STATISTIKA

STATISTICS AND ECONOMY JOURNAL

VOL. **101** (3) 2021

CZECH STATISTICAL OFFICE

CONTENTS

ANALYSES

METHODOLOGY

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

Convergence of Inflation and Unemployment Rates: a Signal of Economic Slowdown?

Stanislava Hronová[1] | *Prague University of Economics and Business, Prague, Czech Republic*
Luboš Marek[2] | *Prague University of Economics and Business, Prague, Czech Republic*
Richard Hindls[3] | *Prague University of Economics and Business, Prague, Czech Republic*

Abstract

In economic theory the Phillips curve presents the relationship between the unemployment rate and inflation rate. The inflation and unemployment rate bring important information about the stages of the economic cycle. This article attempts to find an answer to the question of whether the development of the difference between the unemployment and inflation rate, the so-called signal gap, may be an indicator of changes in the economic cycle. Quarterly data on the Czech Republic, France, Great Britain and the Republic of Korea were used to verify this hypothesis.

Keywords	***JEL code***
Dynamic linear model, unemployment rate, inflation rate, signal gap, GDP	*C22, E37*

INTRODUCTION

The unemployment and inflation rate are among the most important indicators whose values sensitively reflect the changing economic conditions and stages of the economic cycle. The periodicity of determining them (monthly and quarterly, respectively) and the internationally comparable methodology of their estimate place these indicators in the role of certain signalling information on the development of the national economy in the short-term periodicity. The period of a crisis (or recession) is always associated with a high unemployment rate which, with the transition to the stage of recovery, naturally (albeit slowly) falls. The unemployment rate, as a reflection of the growth of consumer prices, tends to be low during a period or crisis, recession respectively, and a return to economic growth is usually accompanied by a price growth expressed by an increasing inflation rate. This empirical evidence results in a hypothesis that during the peak of economic growth the values of these two indicators come closer together,

[1] Department of Economic Statistics, Faculty of Informatics and Statistics, Prague University of Economics and Business, Nám. W. Churchilla 4, 130 67 Prague 3, Czech Republic. Corresponding author: e-mail: hronova@vse.cz. ORCID 0000-0002-3568-9755.

[2] Department of Statistics and Probability, Faculty of Informatics and Statistics, Prague University of Economics and Business, Nám. W. Churchilla 4, 130 67 Prague 3, Czech Republic.

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

i.e. the gap between them narrows. The peak of economic growth is followed by a weakening in economic activity, fall in demand, increase of unemployment and slower price growth. All this leads to the stage of recession which, in terms of the relationship between the unemployment and inflation rate, is displayed as an increase in the difference of their values. If this gap were to be an indicator of stages of the economic cycle, its values could play the role of a certain leading indicator enabling the signalling of changes in the development of the national economy; so we will call it the signal gap.

Some theoretical starting points also point to the fact that the relationship between the unemployment trend and price trend may indicate a change in a stage of the economic cycle. Generally speaking, long-term low unemployment gradually gives rise to a lack of labour force thereby suppressing wage growth. Wage growth is subsequently displayed in an increase of demand, which leads to a price growth that is reflected in the growth of the inflation rate. In view of the slow fall in unemployment, the growth of the inflation rate need not be significant nevertheless sooner or later this situation will force the central bank to a one-off or repeated increase of interest rates. This, in turn, leads to a slow fall in the unemployment rate and to a narrowing in the signal gap.

An increase in interest rates will contribute to a slowdown in economic growth, a fall of share prices and lower availability of foreign sources of financing. The slowdown of economic growth (measured by the GDP growth rate) is accompanied by the level of wages and salaries, whose growth has outpaced labour productivity growth and then led to a further slowdown in economic growth.

The economy is getting into a "vicious circle" resulting in a change of the stage of the economic cycle. Thus the task is to find a period when the falling value of the signal gap begins to indicate a change in the cycle stage, and find a model for such behaviour, including an estimate of what time in advance the determined fall of the signal gap indicates the change of the cycle stage.

The aim of this article is to verify the hypothesis that the value of the signal gap directed at the minimum is an indicator of the peak of economic growth (expressed by the GDP growth rate), which signals the transition to the recession stage. In other words, when the inflation and unemployment rate converge to the same number, a longer economic downturn can apparently often follow. Using methods for a time series analysis, we will then attempt to estimate the size of the time in advance of the minimum of this gap before an approaching recession. To verify this hypothesis, we will use quarterly data for the Czech Republic, France, Great Britain and the Republic of Korea.

1 THEORETICAL STARTING POINTS

The relationship between the unemployment and inflation rate is not a new phenomenon in short-term economic diagnostics. The so-called Phillips curve, as one of the postulates of economic theory, is based on the relationship between these two indicators. Originally A. W. Philips (Philips, 1958) defined the relationship between the trend of wage rates and the unemployment rate; only later did he modify it into the relationship between the inflation and unemployment rate. We know this generalisation as the so-called price-price Phillips curve which represented a significant impetus for later ideas of central banks about the possibilities of the "regulation" of the inflation rate based on the unemployment rate.

The idea of the possibilities of the direction of economic policy, based on Phillips' formulated relationship of two indicators, seemed attractive to the authors of economic policy and representatives of central banks. If it would be possible to influence unemployment by supporting demand, it would be possible to also regulate prices. The period of post-war recovery in Europe played into these ideas when relative price stability did not raise concerns about sharp price growth and high employment was maintained by high economic growth. The change in economic conditions in the 1970s (the oil crisis after the Yom Kippur War in October 1973, changes in the structure of the labour market, growth of labour productivity, expansion of information technologies) led to disillusion with the malfunction of the relationship between the inflation rate and unemployment rate in practical economic policy, particularly in the long-term horizon. One of the most significant economic relationships (although only empirically observed, nevertheless often passed into law), suddenly ceased to apply. The reasons why this was the case was summed up in their paper especially by M. Friedman (Friedman, 1968) and E. Phelps (Phelps, 1967), who showed that unemployment is the outcome of an entire series of economic processes and that they cannot be "regulated" by influencing demand. A significant role in the loss of the importance of the Phillips curve is also played by the policy of central banks based on inflationary expectations. This totally naturally disrupts the relationship between the unemployment and inflation rate, since the expected inflation rate affects economic entities by determining, to a considerable extent, the actual inflation rate. Nevertheless, M. Friedman and E. Phelps admitted that in the short-term horizon the relationship based on the price-price Phillips curve applies, but is accompanied by the growth of inflationary expectations. The validity of the relationship determined by the Phillips curve, or the possibility to regulate the inflation rate based on the influence of demand was later disputed even in the short-term horizon (Atkeson and Ohanian, 2001) or (Lansing, 2002).

Deriving the relationship between the trend of the inflation and unemployment rate was based on (empirically observed) statistical dependence between these variables in that Phillips considered the unemployment rate as the explanatory variable (nevertheless, I. Fisher considered the inflation rate as the explanatory variable, see Fisher, 1973). The discussion concerning the relationship often led to the fact that dependence was interpreted as bilateral and almost no attention was devoted to the economic arguments that opted for the unemployment rate as the explanatory variable.

The long-term observed dependence in one country was therefore incorrectly economically generalised for a different time and different area, and without the mathematical accuracy of the derived relationship being proved, this relationship was used as an instrument of economic policy. It was only a matter of time before such an instrument failed. Statistically proven dependence still does not mean that this is causal dependence. It must also not be forgotten that two variables can only be seemingly statistically dependent, since both are influenced by a third variable (sometimes clearly and sometimes very well concealed) or simply show the identical time trend without any factual context of the limits of the considered variables being justified by this identical time trend.

Another statistical argument to dispute the idea that it is possible to transfer relationships between the unemployment and inflation rate, valid from the mid 19th century to the mid 20th century and even passed into law is the very content of these indicators. The definition and methods of determining these indicators have changed significantly since the time Phillips' paper was published. And this means their values, interpretation and ability to reflect reality as was the case 100 years ago have also changed.

Despite the doubts about the universal validity of the above described economic relationships, it is clear that the unemployment trend and price trend very sensitively react to changes in the economic cycle. So it is necessary to look at the new possibilities that the relationship of the values of these two indictors can bring us.

If it can be assumed that the peak stage of the economic cycle is accompanied by the increased demand for labour force reflected in the low unemployment rate and that the surplus of free funds and high demand causes a price rise (inflation rate), then the signal of an approaching point of "overheating" is the convergence of the values of these two indicators. In other words, the difference of the values of the unemployment and inflation rate which we call the signal gap decreases in the direction of the peak of the stage of economic growth, and is therefore not just the indicator of this state, but also the indicator of the transition to the stage of recession.

If we accept this assumption, then what should apply is that there is a connection between economic development (measured GDP growth rate) and the signal gap, or what should apply is that the signal gap fulfils the role of a leading indicator of change in the stages of the economic cycle. It is clear that a certain time lag must be considered which must be found using methods of time series analysis.

2 DATA AND ANALYSIS METHODOLOGY

To verify the validity of the hypothesis of the role of the signal gap, the selected indicators need to be defined first in terms of their content and periodicity, so that they best correspond to the theoretical assumptions and their values best reflect the development of economic boom during the analysed period. It must also be decided how long such a period should be so that the formal assumptions are met of the use of time series analysis methods, but also factual requirements, i.e. requirements to be able to affect stages of the economic cycle.[4] Last but not least, by using appropriate time series analysis methods to create a model description of the relationship of the GDP growth rate and signal gap, and find the signal gap lead, or lead of its changes before a change in economic development. The definition of indicators and subsequent analysis will rely on data of the Czech Republic, France, Great Britain and the Republic of Korea which are available on the websites of the relevant statistical offices.[5] The choice of these countries is determined to a certain extent by the unavailability of short-term data of the relevant periodicity and length of the time series on the websites of the absolute majority of statistical offices. There is no central source of data from which to obtain data in the required form for the same time period for a random country. So it was necessary to look through the websites of national statistical offices and find the relevant data.

