelegenheiten (Einquartirung, Militärstiftungen).

Außerdem kann der Landesausschuß das statistische Landesamt auch mit anderen statistischen

5

[30] The negotiating memorandum drafted at the 21st meeting of the IInd Annual Meeting of the Bohemian Kingdom from the year 1895, p. 33.

4 LAND STATISTICAL OFFICE IN BOHEMIA 1898–1918

The Land Statistical Bureau started operations at the beginning of the year 1898 under the proved supervision of head knight Kořistka and secretary JUDr. Bernat. The main focus of the work was agricultural statistics (including a brief analysis of meteorological observations in the reference year). Detailed analysis of election results in the Landtag (1895 and 1901) was added, the same as the overviews of land finances, education etc. Starting with the year 1899 the Statistical Bureau annually published in Czech and German the *Zprávy Zemského statistického úřadu království Českého* (News of the Land Statistical Office of Bohemian Kingdom; usually one volume in two issues).

Figure 7 Dobroslav Krejčí (in 20s)[31]

In the following years, the management of the statistical chair changed. At the end of the year 1904 J. Bernat died after a long sickness and one year later also the head K. Kořistka (80 years old at that time). The management of the Bureau was taken over by D. Krejčí, newly were accepted new officials dealing with conceptual matters JUDr. Karel Engliš (1880–1961, in LSO 1904–1908), JUDr. Vilibald Mildschuh (1878–1939, in LSO 1904–1917) and JUDr. Jan Auerhan (1880–1942, in LSO 1906–1919).

It is worth mentioning that there was no statistical study at that time in the monarchy. Universal qualifications for such officials in various state and provincial offices was graduation from the law faculty. Statistics was a marginal subject and had the reputation of a very boring discipline, the lectures rather avoided listeners, the methods of statistical work were not lectured.[32] The above-mentioned persons entered the LSO practically with zero knowledge of statistical work. Theoretical knowledge was complemented by an individual study of literature, especially in German. D. Krejčí after joining the LSO passed the internship from the Central Statistical Commission in Vienna (Inama-Sternegg was still the head) and at the Land Statistical Office in Styria Graz, (led by Prof. Mischler).[33]

In the following years, the number of the staff of the statistical Bureau continued to increase – to 29 in 1909 and 37 at the time of the greatest development before World War I. Mainly as a result of national disputes in Bohemia, also the problems grown in the last ten years before the outbreak of the war. The personality of the knight Kořistka, an emeritus professor of the German Technical University, was acceptable to both nationalities, while the Czech patriot D. Krejčí not. No compromise was reached – the place of the head of the Statistical Bureau remained permanently vacant, and Krejčí managed the office as the deputy head and secretary. Many German districts and municipalities have explicitly refused to provide data until the German was promoted as a head (the reporting duty here, as already mentioned, did not exist). National disputed and obstructions blocked for a long time also the Landtag, and since

[31] Weyr (1937, p. 1).

[32] Lecturers who did not have statistics as major specialization usually read out various aggregate data, particularly about the population and economy in the monarchy and abroad.

[33] Prague native Ernst Mischler (1857–1912), son of prof. Peter Mischler, referred in § 2, studied law in Prague and later lectured here. From 1911 to 1912 he was then the President of the Central Statistical Commission in Vienna.

1913 not meeting of the Land Statistical Commission had been held. Krejčí himself did not contributed to the moderation of national disputes, as he promoted conceptual officials, persons of indisputable qualification and perspective, but only of Czech nationality. Newly it was: JUDr. Josef Mráz (1882–1934, in LSO 1909–1919) and JUDr. František Weyr (1879–1951, in LSO 1909–1912), the only exception was Wilhelm Winkler (1884–1984, in LSO 1909–1914).

In 1905, the Land Statistical Bureau moved from the seat of the Agricultural Council to Malá Strana (Lesser Town) to the rent house on Dražický Sq. No 10/65. In 1912 it rented the premises in the new building in Šeříková Street No. 4/618 (Malá Strana).