The initial indicators for calculating signal gaps are the inflation and unemployment rate, which are methodically internationally comparable indicators. The values of these indicators are normally published in monthly, quarterly and annual periodicity. However, information applying to the development of the national economy (GDP growth rate) is only available in quarterly and annual periodicity. If we are to follow an economic boom in a short-term horizon, it is clear that we can only work with quarterly data.

The inflation rate is a relative increment corresponding to the consumer price index. The quarterly inflation rate is expressed by a percentage change of the price level in the given quarter in comparison with the immediately previous quarter; values are seasonally adjusted. This is formally the share of the base year of the consumer price index in the given quarter and the base year of the consumer price index in the previous quarter whereby the base is the same in both cases (in all stated countries the base is the average of 2015).

The fact that the inflation rate is estimated from the development of consumer prices is beneficial in terms of its international comparability. On the other hand, (given the definition of the household final consumption expenditure indicator), the inflation rate does not include the growth of real estate and construction work prices. This seems insufficient from the point of view of capturing the overall rise in prices. The solution in a given situation could be to use the GDP deflator, which should take into account the overall price movement. Here, however, we would encounter the problem of the interdependence of the GDP growth rate and the GDP deflator, and the definition of a signal gap would then be meaningless.

The unemployment rate is defined as the share of the number of unemployed in the number of economically active aged 15–64 years. In all the analysed countries the quarterly general unemployment rate was used, which is based on the definition of unemployed persons according to the conditions of the International Labour Organisation – ILO; the values are seasonally adjusted.

The signal gap is then defined as the difference between the (quarterly) unemployment rate (in %) and the (quarterly) inflation rate (in %). Its value is therefore expressed in percentage points. Values of this difference approaching zero signal the peak of the economic cycle phase, and are therefore an indicator of an early transition to a deceleration phase.

---

[4] Given the requirement for a reasonable length of the time series from a formal and factual point of view, authors are naturally limited by the availability of data on the websites of statistical offices.

[5] See <*www.czso.cz*>; <*www.insee.fr*>; <*www.ons.gov.uk*>; <*http://kostat.go.kr*>.

To describe an economic boom the quarterly GDP growth rate indicator was used as it is the relative increment corresponding to the index comparing the value of GDP in the given quarter and in the previous quarter; the values are seasonally adjusted. The corresponding GDP values are expressed in the chained prices of the previous year (annual chain linked quarterly data).

When choosing the length of quarterly time series, we again referred to the availability of data on the websites of selected national statistical offices. The limiting factor was always the limited length of the time series of the unemployment and inflation rate. Nevertheless, even in these time limits, we had at least 80 observations from the selected countries. We consider the given length of the time series to be satisfactory not just for using tools of stochastic modelling, but also to describe the stages of the economic cycle.

These stages with varying intensity, in individual countries, also reflect the stages of the development of the world economy (slump primarily of Asian economies in the second half of the 1990s, the post-transformation crisis of the countries of Central and Eastern Europe, the rapid growth of the prices of American technology stock at the start of the millennium as the first signal of the global economic recession then approaching in 2008, the fiscal irresponsibility of developed European countries in years 2001–13 etc.

If we base our hypothesis on the assumption that the signal gap fulfils the role of lead indicator in relation to the GDP growth rate, it is necessary to show the dependence firstly of these time series. Therefore, the first step was the graph showing the progress of these series in which the series trend and their correlation can be seen well. These graphs indicated how strong the dependence will be between the series of the GDP growth rate (in graphs marked GDP) and the signal gap (in graphs marked GAP) for each of the analysed countries.

To prove dependence, including the time delay between both series, we used the cross-correlation function (CCF). This function is normally used as the linear dependence rate of two time series. Its advantage is the fact that apart from the dependence level, it also determines the direction of any linear dependence, i.e. also any time lag. The outcome is a table with values and a correlogram which shows the calculated correlation coefficients not just in the same time point t for both series, but also correlation coefficients in time t for one series and in time $t \pm 1$, $t \pm 2$ for the second series, etc. The correlogram also shows a 95% confidence interval for the values of the correlation coefficient, which considerably simplifies the identification of the statistically important values of this coefficient, including any time lag between the series. The CCF is defined[6] as:

$$\rho_{XY}(k) = \frac{\gamma_{XY}(k)}{\sigma_X \sigma_Y}, \tag{1}$$

where X_t and Y_t are the analysed time series. The CCF is then defined at k as the covariance of X_t and $Y_t + k$ for $k = 0, \pm 1, \pm 2, \ldots$, divided by a product of the standard deviations of both series, where σ_X and σ_Y are the standard deviation values for the series X_t and Y_t (respectively). It is clear that for the CCF the relationship is:

$$\rho_{XY}(k) = \rho_{XY}(-k). \tag{2}$$

Models appropriate for this situation are derived from the class of dynamic linear regression models. These models are based on ARIMA processes. It is assumed that the output series Y_t depends on its past time values $t-1, t-2, \ldots$, then on the values of the input series X_t in time points $t, t-1, t-2\ldots$

[6] The general definition of the CCF and its properties can be found for example in the paper of Wei (2006) or in Box, Jenkins, Reinsel (1994).

and on the values of the so-called noise time series Nt, which is self-regulated by the ARIMA process. The entire theory of linear dynamic models is described in detail[7] and very good results are achieved with it. In general, values of output series can be recorded using the following model,

$$Y_t = c + \nu_0 X_t + \nu_1 X_{t-1} + \nu_2 X_{t-2} + \dots + \nu_K X_{t-K} + \frac{1}{(1-\phi_1(B))(1-\Phi_1(B^T))}\,\varepsilon_t\,, \tag{3}$$

where Y_t is the output series, X_t is the input series, c is constant, ν_i are unknown parameters for $i = 0, \dots, K$, $\phi_1(B)$ is the autoregressive operator of order 1, $\Phi_1(B)$ is the seasonal autoregressive operator of order 1, ε_t is the random variable (white noise), B is the shift operator ($BY_t = Y_{t-1}$), L is the length of season (cf., e.g. Box, Jenkins, Reinsel, 1994).

A whole series of criteria exist for this type of model according to which the model can be verified. Most of them are based on the autocorrelation and partial autocorrelation function. Here we also used other tools such as the unit root test, homoscedasticity test, Dickey-Fuller tests and others.[8]

3 RESULTS OF THE ANALYSIS

We analysed the data from the four selected countries. These are the Czech Republic, France, Great Britain and the Republic of Korea. Generally speaking, these are countries marked for their developed economy and monitor and report indicators over the long-term which we use in our analyses.

Let us look at how the economy developed in the analysed countries in the last ca 30 years, how the time series of quarterly GDP growth rates and signal gaps (GAP), models and total results appeared for the individual countries. We will present the detailed results and analysis procedure, including the CCF, only for the data for the Czech Republic, to show how we proceeded. For the other countries we will proceed analogically and will present only a graph showing the progress of the series and resulting model.

3.1 Czech Republic

We had quarterly data available for the Czech Republic for the period of 1996–2019, so we worked with 95 observations.[9]

The period after 1990 was marked in the Czech Republic by a decline in economic activity in connection with the transition from a centrally planned to a market economy. The short period (up to 1993) was followed by economic growth which peaked in 1995 and 1996 (the value of the signal map close to zero is a good indicator here of the "overheating" of the economy and approaching economic crisis – see Figure 1) and subsequent recession in years 1997 and 1998.[10] The reasons for this economic crisis were internal and involved unresolved privatisation problems (including privatisation of the banks), the slow restructuring of industry, currency and credit crisis of the banking sector and a tough restrictive anti-inflationary policy. The year 2000 saw a turning point in economic development leading to growth and the start of a recovery resulting in the most successful years of economic development in the Czech Republic (which is illustrated by the signal gap's high values – see Figure 1).

---

7 For more see Wei (2006) or Box, Jenkins, Reinsel (1994).

8 The entire analysis was carried out in Stagraphics Centurion, version 16. The model estimate may be carried out differently so the results in various softwares may differ. Obviously, this does not in any way matter provided that the theory is respected according to which the given procedures were programmed.

9 Information for the 1st quarter of 1996 is not available.

10 The low signal gap values (0.0 pp in the 3rd quarter of 1997, 0.9 pp resp. in the 2nd quarter of 1998) is the outcome here of state intervention in the economy in order to avert the growing state budget deficit and stop the downturn in the economy. This led to a sharp price growth in the presented quarters. Thus, in this case the low and isolated signal gap values cannot be considered a tendency that could be an indicator of an approaching recession, but a one-off reaction to state intervention.

However the stage of recovery, which is considered the period of 2001–04, and the following economic boom stage of 2005–07 had different features. The period of 2001–04 was marked by stable economic growth, supported by high rates of growth of industrial and building production, growth of consumption of households and general government, as well as the gradual improvement of foreign trade links, including terms of trade with significant strengthening of the Czech koruna and relatively high, but a stable unemployment rate, lower inflation rate and fall of the prices of industrial manufacturers. In the period of 2005–06 the basic growth factors change: there is a rise in the importance of foreign exchange, Czech currency strengthens significantly, the general government debt level stabilises, the general government deficit decreases and unemployment falls slightly. However, the positive results of the national economy of the Czech Republic ended with the onset firstly of the global financial crisis in years 2008–09 and the subsequent recession in 2010–12 (the signal gap values in 2007 gradually fell significantly to 0.9 p. p. in the first quarter of 2008, and indicated the start of a recession). The high values of the general government deficit with a downturn in economic activity led to a growth of the general government debt to values around 45% of GDP (i.e. relatively high above the long-term approximately 30%) and growth of unemployment.

The year 2014 saw a recovery supported by growth of industrial and building production, retail, growth of business and state investment (without a significant increase of its debt) and displayed by a fall in unemployment, real wage growth, surplus of the current account on the balance of payments. All this came at a low inflation rate, low unemployment rate and decreasing general government debt. From the third quarter of 2017 the Czech Republic shows a slowdown of growth (the average inter-quarter rate in this period is 0.6%) and the signal gap shows a decreasing tendency again during five quarters. The Czech economy comes up against barriers of further growth (the unemployment rate is below 3% and the inflation rate below 1%), and a slowdown in economic growth can be expected which actually came in 2019. However the year 2020 brought unpredictable problems with the COVID-19 pandemic, which will indisputably lead (not only in the Czech Republic, but also in the other analysed countries) to a significant fall of GDP and general government debt growth.

To illustrate the relationship of the trend of GDP and signal gaps, we will show the progress of both time series (see Figure 1). At first glance it is clear from the graph that there is dependence between the series. It appears that the GAP series shows very similar progress to that of the GDP series, nevertheless with a certain time delay. This fact should be confirmed by the CCF values (see Figure 1). The progress of the CCF should indicate a lot in terms of the shape of the model, it is clear that there exists a significant correlation between the studied GDP and GAP series in the time point t (GDP) and $t - 4$ (GAP). Thus, there is a time lag by four time units (a delay by four quarters, that is by an entire calendar year). We will use this fact when creating the model. This significant linear dependence with a time delay by four time units is not only apparent in the graph, but is also confirmed by the correlation coefficient zero value test. The correlation coefficient for the combination of time points t and $t - 4$ resulted as being significantly different from zero (as the only one that is not found in the 95% confidence interval).

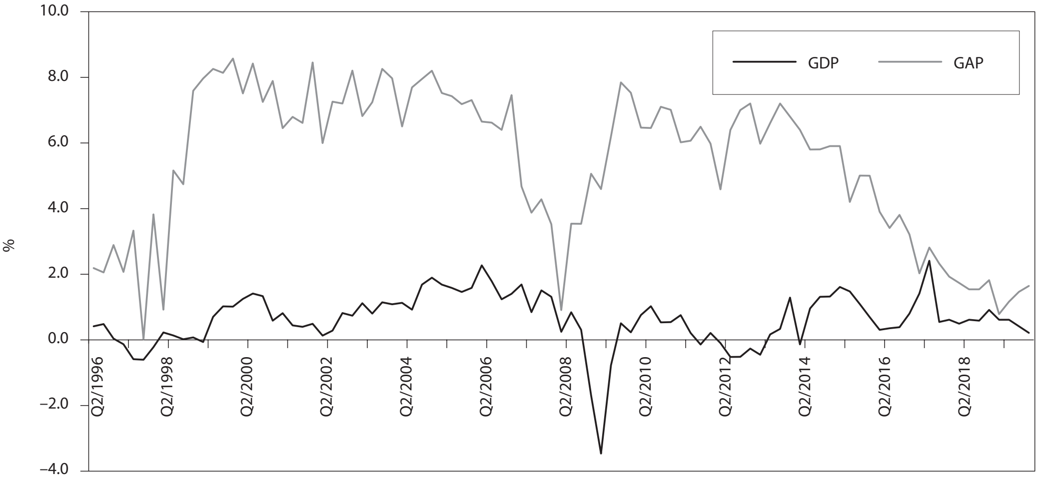
After we had analysed the ARIMA model for the GDP series, we proceeded to construct a transfer function model of the given Formula (3). Finally we reached the ARIMA model (0, 1, 1) with one regressor in the form of

$$Y_t = 0.173X_{t-4} + (1-0.283B)\varepsilon_t \,, \tag{4}$$

where Y_t is the series of the GDP quarterly growth rate after the current differentiating, X_t the series of the signal gap quarterly values, and ε_t is the standard white noise. It must be pointed out that the model went through a whole series of tests (tests of residue, the unit root, homoscedasticity, and Dickey-

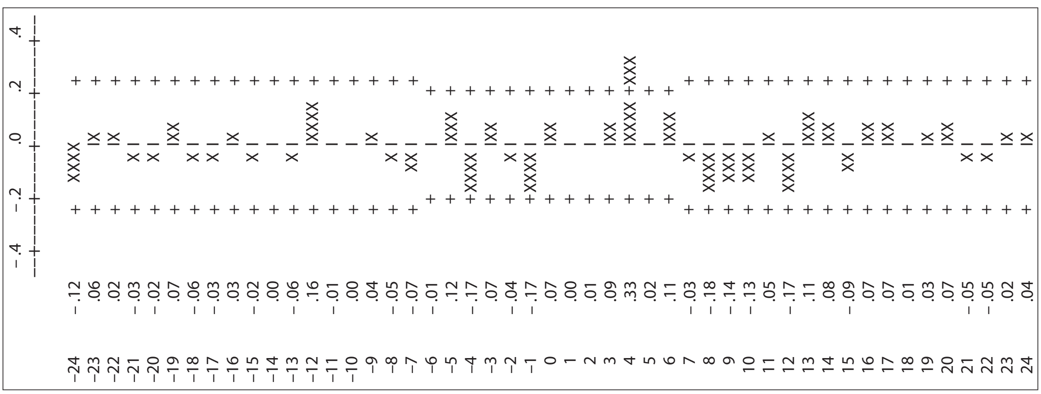
Fuller tests) and was shown to be fully adequate. The output from Stagraphics Centurion is presented in the following table. As model quality criteria, the usual rates were used which we also present and which support the model's adequacy.

Figure 1 Trend of the GDP growth rate and signal gaps – Czech Republic 1996–2019



Source: <*www.czso.cz*>

Figure 2 Graph showing the progress of CCF (Czech Republic) – 95 percent confidence interval



Source: Own calculations, <*www.czso.cz*>

This analysis demonstrates in the examples of the Czech Republic that the hypothesis concerning the prediction potential of the signal gap is valid. The model also shows with what delay after the signal gap values close to zero, a recession is approaching. In addition, the data of the Czech Republic show that the zero (or approaching zero) time isolated signal gap values may be the outcome of one-off state (e.g. fiscal, tax) measures influencing the price jump in the given quarter. Such one-off values cannot (given that they do not represent the tendency being asserted in several quarters) be considered an indicator of an approaching recession.

Table 1 Stagraphics Centurion output – Czech Republic

ARIMA Model Summary

Parameter	*Estimate*	*Stnd. Error*	*t*	*P-value*
MA(1)	0.283404	0.103777	2.7309	0.007571
LAG(Gap;4)	0.173216	0.0534385	3.2414	0.001657

Backforecasting: yes
Estimated white noise variance = 0.353578 with 92 degrees of freedom
Estimated white noise standard deviation = 0.594624
Number of iterations: 5

Forecast Summary

Parameter	*Estimation*
Statistic	*Period*
RMSE	0.594426
MAE	0.419092
ME	–0.00307198

Source: Own calculations, <*www.czso.cz*>

3.2 France

For France we had complete quarterly data available for the period of 1990–2019, which consist of 120 observations.

France is the sixth biggest world economy[11] and the second biggest in the Eurozone[12] (measured by nominal GDP). Its economic development after 1990 is characterised by low, but relatively stable GDP growth rates (the average annual GDP growth rate is 1.6%), low inflation rate (the average annual information rate 1.5%), high unemployment rate (with the lowest value of 7.4% in 2008 and the highest of 10.7% in 1997) and rising general government debt (from the value of 35.6% of GDP in 1990 to 98.1% of GDP in 2019).[13]

In terms of quarterly GDP growth rates, the critical periods were between 1992 and 1993, the second quarter of 2003 and the period from the second quarter of 2008 to the second quarter of 2009 when there was an inter-quarter fall in GDP. This was preceded by decreasing signal gap values during 1991, during 2002 respectively, during 2007 respectively (see Figure 3).

The reason for the economic decline in the early 1990s was the postponement of major reform steps which, in an aging population, disrupted the economic and financial balance. Subsequently, the reform of pension and health insurance was adopted and measures were taken to boost the economy and reduce unemployment.

The loss of competitiveness on foreign markets had a negative impact on France's economic growth in the first years of the 21st century. Therefore, the French government adopted a stabilisation programme for years 2004–06, which led to the greater success of French exporters, primarily to Asian markets and a return to the stage of economic growth. The economic and fiscal crisis of 2008–09 had a heavy impact on the French economy, GDP fell continuously for five quarters, the general government deficit rose to 7.2% of GDP, the general government debt rose year-on-year by 14 p. p. in relation to GDP, falling demand suppressed price growth, which resulted in deflation (between 2008 and 2009).[14]

---

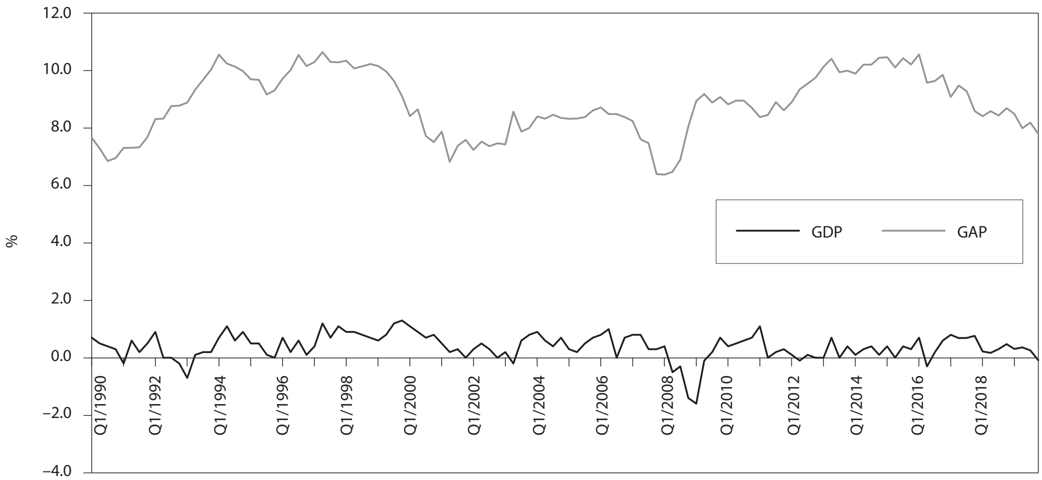
[11] Data for 2018, viz <*https://databank.worldbank.org/data/download/GDP.pdf*>.
[12] <*www.eurostat.eu*>.
[13] For these and the following data see <*www.insee.fr*>.
[14] See INSEE (2014).