The Land Statistical Bureau at that time showed ever more extensive activity, publishing included. Each year, several, often very large workbooks appeared in both language versions: *Zprávy Zemského statistického úřadu* (News of Land Statistical Office). Processing of agricultural statistics was taken by D. Krejčí and W. Winkler after Kořistka. J. Auerhan was specialized particularly in statistics of the inhabitants and self-government in land, K. Engliš focused mainly on education and social care, V. Mildschuh and J. Mráz on economic and financial statistics etc.

Two issues were of utmost importance: *Statistická příručka království Českého* (Statistical handbooks of Bohemian Kingdom) (1909 – 499 p.; 1913 – 540 p.), publishing in hundreds of tables various available data on Bohemia, for comparison often with similar data from Moravia, Silesia and the entire Austrian part of the monarchy. In many cases, comparable older data were also published.

Land Statistical Bureau continued its activities to a somewhat limited extent during the World War I., as the Vienna authorities were particularly interested in statistics on agriculture and food production. These data in a small number were also published in the printed *Zprávy* (Reports), the last editions were issued only after 1918, when the Land Statistical Bureau was taken over by the new State Statistical Office.

5 LAND STATISTICS IN MORAVIA AND SILESIA

Let's look briefly at the statistical service of the Land Self-Government in Moravia, which existed for only 15 years and its activity was weaker. Agricultural statistics in Moravia in 19th century was made by historical–statistical section of Economic society, the agricultural council was only established here only in the year 1897. It took over the harvest statistics, which received a flat-rate contribution from the Viennese Department of Agriculture. At that time, the issue of establishment of the Land Statistical Office was discussed and the Landtag initiated a meeting in Vienna on the unification of the autonomous statistical service methodology in individual countries (Conference on Land Statistics – see above).

Representatives of the Provincial Committee visited in the years 1897–1898 the Land Statistical Offices in Styria and in Bohemia and after further negotiations the Landtag decided to establish the Statistical Office of the Moravian Margraviate on 24th March 1899.[34] The announcement was sent on 4th May 1899 to the representatives of all municipalities in the country with the bilingual list of the High Representative of the Land Self-government, by the Landeshauptmann earl Vetter z Lilie (see Figure 8).[35]

The Office started its activity on 16th April 1899. The Provincial Committee appointed Albert Bervid, who has already been involved in the preparation for the establishment of the LSO in previous years. The initial status of four officials increased gradually in the following years, with a maximum of 10. The Office moved several times to various buildings around the Chamber of Deputies in Joštova street (the Constitutional Court is housed here today). LSO eventually released only three statistical publications – bilingual on credit institutions in Moravia (1902) and two large-format publications on election results in year 1906 – in Czech on the election of Czech deputies and in German on the election of German deputies (1907). Other publishing plans have failed due to a lack of finance and qualified staff.

[34] The usually stated year 1893 is wrong – see Podzimek (1974, p. 53) and elsewhere.
[35] See Závodský (2004), where are also citations from the Moravian Regional archives.

Figure 8 Announcement to the Moravian municipalities about the origin of Land Statistical Office and on the obligations of municipalities towards to it[36]

Mährischer Landesausschuss.

Z. 25446 ai 1899.

An die Gemeindevorstände der sämmtlichen Gemeinden in Mähren und in den mähr. Enclaven in Schlesien.

Zufolge Beschlusses des mähr. Landtages vom 24. März 1899 wurde behufs systematischer Führung der Landesstatistik ein eigenes Amt in Brünn unter dem Titel „Statistisches Landesamt der Markgrafschaft Mähren" errichtet. Die Aufgabe dieses Amtes ergibt sich:

1. aus der Sammlung und Verarbeitung landesstatistischer Daten in dem Umfange und nach der Methode, wie diese durch die Beschlüsse der Wiener Conferenzen für Landesstatistik jeweilig festgestellt werden;
2. aus besonderen Aufträgen des m. Landesausschusses, und
3. aus dem Umkreise der sich durch die sonstigen, in den Tabellen der Wiener Conferenzen nicht inbegriffenen Landesangelegenheiten bestimmenden „Landesstatistik".

Die Erhebungen und Arbeiten des statistischen Landesamtes werden zum großen Theile verschiedene Gemeindeangelegenheiten zum Gegenstande haben (z. B. Armenwesen, Sanitätspflege, Humanitätsanstalten, Stiftungen, Schulen, Gemeindefinanzen u. s. w.) und es werden bezüglich eines jeden in die Erhebung einbezogenen Gegenstandes besondere Fragebögen seitens des statistischen Landesamtes versendet werden.

Hievon werden die Gemeindevorstände aller Gemeinden Mährens und der mähr Enclavgemeinden Schlesiens mit dem Auftrage verständigt, die ihnen seitens des statistischen Landesamtes von Fall zu Fall zugesendeten Fragebögen stets gewissenhaft auszufüllen und so rasch als möglich abzuliefern.

Brünn, am 4. Mai 1899.

Der Landeshauptmann:

Felix Graf Vetter von der Lilie.

Moravský výbor zemský.

Čís. 25446 r. 1899.

Představenstvům veškerých obcí na Moravě a v mor. obvodech ve Slezsku.

Následkem usnesení sněmu markrabství Moravského ze dne 24. března 1899 zřízen byl v Brně pro soustavné vedení zemské statistiky vlastní úřad s titulem **„Statistický zemský úřad markrabství Moravského"**. Úloha tohoto úřadu záleží:

1. ze sbírání a zpracování zemských statistických dat v tom rozsahu a dle té methody, jak to usneseními vídeňských konferencí pro zemskou statistiku občasně bude určováno;
2. ze zvláštních, na rozkaz zemského výboru konaných šetření;
3. z oboru „zemské statistiky", vyplývající z ostatních záležitostí, jež nejsou v tabelách vídeňských konferencí obsaženy.

Vyšetřování a práce statistického zemského úřadu budou se z velké části vztahovati také na záležitosti obecní (ku př. chudinství, zdravotnictví, humanitní ústavy, nadace, školy, obecní finance atd.) a budou statistickým zemským úřadem v příčině každého předmětu, o němž vyšetřování konati se bude, dodány obecnímu představenstvu zvláštní dotazníky.

O tom dává se obecním představenstvům všech obcí na Moravě a v obvodních obcích Slezska věděti a nařizuje se, aby dotazníky, které od případu k případu statistickým zemským úřadem jim budou dodány, vždy svědomitě vyplněny a co nejrychleji odeslány byly.

V Brně, dne 4. května 1899.

Zemský hejtman:

Felix hrabě Vetter z Lilie.

Hope to improve the state of affairs emerged after the election to the Landtag in June 1913. One of the new deputies was JUDr. Karel Engliš (1880–1961), at that time an extraordinary professor of the national economy and statistics at the Czech Technical University in Brno. On 3rd February 1914 Engliš

[36] *Moravský zemský archiv* (Moravian Regional archive), carton 178.

passed with the other deputies to the Landtag a draft of statistical law for Moravia. It established (for the first time in the territory of the Czech Lands) a general reporting obligation for all investigations ordered or approved by the Provincial Committee. This obligation applied to municipalities and District Road Committees.[37] Engliš's proposal was approved with a slight change both by the Landtag and by the emperor. It was published in the land code on 18th December. However, the practical consequences of this law were none, because at the same time with the outbreak of the World War, the Landtag and Land Statistical Office were closed. Another Engliš's proposal on issuing new LSO statute could not have been discussed.

Even the smallest of the countries of Cisleithania – Silesia has established its own statistical service. The first proposal in this regard came from the Member of the Landtag, Baron Spens von Boden, who already in 1887 proposed the establishment of the Land Statistical Service and the annual publication of the land statistical yearbook. The Landtag approved this proposal in principle, but for many years there was discussion about the legal and organizational classification of the Land Statistical Service, its relation to the Vienna Central Statistical Commission, etc. As mentioned above, the initiative of the Silesian and Moravian Landtag of January 1894 led to the Conference on Land Statistics.