In 2012 and 2013 the French economy was marked by stagnation (0.3%, 0.6 % year-on-year GDP growth respectively). Subsequent recovery in France came slowly, more than 2% of the year-on-year GDP was not achieved until 2017; however 2018 was again marked by a slowdown to 1.7% of year-on-year growth and 2019 to 1.0%, which was preceded by a decrease in the value of the signal gap during 2017 and 2018 (see Figure 3).

In the long-term high unemployment rate and relatively low inflation rate the signal gap values are significantly higher than in the Czech Republic. Nevertheless, their fall in several quarters is a signal of an approaching slowdown of economic growth, recession respectively.

The CCF values even in the case of France showed a linear dependence with a time lag of four time units, which is incidentally also apparent from the progress of analysed series illustrated in the graph.

Figure 3 Trend of the GDP growth rate and signal gaps – France 1990–2019



Source: <*www.insee.fr*>

An appropriate model was again showed to be the ARIMA model (0, 1, 1) with one regressor, this time in the form of

$$Y_t = 0.256X_{t-4} + (1-0.592B)\varepsilon_t\ , \tag{5}$$

where individual symbols have the same significance as in the previous model. The model again passed all the stages of verification successfully and the characteristics of the model's quality also came out positively.

The analysis of the data of France shows that the hypothesis concerning the prediction potential of the signal gap is valid and a recession is approaching with a delay with decreasing signal gap values (values approaching 6 p. p.). So unlike the Czech Republic, these are not signal gap values approaching zero, in view of the long-term high unemployment rate (often exceeding 10%) and very low inflation rate. However, here too the decrease is apparent of the signal gap values before the approaching economic slowdown.

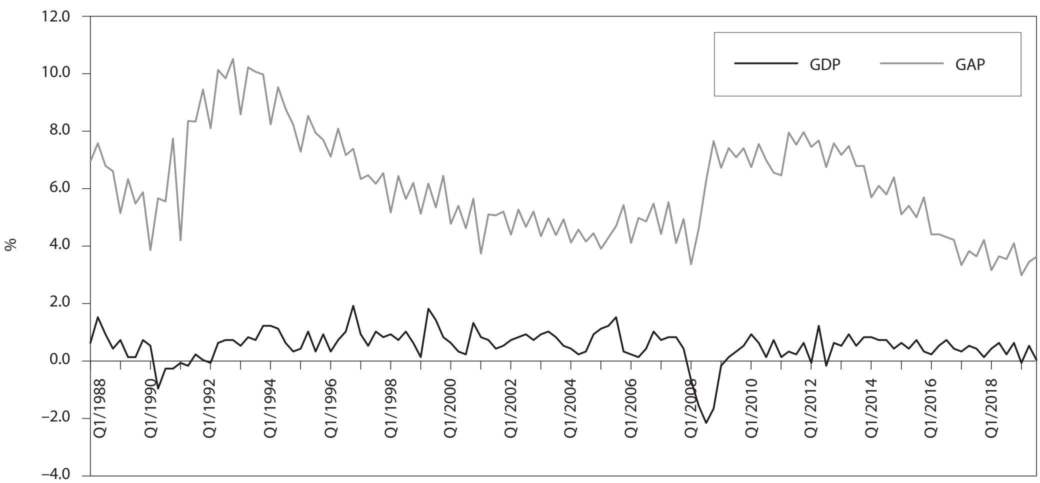
3.3 Great Britain

In the case of Great Britain, we worked with 128 observations, which are quarterly data for the period of 1988–2019; this was the longest time series in our analyses.

The economy of Great Britain is the fifth biggest world economy[15] and is the biggest financial centre in the world. Financial services are the most important export commodity and services account for almost 80% of GDP.

The economic development of Great Britain after 1990 is characterised gradually by the decreasing GDP growth rate (from growth rates exceeding 3% in the second half of the 1990s to a growth rate not exceeding 2% in years 2016–19; the average annual growth rate of GDP for the entire analysed period is 2.1%), a relatively unstable inflation rate (average annual inflation rate 2.6%), sooner a higher unemployment rate (with the lowest value of 3.9% in 2019 and the highest exceeding 10% after the crisis in 1991) and rising general government debt (from a value of 34.2% of GDP in 2002 to 86.0% of GDP in 2019).[16]

Figure 4 Trend of the GDP growth rate and signal gaps – Great Britain 1988–2019



Source: <*www.ons.gov.uk*>

In terms of quarterly GDP growth rates, the critical stages were the period from mid 1990 to mid 1992 (negative inter-quarter growth rate), the period from the second quarter of 2008 to the second quarter of 2009, the year 2012 with a fall in GDP in the second and fourth quarter and the year 2019 with a fall in GDP in the second quarter. This was preceded by decreasing signal gap values between 1989 and 1990, during 2007 respectively, during 2018 respectively (see Figure 4).

After economically successful years (from 1982, with peak growth in 1988) the start of the 1990s were marked with a transition from crisis. Some economic decisions of the final years of Margaret Thatcher's government led the economy into a vicious circle, which made it even difficult for big businesses and higher income sections of the population to get out of. A number of successful companies up to this time also went bankrupt, lower income groups, which were forgotten by the relative economic boom of the first decades of Thatcher, fell into poverty. The consequence of the crisis led to a rise in unemployment (from 7.0% in the first six months of 1990 to more than 10.0% from the second half of 1992 at a very low inflation rate).

The years 1993–2007 are a period of economic growth at a gradually falling unemployment rate and continuous low inflation rate. So there is a gradual narrowing of the signal gap, which ushers

[15] For data for 2018, see <*https://databank.worldbank.org/data/download/GDP.pdf*>.
[16] For these and the following data see <*www.ons.gov.uk*>.

in the approaching crisis. The threat of the financial and mortgage crisis imported from the US was averted up to a certain extent by the preventive measures of the British government. This also put a stop to the wave of distrust in the British banking sector whose shocks would have seriously damaged the entire economy. Nevertheless, GDP fell by 4.2% in 2009 and recovery came slowly. State intervention helped to reduce the period of crisis, but to a certain extent complicated the transition to the stage of recovery – the British economy again showed a fall in GDP in the two quarters of 2012.[17]

Warning signals (here in the form of the narrowing signal gap – see Figure 4) about an approaching recession in 2018 were displayed with a fall in GDP in the second quarter of 2019 (by 0.1% as opposed to the previous three months). This development, among other, was reflected in stockpiling, with the March deadline for Great Britain's exit from the European Union and the anticipation of problems in preparation for Brexit at the end of October 2019. Incidentally, the British economy had already slowed down from mid 2016, when the Brits decided in a referendum that the country would leave the European Union (on 23 June 2016).

The progress of the series (and CCF values) also, in the case of Great Britain, confirmed the dependence between the series with a time delay. We identified an appropriate model in the form of:

$$Y_t = 0.106X_{t-4} + (1-0.361B)\varepsilon_t \, . \qquad (6)$$

This again is the ARIMA model (0, 1, 1) with one regressor. The model passed all stages of testing confirming its adequacy.

The analysis of the data of Great Britain also demonstrates that the hypothesis concerning the prediction potential of the signal gap is valid and a delayed recession is coming with decreasing signal gap values (values approaching 3 p. p.). Given a higher unemployment rate and a sooner higher inflation rate, these are not signal gap values approaching zero (similar as in the case of France).

3.4 Republic of Korea

We obtained data for the national economy of the Republic of Korea for the period of 1999–2019; we have data for 1999 from the third quarter, therefore we are working with 82 observations. The reason for having a shorter time series than in the previous cases is because of the unavailability of quarterly data on unemployment before 1999.

The economy of the Republic of Korea (hereinafter "Korea") is the twelfth biggest world economy[18] and is a developed country focused primarily on the car and electrical engineering industry. Korea's economy is marked by a specific business environment (the position of big companies, so-called chaebõls,[19] which have a large microeconomic impact and are significantly interlinked with political power). After learning from the crisis in the late 1990s, the Korean government maintains a low level of public debt (36% of GDP in recent years) and also a reserve fund which it uses in case of a crisis. These measures also had an effect on the small impact of the global crisis in years 2008–2009, when the Korean economy, as only one of a few, grew during these years of crisis (year-on-year growth of GDP 3.0% in 2008 and 0.8 % in 2009).[20]

[17] There was a similar difficult post-crisis trend in the Czech Republic for example. The reason was the application of inappropriately selected instruments of state economic policy, nevertheless different than in Great Britain.

[18] For data for 2018, see <*https://databank.worldbank.org/data/download/GDP.pdf*>.

[19] Chaebõl is a specific type of group operating in various sectors and based on the interconnection of business and financial controls; it is usually owned by a single businessman and members of his family – e.g. Samsung, Hyundai, Daewoo, Samyang; see Chang and Lee (2006).

[20] For these and the following data see <*https://kosis.kr/eng*>.

At the end of 1997 Korea (just like other Asian countries) was affected by the financial crisis (in 1998 GDP fell by 5.1%). The economy suffered a high foreign debt, lacked foreign exchange reserves and the government had to ask the International Monetary Fund for an emergency loan. The crisis resulted in the bankruptcy of a number of companies (including some chaebõls such as Daewoo), as the banks refused to finance them. The government therefore carried out a number of reforms of the financial and labour market, and the public sector.[21] The measures were effective and by 1999 Korea recorded an 11.5% growth of GDP.