Establishment of the Land Statistical Service (Landesstatistischer Dienst) in Opava was approved by Landtag on 26th January 1898 and this service started on 1st May that year. From the beginning it was headed by Dr. Karl Berthold (has been formally appointed by the Provincial Committee on 6th February 1900). Initially, the office had only two workers, later one more was accepted. On 20th February 1907, its office became an independent part of the Provincial Office as the Land Statistical Office (Landesstatistisches Amt).[38]

The Office had been carrying out various surveys since 1898, both according to the recommendations of the Conference on Land Statistics and on the instructions of the Provincial Committee. It was also preparing the publication of the Silesian statistical handbook, which began to be published in 1899 – until 1905 every year, then only irregularly. In total, 10 volumes were published (the last one after the formation of Czechoslovakia in 1919) under the name *Statistisches Handbuch für die Selbstverwaltung in Schlesien*. All publications of Silesian statistics were published in German only up to year 1918.

The small Silesian statistical office continued its work after the birth of Czechoslovakia, cooperating on statistical data for negotiations on the new border also with Poland (division of Teschen land). It published two more extensive works on the population of Silesia on the basis of the Population Census of 1921 and on the development and up-to-date status of education in Silesia. When it turned out that the new Czechoslovak statistical office did not count with any out-of-Prague branches, the Land Statistical Office of Silesia was closed down in June 1925. K. Berthold was then an active member of the State Statistical Council and Czechoslovakian statistical society.

CONCLUSION

In Bohemia, the Land Statistical Office existed for 21 years. The activities of its Statistical Bureau were limited by competencies, allocated funds and difficulties due to the national disputes. Nevertheless, this activity can be considered as successful.

The Land Statistical Office under the leadership of K. Kořistka and mainly of D. Krejčí elaborated a number of proper statistical analyses, and published, in addition to a long series of (often very extensive) workbooks *Zprávy (Reports),* also two issues of *Statistická příručka království Českého (Statistical handbook*

[37] Unlike in Bohemia, in Moravia the autonomous authorities at the district level did not exist. There were only district road committees, organizing only the construction and management of district roads.

[38] See Závodský (2004, pp. 298–299).

of the Kingdom of Bohemia). Future university professors (K. Engliš, V. Mildschuh, F. Weyr and W. Winkler) as well as representatives of the Czechoslovak State Statistics (F. Weyr, J. Auerhan and Josef Mráz) went through the statistical practice here.

Czech statisticians in January 2019 will celebrate the 100th anniversary of the establishment of the State Statistical Office of the new Czechoslovak Republic by the law, which was prepared by D. Krejčí and K. Engliš. The Land Statistical Office in Prague, with a number of experienced staff (including facilities and library), has become the basis for building a new office, whose direct successor is today's Czech Statistical Office. We are planning another historic treatise on that anniversary.

ACKNOWLEDGMENT

This paper was supported by the Internal Grant of the University of Economics Prague No. 39/2017 "The Initial Development of Activities of the State Statistical Office".

We thank to Mgr. Petr Houdek from the Parliamentary Library of the Czech Republic for his help with searching for documents.

References

ADÁMEK, K. *O statistice*. Chrudim: Eckert, 1899.

BERTHOLD, K. Die Entstehung, Entwicklung und Tätigkeit der Konferenz für Landesstatistik. *Statistische Monatschrift,* Neue Folge, 1909, Vol. 14, pp. 593–613.

Geschichte und Ergebnisse der zentralen amtlichen Statistik in Österreich 1829–1979. Wien: Österreichisches Statistisches Zentralamt, 1979.

HERBEN, J. *T. G. Masaryk I.* Prague: Nákladem Spolku výtvarných umělců Mánes, 1926.

KREJČÍ, D. Kořistka jako statistik. In: *Sborník Čs. společnosti zeměpisné,* 1925, Vol. XXXI, pp. 26–30.

KREJČÍ, D. Přehled vývoje správní statistiky v Čechách. *Čs. statistický věstník,* 1920, Vol. I, Iss. 1–2, pp. 37–49; Iss. 3–5, pp. 73–89.

KREJČÍ, D. Vznik a vývoj zemědělské statistiky v Čechách. In: *Zprávy zemského statistického* úřadu *království Českého,* 1905, Vol. VI, Iss. 2, pp. X–XXVI.

Letem českým světem: půl tisíce fotograf. pohledů z Čech, Moravy, Slezska a Slovenska. Prague: J. R. Vilímek, 1898.

Moravský zemský archive. Fond A 9 – Zemský výbor, karton 64, pp. 178–183.

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

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

Společná česko–slovenská digitální parlamentní knihovna – stenoprotokoly [online]. [cit. 15.12.2017]. <http://www.psp.cz/eknih/index.htm>.

Statistika: Od historie po současnost [online]. 2006. [cit. 15.12.2017]. <https://www.czso.cz/documents/10180/23169768/historie_csu.pdf>.

Statistics: From history to present [online]. 2006. [cit. 15.12.2017]. <https://www.czso.cz/csu/czso/history–of–the–statistical–office>.

WEYR, F. *Dobroslav Krejčí.* Prague: Nákladem České akademie věd a umění, 1937.

ZÁVODSKÝ, P. Karel rytíř Kořistka a statistika. In: *Mundus symbolicus*, 1997a, Vol. 5, pp. 99–105.

ZÁVODSKÝ, P. 100 let Zemského statistického úřadu Království českého. *Statistika*, 1997b, Vol. 34, Iss. 3, pp. 93–98.

ZÁVODSKÝ, P. Provincial Statistical Offices – Predecessors of the State Statistial Office. In: KLÍMA, M., eds. *Applications of mathematics and statistics in economy.* Prague: Professional Publishing, 2004, pp. 292–300.

ZÁVODSKÝ, P. *Vývoj statistické teorie na území Československa do roku 1848.* Prague: FSÚ and Infostat, 1992.

ZÁVODSKÝ, P. AND ŠIMPACH, O. Karel Kořistka and the Provincial Statistical Service in the Czech Lands. In: *Applications of Mathematics and Statistics in Economics 2016,* Banská Bystrica: Občianske združenie Financ, 2016, pp. 387–398.

Periodicals, published by the Patriotic-Economic Society, the Agricultural Council in Bohemia and by the Land Statistical Office:

Abhandlungen die Verbesserung der Landwirtschaft betreffend (1797–1806).

Neue Schriften der k. k. patriotisch–ökonomischen Gesellschaft im Königreiche Böhmen (1825–1847).

Centralblatt der Land– und Forstwirtschaft in Böhmen (1850–1852).

Centralblatt für die gesammte Landeskultur (1853–1870).

Centralblatt für die gesammte Landeskultur des In– und Auslandes (1871).

Verhandlungen und Mittheilungen der k. k. patriotisch–ökonomischen Gesellschaft in Böhmen (1849–1868, since 1850 it was published as an annex to the previous three cited periodicals).
Zpráva o činnosti ústředního výboru pro statistiku polního a lesního hospodářství v království Českém za rok … (1867–1872).[39]
Zprávy kanceláře pro statistiku polního a lesního hospodářství v království Českém za rok 1872, Iss. 1–2 (1872–1873).
Zprávy výboru pro statistiku polního a lesního hospodářství v království Českém za rok … (1874–1891).
Zprávy statistické kanceláře zemědělské rady pro království České za léta… (1893–1897).
Zprávy Zemského statistického úřadu království Českého, sv. I–XXVI (1899–1923).

[39] Beginning with the *Report of 1867*, all volumes were published in Czech and German versions.

Recent Publications and Events

New publications of the Czech Statistical Office

Consumer Price Indices 2017. Prague: CZSO, 2018.
Czech Republic in International Comparison (Selected Indicators). Prague: CZSO, 2018.
Export and Import Price Indices in the Czech Republic in 2017. Prague: CZSO, 2018.
Statistical Yearbook of Prague 2017. Prague: CZSO, 2017.

Other selected publications

A Concise History of World Population. 6^{th} Ed. Oxford: Wiley, 2017.
Applied Statistics for Social and Management Sciences. Singapore: Springer, 2016.
Main Economic and Social Indicators of the Czech Republic 1990–2016. Prague: VUPSV, 2017.