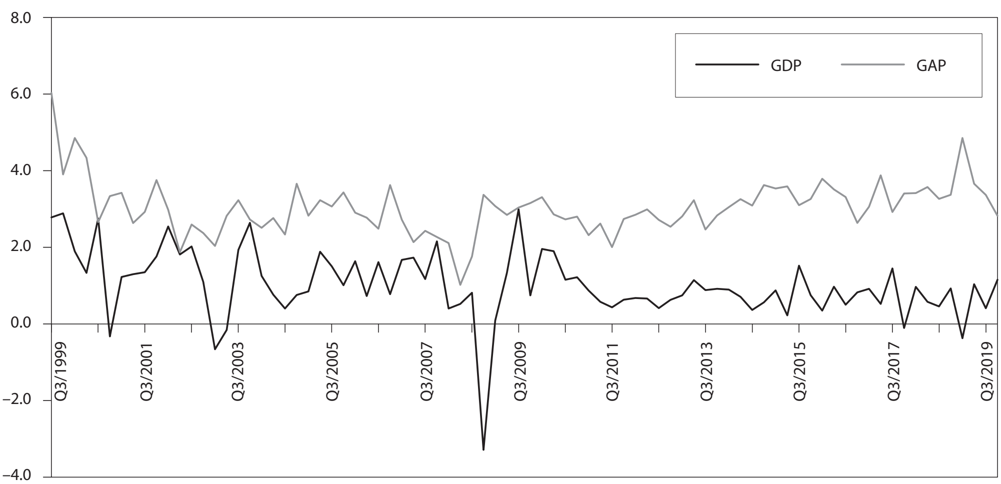
The Korean economy has been recording a high and relatively stable GDP growth rate (an annual average of 4.4%); the price trend has also been similarly stable (the average annual inflation rate is 2.3%) as has unemployment (the annual unemployment rate ranges between 3.1–4.4%).

If we look at the quarterly data in the period of 1999–2019, it is apparent (see Figure 5) that the low signal gap value at the end of 1999 ushered in a fall of GDP in the last quarter of 2000 (by 0.3%); the same was the case in 2002, when the low signal gap value warned of a possible economic slump which came in the first and second quarter of 2003 as a reaction to the recession in the US. This period is characterised in Korea, among other, by the rapid growth of household consumption and associated debt levels.

The period of 2004–07 marks a stage of prosperity in Korea with a GDP growth rate exceeding 4% per annum. Recovery came as a consequence of the growing success of exports (particularly electrical engineering) and was also supported by the economic growth of China as the traditional economic partner. The years of the economic boom is seen in the extent of the signal gap (see Figure 5).

The year 2008 brought a fresh warning for the Korean economy that without reforms to the business environment, a prosperous economy cannot be sustained. Therefore the government proceeded with fiscal and tax changes, support for small and medium enterprises, for low-income households and for building the infrastructure. At the end of 2009 the crisis had already been overcome.[22] Decreasing signal gap values in 1997 (see Figure 5) again provide a very good indication of an approaching crisis.

Figure 5 Trend of the GDP growth rate and signal gaps – Korea 1999–2019



Source: <*http://kostat.go.kr*>

[21] See Khayyat (2015), Chang and Lee (2006).
[22] See OECD (2018).

A fall in quarterly GDP reappeared in the first quarter of 2019 as a consequence of lower export performance, which was reflected in a slowdown of the global economy,[23] growing protectionist tendencies and falling export prices. In addition, the low GDP growth rate was accompanied by a fall in prices (deflation in the first and third quarter of 2019).

A significant correlation with a time delay of four units was again confirmed (a significant CCF value) even in the case of Korea. The trend of the series in a given period is shown in Figure 5.

After a thorough analysis also in the case of Korea, we opted form the ARIMA model (0, 1, 1) with one regressor in the form of:

$$Y_t = 0.221X_{t-4} + (1-0.948B)\varepsilon_t \,. \tag{7}$$

It is interesting that the same model was used for all the countries, albeit with different parameter values. Of course, this does not have to be the general rule.

The analysis of the data for the Korean national economy also demonstrated that the hypothesis concerning the prediction potential of the signal gap is valid and a delayed recession is coming with decreasing signal gap values (values approaching 2 p. p.).

CONCLUSIONS

Economic development in each country is the outcome of the effect of a series of economic factors. Among the most significant short-term indicators informing about the progress of the stages of the economic cycle of the national economy is indisputably the unemployment and inflation rate. The difference in the values of these indicators, here called the signal gap, has a tendency to fall (a tendency to approach zero respectively) in the stage of the economic cycle peak, when the economy has exhausted its further growth resources and is followed by a slowdown in economic development, i.e. a transition to the stage of recession (or crisis). The article on the quarterly data of the Czech Republic, France, Great Britain and the Republic of Korea for the period from the 1990s to 2019, i.e. for the period when each of the analysed countries went through, among other, a global financial and economic crisis in years 2008–09 and a subsequent recession, demonstrates that such a hypothesis applies. The model then makes it possible to determine with what delay recession arrives with decreasing signal gap values (values approaching zero respectively).

The analyses based on the CCF and regressive dynamic models showed in all the analysed countries that recession comes in four quarters after the lowest signal gap value. Therefore a significant linear dependence in time points t a $t - 4$ exists between the GDP and GAP time series. This means that the GDP time series "precedes" the GAP series by 4 time units. We successful described this dependence using an appropriate linear dynamic model, specifically the so-called transfer function model. We analysed the data of the four selected countries and in all cases the most appropriate turned out to be the ARIMA model (0, 1, 1) with one regressor. This regressor was the GAP series with a time delay of four time units (four quarters). This means that if we know the GAP series value in time point t, we can use it in the stated models and obtain a very good estimate of the GDP quarterly values for the next year. If we carry out this procedure for each of the four quarters of the last known year, we obtain a solid prediction of the possible trend (including any qualitative changes) for the next year.

The analysis of the inflation and unemployment rates data used and the verification of the hypothesis of their convergence as a signal of change in the stage of the economic cycle therefore show a certain potential of how to use the signal gap as an indicator of the turning point in economic development. This at least in a situation of gradual cyclical economic development – both used rates and their

[23] See, among other, the situation in other countries analysed here.

convergence are the outcome of the changing economic potential in a standard economic cycle. The results of the analysis can therefore be used in short-term forecasts of economic development and in analyses of the economic cycle. It turns out that the indicators of the inflation rate and the unemployment rate must continue to be considered not only as separate and important indicators of the state of the economy, but it is also necessary to evaluate the development of the difference in their values over time. We thus obtain considerable information about the phase of the economic cycle.

Nevertheless, the article was written during the period of the aggressive onset of the coronavirus crisis which is currently a dominant process of intervention of the highest intensity and its impact on the economy cannot, at present, be accurately described in a model. At this time it is not even known, from a medical and economic point of view, what the further development and intensity of the impact of this viral pathogen will be. Neither are the outlines of the individual stages of the cycle known and how they will build up after the coronavirus pandemic. Despite this – or perhaps because of this – it would be expedient, once the crisis subsides, to also apply the above analysis to the new data from the post-pandemic period. It is evident that to analyse indicators entering the signal gap and later analyse the gap alone may turn out to be a useful tool for detecting changes in the economic cycle even after the stage of the unusually strong intervention of a pandemic.

References

ANDRE, C., GARCIA, C., GIUPPONI, G., PARELIUSSEN, J. Labour Market, Welfare Reform and Inequality in the United Kingdom. *OECD Economics Department Working Papers*, Paris: OECD Publishing, 2013, No. 1034.

ATKESON, A. AND OHANIAN, L. E. Are Phillips Curves Useful for Forecasting Inflation? *FR Minneapolis Quarterly Review*, 2001 (Winter), pp. 2–11.

BOX, G. E. P., JENKINS, G. M., REINSEL, G. C. *Time series analysis, forecasting and control.* 3rd Ed. New Jersey: Prentice-Hall, Inc. Englewood Cliffs,1994.

CHANG, Z.-S. AND LEE, S. H. *Transformation in Twentieth Century Korea.* 1st Ed. New York: Routledge, 2006.

FERREIRA LOPES, C. E. AND CROSS, N. J. G. New Insights into Time Series Analyses. *A&A*, 2017, 604, A121.

FOX, J. *Applied regression analysis and generalized linear models.* London: Sage Publications, 2016.

FISHER, I. A Statistical Relation Between Unemployment and Price Changes, International Labor Review, 1926. Reprinted as: I Discovered the Phillips Curve, *Journal of Political Economy*, 1973, 81, pp. 496–502.

FRIEDMAN, M. The Role of Monetary Policy. *American Economic Review*, 1968, 58, pp. 1–17.

HINDLS, R. AND HRONOVÁ, S. Unemployment and inflation rates in the years of economic turbulences or do the Phillips curve relationships still hold? In: *8th Annual Conference of the EuroMed Academy of Business*, EuroMed Press, Verona, 2015, pp. 861–870.

HRONOVÁ, S., HINDLS, R., MAREK, L. Does the Signal Gap Indicate a Growing Risk of Economic Recession? In: *12th Annual Conference of the EuroMed Academy of Business*, EuroMed Press, Thessaloniki, 2019, pp. 388–399.

INSEE. *Trente ans de la vie économique et sociale.* Edition 2014, Paris.

KHAYYAT, N. T. History of Economic Development in South Korea. In: *Energy Demand in Industry*, Springer Dordrecht, 2015, pp. 21–29.

OCDE. *Etudes Economique de l´OCDE: France 2019.* Paris: Editions OCDE, 2019.

OECD. *OECD Economic Surveys: Korea 2018.* Paris: OECD Publishing, 2018.

OECD. *OECD Economic Surveys: United Kingdom 2017.* Paris: OECD Publishing, 2017.

LANSING, K. J. Can the Phillips Curve Help Forecast Inflation? *FRBSF Economic Letter*, 2002 (October), 4, No 2002-29.

PHELPS, E. S. Phillips Curves, Expectations of Inflation, and Optimal Unemployment Over Time. *Economica*, 1967, 34, pp. 254–281.

PHILLIPS, A. W. The Relation between Unemployment and the Rate of Change of Money Wage Rates in the United Kingdom, 1861–1957. *Economica*, 1958, 25, pp. 283–299.

SAMUELSON, P. A. AND NORDHAUS, W. D. *Economics*. 19th Ed. Boston: McGraw-Hill/Irwin, 2010.