Conferences

The ***European Conference on Quality in Official Statistics (Q2018)*** will be held in **Kraków, Poland, during 26–29 June 2018.** More information available at: <*http://www.q2018.pl*>.

The ***21^{st} AMSE 2018 Conference*** will take place in **Kutná Hora, Czech Republic, from 28^{th} August to 2^{nd} September 2018.** More information available at: <*http://www.amse-conference.eu*>.

The ***26^{th} Interdisciplinary Information Management Talks (IDIMT 2018 Conference)*** will be held in **Kutná Hora, Czech Republic, during 5–7 September 2018.** More information available at: <*http://idimt.org*>.

The ***12^{th} International Days of Statistics and Economics (MSED 2018 Conference)*** will take place in **the University of Economics, Prague, Czech Republic, from 6^{th} to 8^{th} September 2018.** The conference is jointly organized by the Department of Statistics and Probability and the Department of Microeconomics, University of Economics, Prague, Czech Republic; Faculty of Economics, the Technical University of Košice, Slovakia; and the Ton Duc Thang University, Vietnam. The aim of the conference is to present and discuss current problems of statistics, demography, economics and management and their mutual interconnection. More information available at: <*https://msed.vse.cz*>.

The ***36^{th} International Conference on Mathematical Methods in Economics (MME 2018)*** will be held in **Jindřichův Hradec, Czech Republic, during 12–14 September 2018.** The conference is a traditional meeting of professionals from universities and businesses interested in the theory and applications of operations research and econometrics. More information available at: <*http://mme2018.fm.vse.cz*>.

Papers
We publish articles focused at theoretical and applied statistics, mathematical and statistical methods, conception of official (state) statistics, statistical education, applied economics and econometrics, economic, social and environmental analyses, economic indicators, social and environmental issues in terms of statistics or economics, and regional development issues.

The journal of *Statistika* has the following sections:

The *Analyses* section publishes high quality, complex, and advanced analyses based on the official statistics data focused on economic, environmental, and social spheres. Papers shall have up to 12 000 words or up to twenty (20) 1.5-spaced pages.

The *Discussion* section brings the opportunity to openly discuss the current or more general statistical or economic issues; in short, with what the authors would like to contribute to the scientific debate. Discussions shall have up to 6 000 words or up to 10 1.5-spaced pages.

The *Methodology* section gives space for the discussion on potential approaches to the statistical description of social, economic, and environmental phenomena, development of indicators, estimation issues, etc. Papers shall have up to 12 000 words or up to twenty (20) 1.5-spaced pages.

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

In the *Information* section we publish informative (descriptive) texts. The maximum range of information is 6 000 words or up to 10 1.5-spaced pages.

Language
The submission language is English only. Authors are expected to refer to a native language speaker in case they are not sure of language quality of their papers.

Recommended Paper Structure
Title (e.g. On Laconic and Informative Titles) — Authors and Contacts — Abstract (max. 160 words) — Keywords (max. 6 words / phrases) — JEL classification code — Introduction — … — Conclusion — Annex — Acknowledgments — References — Tables and Figures

Authors and Contacts
Rudolf Novak*, Institution Name, Street, City, Country
Jonathan Davis, Institution Name, Street, City, Country
* Corresponding author: e-mail: rudolf.novak@domain-name.cz, phone: (+420) 111 222 333

Main Text Format
Times 12 (main text), 1.5 spacing between lines. Page numbers in the lower right-hand corner. *Italics* can be used in the text if necessary. *Do not* use **bold** or <u>underline</u> in the text. Paper parts numbering: 1, 1.1, 1.2, etc.

Headings
1 FIRST-LEVEL HEADING (Times New Roman 12, bold)
1.1 Second-level heading (Times New Roman 12, bold)
1.1.1 Third-level heading (Times New Roman 12, bold italic)

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

References in the Text
Place reference in the text enclosing authors' names and the year of the reference, e.g. "White (2009) points out that…", "… recent literature (Atkinson et Black, 2010a, 2010b, 2011; Chase et al., 2011, pp. 12–14) conclude…". Note the use of alphabetical order. Include page numbers if appropriate.

List of References
Arrange list of references alphabetically. Use the following reference styles: [for a book] HICKS, J. *Value and Capital: An inquiry into some fundamental principles of economic theory.* 1st Ed. Oxford: Clarendon Press, 1939. [for chapter in an edited book] DASGUPTA, P. et al. Intergenerational Equity, Social Discount Rates and Global Warming. In: PORTNEY, P. AND WEYANT, J., eds. *Discounting and Intergenerational Equity*. Washington, D.C.: Resources for the Future, 1999. [for a journal] HRONOVÁ, S., HINDLS, R., ČABLA, A. Conjunctural Evolution of the Czech Economy. *Statistika: Statistics and Economy Journal,* 2011, 3 (September), pp. 4–17. [for an online source] CZECH COAL. *Annual Report and Financial Statement 2007* [online]. Prague: Czech Coal, 2008. [cit. 20.9.2008]. <http://www.czechcoal.cz/cs/ur/zprava/ur2007cz.pdf>.

Tables
Provide each table on a separate page. Indicate position of the table by placing in the text "<u>insert Table 1 about here</u>". Number tables in the order of appearance Table 1, Table 2, etc. Each table should be titled (e.g. Table 1 Self-explanatory title). Refer to tables using their numbers (e.g. see Table 1, Table A1 in the Annex). Try to break one large table into several smaller tables, whenever possible. Separate thousands with a space (e.g. 1 528 000) and decimal points with a dot (e.g. 1.0). Specify the data source below the tables.

Figures
Figure is any graphical object other than table. Attach each figure as a separate file. Indicate position of the figure by placing in the text "<u>insert Figure 1 about here</u>". Number figures in the order of appearance Figure 1, Figure 2, etc. Each figure should be titled (e.g. Figure 1 Self-explanatory title). Refer to figures using their numbers (e.g. see Figure 1, Figure A1 in the Annex).

Figures should be accompanied by the *.xls, *.xlsx table with the source data. Please provide cartograms in the vector format. Other graphic objects should be provided in *.tif, *.jpg, *.eps formats. Do not supply low-resolution files optimized for the screen use. Specify the source below the figures.

Formulas
Formulas should be prepared in formula editor in the same text format (Times 12) as the main text and numbered.

Paper Submission
Please email your papers in *.doc, *.docx or *.pdf formats to statistika.journal@czso.cz. All papers are subject to double-blind peer review procedure. Articles for the review process are accepted continuously and may contain tables and figures in the text (for final graphical typesetting must be supplied separately as specified in the instructions above). Please be informed about our Publication Ethics rules (i.e. authors responsibilities) published at: http://www.czso.cz/statistika_journal.

Managing Editor: Jiří Novotný
phone: (+420) 274 054 299
fax: (+420) 274 052 133
e-mail: statistika.journal@czso.cz
web: www.czso.cz/statistika_journal
address: Czech Statistical Office | Na padesátém 81 | 100 82 Prague 10 | Czech Republic

Subscription price (4 issues yearly)
CZK 372 (incl. postage) for the Czech Republic,
EUR 117 or USD 174 for other countries.
Printed copies can be bought at the Publications Shop of the Czech Statistical Office (CZK 66 per copy).
address: Na padesátém 81 | 100 82 Prague 10 | Czech Republic

Subscriptions and orders
MYRIS TRADE, s. r. o.
P. O. BOX 2 | 142 01 Prague 4 | Czech Republic
phone: (+420) 234 035 200,
fax: (+420) 234 035 207
e-mail: myris@myris.cz

Design: Toman Design
Layout: Ondřej Pazdera
Typesetting: Václav Adam
Print: Czech Statistical Office

98th year of the series of professional statistics and economy journals of the State Statistical Service in the Czech Republic: ***Statistika*** (since 1964), ***Statistika a kontrola*** (1962–1963), ***Statistický obzor*** (1931–1961) and ***Česko-slovenský statistický věstník*** (1920–1930).

Published by the Czech Statistical Office
ISSN 1804-8765 (Online)
ISSN 0322-788X (Print)
Reg. MK CR E 4684