WEI, W. W. S. *Time series analysis – Univariate and multivariate methods.* New York: Pearson, Addison Wesley Publishing, 2006.

WOODWARD, W. A., GRAY, H. L., ELLIOTT, A. C. *Applied Time Series Analysis.* CRC Press, 2012.

Inflation Forecasting and Targeting: Experience from Central Europe

Jiří Šindelář[1] | *University of Finance and Administration, Prague, Czech Republic*
Petr Budinský[2] | *University of Finance and Administration, Prague, Czech Republic*

Abstract

This paper deals with inflation forecasting and targeting performance of selected Central and Eastern-European central banks. Using battery of absolute and scaled forecasting errors along with significance tests, we have evaluated inflation predictions on the optimal monetary policy transmission horizon (14–16 months), as well as adherence to long-term inflation targets. Out of the evaluated Czech, Hungarian and Polish central banks, complemented by the European Central Bank for comparison, it was found, that even though the bank´s performance improved during the last decade, notably with the forecasting component, some issues are still present. These are mostly connected to the inflation targeting mechanism, which was found to contain systemic bias in the case of the Czech national bank, as well as failing in comparison with the naïve benchmark in the case of the European Central Bank. Both outcomes pave the way for further investigation in a wider economical context.

Keywords	**JEL code**
Central bank, inflation forecasting, inflation targeting, forecasting error	*E58, E37, E31*

INTRODUCTION

Inflation targeting (IT) has become the central method of monetary policy in most central banks around the globe over past decades. Relying heavily on inflation forecast, it applies point or interval target that the bank tries to achieve with tools at its disposal (Svensson, 2010). While at the start of 2010, some twenty-seven banks were considered "fully fledged" targeters (Hammond, 2012), a decade later the number grew almost two-fold only among the OECD member states. Performance of the targeting mechanism itself was subject to intense scrutiny, both in direct and indirect terms. When it comes to (lower) inflation stabilisation and economic growth, a strong majority of studies agree on its beneficial effect (e.g. Mishkin and Schmidt-Hebel, 2001; Walsh, 2009; Roger, 2010; or Bernanke et al., 2018), overriding older papers suggesting inconclusive evidence (e.g. Ball and Sheridan, 2004), even with selection bias allegedly present (Balima et al., 2020).

[1] University of Finance and Administration, Estonská 500, Prague 10, Czech Republic. Corresponding author: e-mail: sindelar@mail.vsfs.cz.
[2] University of Finance and Administration, Estonská 500, Prague 10, Czech Republic. E-mail: petr.budinsky@vsfs.cz.

In central and eastern Europe, most central banks ascended inflation targeting in the early 2000s. Out of the so-called Visegrad countries, the Slovak Republic adopted the Euro, thus becoming part of the Eurosystem, while the Czech Republic, Poland and Hungary gathered experience with their own targeting regime. Over the years, their performance was surveyed, both internally and externally. The evaluation generally gained favourable results, confirming IT success predominantly in the disinflation area (Krušec, 2014; Jonas and Mishkin, 2004; Mackiewicz-Łyziak, 2016). While many studies focus on the long-term macroeconomic benefits of the targeting itself, they seldom aspire to answer how successful the banks were in hitting the target and forecasting the inflation accurately. Relevant evaluations are rather limited in scope, focusing mostly on methodology issues (e.g. Faust and Wright, 2012), transparency (e.g. Woodford, 2013) or (positive) synergy with the IT itself (Hall and Jääskelä, 2009; Diron and Mojon, 2005). The accuracy aspects, however, are no less fundamental than the overall macroeconomic benefits of the IT mechanism, indicating monetary policy overall effectiveness and accountability. For other economic agents, these are crucial traits (Hubert, 2015).

The Czech national bank (CNB), as the main subject of our study, regularly evaluates its forecasting and targeting performance in periodic inflation reports (CNB, 2020). Aside from these, dedicated one-time assessments are also conducted, indicating undershooting of target in the first decade of IT (Šmídková, 2008), along with potential biases in forecasts (Babecký and Podpiera, 2008) yet with improving accuracy over time (Antal et al., 2008). The mechanism experience on a longer time-frame mostly confirmed those results (Rusnok, 2018), with the Polish (Grostal and Niedźwiedzińska, 2019) central bank reaching a similar conclusion – i.e. the target was mostly undershot, forecasting error gradually decreased over time. On the contrary, the Hungarian experience on the 15-year IT period led to mostly undershooting of the target (MNB, 2020), providing a different record of accomplishment.

With respect to the above, this paper seeks to comprehensively evaluate the accuracy of inflation targeting & forecasting in the Czech, Polish and Hungarian central banks. In order to achieve this, the paper is divided into four sections. First, we define data sources and describe the inflation targeting mechanism, along with methods used in the analysis. Then, results of the analysis are presented, in terms of both error measures and their significance testing (systemic bias, differences between institutions, and improvement over time). The final part consists of a results discussion, along with a synthesis of conclusions and policy recommendations.

1 DATA AND METHOD

We use three main data sets for each of the central banks analysed, with the ECB being the final addition. The first set of data represents predictions of annual inflation produced by each institution in its preceding year autumn forecast (usually September–November). This forecast was chosen because of its vital role in other agents decision-making (Jain and Sutherland, 2020), and also because it combines a 14–16 month horizon of optimal monetary policy transmission in the final quarter, along with shorter horizons in the preceding quarters of the year being forecast.[3] Second, the dataset is comprised of inflation targets set by banks in individual years. The third line represents real inflation data, with a precise indicator being chosen for each bank according to its forecasting & targeting methodology. Although inflation data are usually not subject to later revisions, we used values produced by the autumn report in the following year (i.e. first outturn).

As evinced by Table 1, the forecasting horizon evaluated with the inflation predictions oscillates between 15 months (September forecasts) and 13 months (December forecasts). Regarding range, we always use the maximum length of the time-line available, marking first natural distinction from many

[3] Because the CNB predicts inflation only on a quarterly basis, the annual forecast has been created synthetically as a simple average of these quarterly forecasts for a given year.

Table 1 Data specification

Subject	Forecast data		Inflation target data		Real inflation data	
	Data set range	**Source**	**Data set range**	**Indicator targeted**	**Data set range**	**Source**
Czech national bank (CNB)	2006–2020	Autumn forecast (2006–2008 October, 2009 on November)	1998–2020	Net inflation (1998–2001) Monetary policy-relevant inflation (2002–2020)*	1998–2020	Inflation report (CNB, 2020) (2006–2008 October, 2009 on November)
Hungary national bank (MNB)	2002–2020	Autumn forecast (2002–2011 November, 2012 on September)	2001–2020	Consumer price index (CPI) (2001–2020)	2001–2020	Inflation report (MNB, 2020) (2002–2011 November, 2012 on September)
Polish national bank (NBP)	2006–2020	Autumn forecast (2006–2020 October)	1999–2020	Consumer price index (CPI) (1999–2020)	2001–2020	Inflation report (NPB, 2020) (1999–2020 October)
European central bank (ECB)	2001–2020	Autumn forecast (2001–2020 September)	2001–2020	Harmonised consumer price index (HCPI) (2001–2020)	2001–2020	ECB statistics database (ECB, 2020)

Note: * The CNB targeted net inflation in 1998–2001. In 2001–2007, it targeted overall inflation, yet with a permanent subtraction of the primary effect of indirect tax change, hence using *de facto* monetary policy-relevant inflation. This was explicitly targeted from 2008 onwards (CNB, 2020).
Source: Own research

older, more restricted studies (e.g. Antal et al., 2008; or Jonas and Mishkin, 2004). Because of different timing of targeting/forecasting, this creates slightly longer datasets for some institutions than for others, which needs to be taken into account during the results interpretation. As our first analytical step, we used comprehensive battery of three forecasting errors covering both magnitude of forecasting error, systemic bias and performance in changes. This marks the second distinction from the available studies (e.g. Babecký and Podpiera, 2008), which rely on a limited range of measures, typically absolute and relative errors. Denoting Y_t as the real value in the year being forecast (targeted) and F_t as the forecast (target) value, we define forecasting error E_t as $(Y_t – F_t)$, leading to the following error measures definitions:

- Mean Absolute Error (MAE),

$$MAE = mean(|E_t|),$$

- Root Mean Squared Error (RMSE),

$$RMSE = \sqrt{mean\ (E_t^2)},$$

- Mean Average Scaled Error (MASE),

$$RMSE = mean\ \frac{E_t}{\frac{1}{n-1}\sum_{i=2}^{n}|Y_i - Y_{i-1}|}.$$

Finally, we used battery of statistical methods to identify significant patterns inside data, in three crucial dimensions. In order to determine systemic bias, we utilised **Wilcoxon signed rank test** (Wilcoxon, 1945) testing the null hypothesis that the distribution of forecasting errors is symmetric around zero against the alternative hypothesis that such distribution is not symmetric around zero. Potentially significant differences between the forecasting errors of individual institutions were evaluated using **Kruskal-Wallis test**[4] testing null hypothesis that the distribution of forecasting errors is the same for all individual institutions against the alternative hypothesis that such distributions are not the same. Lastly, in order to detect incremental improvement, i.e. whether the average error size reduces as the horizon advances, the **Page trend test** was applied (Page, 1963) to test the null hypothesis that there is no shift in the forecasting error between the two studied horizons against the alternative hypothesis that such shift exists. P-values less than 0.05 were considered statistically significant and the analysis was conducted using the R statistical package, version 3.2.3.

2 RESULTS

How did the individual institutions fare, when it comes to the "raw" forecasting and targeting accuracy? Let us begin the results presentation with the Czech national bank (CNB).

Table 2 Error measures – CNB

Period	MAE		RMSE		MASE	
	Forecast	Inflation target	Forecast	Inflation target	Forecast	Inflation target
1998–2000	-	2.24%	-	2.82%	-	0.781
2001–2005	-	1.74%	-	2.14%	-	1.162
2006–2010	1.13%	1.18%	1.23%	1.28%	0.977	1.017
2011–2015	0.69%	1.36%	0.73%	1.51%	0.883	1.744
2016–2020	0.62%	0.60%	0.64%	0.80%	0.733	0.714
Total period	**0.81%**	**1.35%**	**0.90%**	**1.74%**	**0.877**	**1.038**

Source: Own research

As evinced by Table 2, aside from total period results, we also divided the timeframe into five shorter – mostly five year – intervals. Overall, the results indicate three main findings. Firstly, the error magnitude generally decreased over the years (both simple and squared), with the largest values being found in the first two sub-periods. In most of the years, real inflation undershot the target, as well as the predicted value. Second, the forecast errors were on average notably lower than the deviations connected to the inflation target, which in turn generally surpassed a one percentage-point toleration interval. Finally, scaled MASE metrics has shown, that while the bank made predictions in every period surpassing the accuracy of the in-sample naïve forecast (i.e. MASE value lower than 1), the naïve version of the inflation target (i.e. previous year inflation real value) made a better predictor than the actual inflation target. This implies that inertia inside the system might have been sometimes stronger than the monetary policy tools applied by the bank.

The Hungarian central bank results presented a largely different image, both in total (shorter) period and individual sub-periods.

The MNB exhibited slightly worse performance when it comes to its MAE & RMSE forecast errors and to its ability to target inflation. Contrary to the CNB, it was generally able to forecast and target inflation more accurately in the earlier post-transformation years, yet exhibited visibly worse accuracy in later periods. Its forecasting performance vs. the naïve benchmark was notably better, with forecast

[4] Traditional methods, i.e. Diebold-Mariano test, exhibit substantial problems in dealing with serial persistence present in the data (rejecting null too often – oversized type I error), making them unsuitable – see Christensen (2007) for details.

Table 3 Error measures – MNB

Period	MAE		RMSE		MASE	
	Forecast	Inflation target	Forecast	Inflation target	Forecast	Inflation target
2001–2005	0.92%	1.46%	1.07%	1.63%	0.444	0.704
2006–2010	0.56%	1.48%	0.69%	1.79%	0.445	1.182
2011–2015	1.53%	1.69%	1.74%	2.04%	0.900	0.999
2016–2020	0.73%	0.88%	0.86%	1.24%	0.836	1.007
Total period	**0.93%**	**1.38%**	**1.16%**	**1.70%**	**0.618**	**0.935**

Source: Own research

mostly overshooting real inflation, which was in turn generally higher than the inflation target. MASE value of inflation targeting again suggested high inertia inside the economy, almost making the previous year inflation a better predictor than the target itself.

The Polish experience with inflation forecasting and targeting was probably the most volatile among the institution set, with forecasting errors highest across the field.

Table 4 Error measures – NBP

Period	MAE		RMSE		MASE	
	Forecast	Inflation target	Forecast	Inflation target	Forecast	Inflation target
1998–2000	-	2.05%	-	2.83%	-	0.562
2001–2005	-	1.64%	-	1.89%	-	0.612
2006–2010	0.91%	0.86%	1.00%	1.11%	0.769	0.729
2011–2015	1.44%	2.10%	1.52%	2.24%	1.043	1.522
2016–2020	0.90%	1.10%	1.00%	1.51%	0.900	1.100
Total period	**1.12%**	**1.48%**	**1.23%**	**1.86%**	**0.933**	**0.847**

Source: Own research

Both simple and squared errors of the Polish NBP forecasts and target deviations markedly surpassed one percentage point. Inflation undershot both projections in the given period, marking similarity with the CNB, a fact in the 2011–2015 five-year period. Regarding the performance versus the naïve forecast, MASE values suggest a slightly better performance in the targeting sphere, with notably worse performance related to inflation forecasting. Here, the value-added by the bank´s prediction compared to the in-sample naïve benchmark was very limited, albeit on the shortest timeframe.

The ECB formed a sort of control element in our central bank sample. As such, it was evaluated on a shorter range of data, similarly to the NBP, yet still with interesting results.

With the aforementioned shorter data set in mind, we can summarize that both MAE and RMSE measures indicate the ECB´s superior accuracy among surveyed institutions. Simple and squared errors did not surpass a one percentage point deviation, with the real inflation again mostly undershooting the forecast and inflation target. The MASE metrics, however, paints a different picture. The in-sample naïve forecast was found to be a better inflation predictor than the bank´s own forecasts and set inflation target. This suggests inferior forecasting value added as well as questionable monetary policy effectiveness.

Table 5 Error measures – ECB

Period	MAE		RMSE		MASE	
	Forecast	Inflation target	Forecast	Inflation target	Forecast	Inflation target
1998–2000	-	-	-	-	-	-
2001–2005	-	0.20%	-	0.22%	-	2.000
2006–2010	0.92%	0.74%	1.21%	0.98%	0.821	0.661
2011–2015	0.82%	1.04%	0.84%	1.17%	1.139	1.444
2016–2020	0.48%	1.00%	0.57%	1.19%	0.774	1.613
Total period	**0.65%**	**0.75%**	**0.82%**	**0.97%**	**1.008**	**1.164**

Source: Own research

Following the evaluation of error deviances, we now approach the crucial question of suggested traits´ statistical significance. Table 6 summarizes the first batch of tests undertaken in this regard.

Table 6 Wilcoxon test and Page trend test results (p-values)

Period	Wilcoxon test		Page trend test
	Forecast	Inflation target	
CNB	0.639	0.005	0.342
MNB	0.258	0.133	0.002
NBP	0.140	0.313	0.033
ECB	0.913	0.217	0.967

Source: Own research

The introductory group of tests is concerned with the existence of systemic bias (Wilcoxon test) and gradual improvement of accuracy from the inflation target to the inflation forecast (Page trend test). In the first aspect, the hypothesis of the systemic character of the forecasting error was not rejected in regards to the inflation targeting of the CNB. With all other items reaching the opposite outcome, our results suggest that only the Czech central bank error-pattern contained systemic bias, related to the aforementioned real inflation undershooting the target. There was no such finding associated with the inflation forecasts surveyed. As for incremental improvement, Page test suggests that with two institutions, the MNB and the NBP, the quality of their projections improved with a shortening horizon between (longer-term) the inflation target and the inflation forecast.

Table 7 Kruskal-Wallis test results (p-values)

Kruskal-Walis test – forecast (whole sample)			0.654
Kruskal-Walis test – infl. target (whole sample)			0.014
IT differences decomposition	**CNB**	**ECB**	**MNB**
ECB	0.274	-	-
MNB	0.020	0.091	-
NBP	0.646	0.798	0.274

Source: Own research

Was there a significant difference between the forecasting/targeting accuracy of individual institutions? Results in this regard are summarized in Table 7.

The application of the K-W test across the whole sample indicated a significant difference between the accuracy of inflation targeting among our central banks. Specifically, the CNB was found to be a significantly more accurate targetter than the MNB, with the ECB nearing a similar result (p-value would be significant at 0.1 level). That was, however, the sole case we have detected.

3 DISCUSSION

Our results point out a rather positive picture of the inflation forecasting and targeting performance of the surveyed central banks. The direction of the targeting error mostly confirms a disinflation tendency of the system suggested by Krušec (2014), Jonas and Mishkin (2004), or Mackiewicz-Łyziak (2016). In the majority of the institutions, the magnitude of the forecasting and targeting (absolute) error gradually decreased to a circa one percentage point in the post-2015 period, with conducted tests not verifying the inflation-forecast bias indicated earlier (Babecký and Podpiera, 2008). Worse outcomes, however, were also found. Some central banks, namely the NBP and the ECB, struggled to make their inflation target a better predictor than the naïve in-sample forecast, utilised by the MASE metric. This might lead to concerns over the effectiveness of the targeting concept as a whole, outlined by e.g. Ball and Sheridan (2004). Systemic bias detected with the CNB´s inflation targeting accuracy add to this scepticism.

Comparing the performance of individual central banks is precarious. The results should not only be interpreted with different data-ranges in mind (the CNB being analysed on the longest one), but also with respect to a different inflation targeting framework being used. This important variable, as evinced by e.g. Baxa et al. (2015), determines not only the resulting accuracy, but also its subsequent interpretation. While the CNB, for example, centres its policy actions to achieve the inflation target, which is typically identical to the same-horizon forecast, other banks employ different paradigms. For the ECB both forecast and target diverge frequently and the MNB with the NPB have a wider set of aspirations in their function. With this in mind, our results show that the CNB is a significantly more successful targetter than the MNB, when it comes to overall accuracy on the surveyed period. Other significant differences were not found, adding to the general targeting-success-thesis formulated earlier.

Comparing the accuracy of inflation forecasting and inflation targeting between themselves yielded interesting outcomes. The banks in question exhibited decent forecasting accuracy, compared to their earlier scores (Roger and Stone, 2005) or GDP predictions segment (Šindelář, 2017). No systemic bias and some incremental improvement on shortening the time horizon were observed. While forecasting itself can be considered a purely analytic exercise, inflation targeting is where real monetary actions come into play. The results were more flawed with this activity, indicating not only higher error deviances (frequently over toleration interval), but also sub-par performance versus the naïve benchmark and in a single case, systemic bias. Successfully executing monetary policy through such optics is obviously challenging for CEE countries and even though the ECB reached lower absolute deviances from its inflation target, it did not represent a statistically significant difference.

CONCLUDING REMARKS

The goal of this paper was to evaluate the effectiveness of CEE central banks in inflation forecasting and targeting. The evaluation was done on the basal accuracy level and as such, it indicates a rather satisfactory result. Our analysis shows that while the Czech, Polish and Hungarian central banks struggled with forecasting and targeting accuracy in the earlier period, since 2015 they have become notably more efficient in this regard – including the Covid-19 affected year 2020. The ECB performance, on the contrary, lacked such an improvement trajectory, despite attaining comparable error sizes and not differing significantly from the rest of the sample.

The issues worth further investigation contain systemic bias found in the CNB targeting track record and subpar performance versus the naïve benchmark, mainly in the ECB case (both forecasting and inflation targeting). In the CNB context, the thesis of a small open-economy central bank not being able to decisively execute the monetary policy seems disproved by the MNB and the NBP results. In the Eurozone, though, the ECB´s outcome is puzzling. Our study here is bound by a strict focus on the accuracy itself and a wider macroeconomic investigation of the problem is viable. Incorporating particularly economic growth and potentially unemployment as the most policy (and politically) sensitive elements.

ACKNOWLEDGEMENTS

This paper was supported by institutional support for long-term conceptual development of research organization University of Finance and Administration.

References

ANTAL, J., HLAVÁČEK, M., HORVATH, R. Prediction Bias and Undershooting of the Inflation Target. In: *Evaluation of the Fulfilment of the CNB's Inflation Targets 1998–2007*, Czech National Bank, 2008.

BABECKÝ, J. AND PODPIERA, J. Inflation forecasts errors in the Czech Republic: Evidence from a panel of institutios. In: *Evaluation of the Fulfilment of the CNB's Inflation Targets 1998–2007*, Czech National Bank, 2008.

BALIMA, H. W., KILAMA, E. G., TAPSOBA, R. Inflation targeting: Genuine effects or publication selection bias? *European Economic Review*, 2020, 128, pp. 103–120.

BALL, L. M. AND SHERIDAN, N. Does inflation targeting matter? In: *The inflation-targeting debate*, University of Chicago Press, Ill., 2004, pp. 249–282.

BAXA, J., PLAŠIL, M., VAŠÍČEK, B. Changes in inflation dynamics under inflation targeting? Evidence from Central European countries. *Economic Modelling*, 2015, 44, pp. 116–130.

BERNANKE, B. S., LAUBACH, T., MISHKIN, F. S., POSEN, A. S. *Inflation targeting: lessons from the international experience.* Princeton, NJ: Princeton University Press, 2018.

CIŻKOWICZ-PĘKAŁA, M., GROSTAL, W., NIEDŹWIEDZIŃSKA, J., SKRZESZEWSKA-PACZEK, E., STAWASZ-GRABOWSKA, E., WESOŁOWSKI, G., ŻUK, P. Three decades of inflation targeting. *NBP Working Paper No. 314*, Narodowy Bank Polski, 2019.

CZECH NATIONAL BANK (CNB). *Inflation reports* [online]. CNB, 2021. <https://www.cnb.cz/cs/menova-politika/zpravy-o-inflaci>.

DIRON, M. AND MOJON, B. Forecasting the central bank's inflation objective is a good rule of thumb. *ECB Working Pape*r, No. 564, European Central Bank, 2005.

EUROPEAN CENTRAL BANK (ECB). *Inflation forecasts* [online]. ECB, 2021. <https://www.ecb.europa.eu/stats/ecb_surveys/survey_of_professional_forecasters/html/table_hist_hicp.en.html>.

FAUST, J. AND WRIGHT, J. H. Forecasting inflation. In: *Handbook of economic forecasting*, Elsevier, 2013, Vol. 2, pp. 2–56.

HALL, J. AND JÄÄSKELÄ, J. Inflation Volatility and Forecast Accuracy. *Research Discussion Paper 2009–06*, Reserve Bank of Australia, 2009.

HAMMOND, G. *State of the art of inflation targeting.* 4th Ed. Centre for Central Banking Studies, Bank of England, 2012.

HUBERT, P. Do central bank forecasts influence private agents? Forecasting performance versus signals. *Journal of Money, Credit and Banking*, 2015, 47(4), pp. 771–789.

CHRISTENSEN, J. H., DIEBOLD, F. X., RUDEBUSCH, G., STRASSER, G. Multivariate Comparisons of Predictive Accuracy. *University of Pennsylvania working paper*, University of Pennsylvania, 2007.

JAIN, M. AND SUTHERLAND, C. S. How do central bank projections and forward guidance influence private-sector forecasts? *Bank of Canada Staff Working Paper No. 2018-2*, Bank of Canada, 2018.

JONAS, J. AND MISHKIN, F. S. Inflation targeting in transition economies experience and prospects. In: *The Inflation-Targeting Debate*, University of Chicago Press, Ill., 2004, pp. 353–422.

KRUŠEC, D. Is inflation targeting effective? Monetary transmission in Poland, the Czech Republic, Slovakia, and Hungary. *Eastern European Economics*, 2011, 49(1), pp. 52–71.

MACKIEWICZ-ŁYZIAK, J. Active and passive monetary policy in CEE countries with inflation targeting: the case of the Czech Republic, Hungary, and Poland. *Eastern European Economics*, 2016, 54(2), pp. 133–152.

MAGYAR NEMZETI BANK (MNB). *Inflation reports* [online]. MNB, 2021. <https://www3.mnb.hu/en/publications/reports/inflation-report>.

MISHKIN, F. S. AND SCHMIDT-HEBBEL, K. One decade of inflation targeting in the world: What do we know and what do we need to know? *Working paper 8397*, NBER, Cambridge, MA, 2001.

NARODOWY BANK POLSKI (NBP). *Inflation reports* [online]. NBP, 2021. <https://www.nbp.pl/homen.aspx?f=/en/publikacje/raport_inflacja/raport_inflacja.html>.

PAGE, E. B. Ordered hypotheses for multiple treatments: a significance test for linear ranks. *Journal of the American Statistical Association*, 1963, 58(301), pp. 216–230.

ROGER, S. Inflation targeting turns 20. *Finance and development*, 2010, 47(1), pp. 46–49.

ROGER, S. AND STONE, M. On target? The international experience with achieving inflation targets. *IMF Working paper 05/163*, Washington D.C., USA: IMF, 2005.

RUSNOK, J. Two Decades of Inflation Targeting in the Czech Republic. *Finance a Uver*, 2018, 68(6), pp. 514–517.

SVENSSON, L. E. Inflation targeting. In: *Handbook of monetary economics*, Elsevier, 2010, Vol. 3, pp. 1237–1302.

ŠINDELÁŘ, J. GDP forecasting by Czech institutions: an empirical evaluation. *Prague Economic Papers*, 2017, 26(2), pp. 155–169.

ŠMÍDKOVÁ, K. *Evaluation of the Fulfilment of the CNB's Inflation Targets 1998–2007*. Czech National Bank, 2008.

WALSH, C. E. Inflation targeting: what have we learned? *International Finance*, 2009, 12(2), pp. 195–233.

WILCOXON, F. Individual comparisons by ranking methods. *Biometrics*, 1945, 1(6), pp. 80–83.

WOODFORD, M. Forward guidance by inflation-targeting central banks. *Discussion Paper No. 1314–15*, Columbia University, NY, 2013.

Sensitivity Analysis of Price Indices in Models of Demand Systems

Renata Benda-Prokeinová[1] | *Slovak University of Agriculture in Nitra, Slovakia*
Martina Hanová[2] | *Slovak University of Agriculture in Nitra, Slovakia*
Natalia Turčeková[3] | *Slovak University of Agriculture in Nitra, Slovakia*

Abstract

The primary motivation of the paper is to point out the sensitivity of price indices calculated by the model of demand systems in different price and quantitative levels. We simulated different prices and quantities, so increasing the values of the variables reduced their variance. We would like to point out the significant role of variability and thus the deviation of input data, which is the basis for identifying consumer behavior.

From the methodology, we used the Linearized Almost Ideal Demand System and focused on the partial output of specific expenditure elasticities calculated by price indices – Stone, Laspeyres and Törnqvist index. Following we are wondering which index can consider as the most trustworthy?

In the analysis, we realized that the price variance would affect the indices' values more significantly, than the more considerable variance of the quantity consumed. It means that elasticities characterize consumer behavior in terms of prices, not in terms of quantity consumed.

Keywords	***JEL code***
Variance, expenditure elasticity, consumer demand model, meat items	*D11, D12, C43*

INTRODUCTION

The review of the literature is focusing on the presentation of experience with price indices calculated in models of demand systems.

The Linearized Almost Ideal Demand System (AIDS) developed by Deaton and Muellbauer (1980) remains one of the most popular systems in the applied demand analysis. This seeming nonlinear system's linearization usually uses the Stone price index to approximate the model's accurate price index, Buse and Chan (2000). Deaton and Muellbauer (1980) found that the Stone price index provided an excellent approximation to the valid price index given high prices' high positive collinearity.

1 Department of Statistics and Operation Research, Faculty of Economics and Management, Slovak University of Agriculture in Nitra, Tr. A. Hlinku 2, 949 01 Nitra, Slovakia. Corresponding author: e-mail: renata.prokeinova@uniag.sk.
2 Department of Statistics and Operation Research, Faculty of Economics and Management, Slovak University of Agriculture in Nitra, Tr. A. Hlinku 2, 949 01 Nitra, Slovakia. E-mail: martina.hanova@uniag.sk.
3 Department of Economics, Faculty of Economics and Management, Slovak University of Agriculture in Nitra, Tr. A. Hlinku 2, 949 01 Nitra, Slovakia. E-mail: natalia.turcekova@uniag.sk.

The choice of price index did not influence the likelihood ratio tests. Anderson and Blundell (1987) confirmed these features. Buse and Chan (2000) stated that Pashardes (1993), and Buse (1994; 1998) were interested in the presumption that the Stone index will always provide a good approximation. Buse (1994) showed that the Seemingly Unrelated Regression (SUR) estimator was inconsistent and that it was impossible to derive a consistent instrumental variable estimator for this model. In a subsequent study of tests of homogeneity in AIDS, Buse (1998) examined the role of price collinearity and concluded that high collinearity was neither necessary nor sufficient to obtain a good approximation of the actual index by the Stone index.

Moschini (1995) offered and described other deficiency of the Stone index. He referred that if the prices were given in natural units, as opposed to indices, the Stone index and derived parameters such as elasticities are not invariant to measurement units. Three alternative indexes proposed to deal with this deficiency. Using a single experiment in Monte Carlo, Moschini confirmed the invariant indices' adequacy, as only the Stone index could not represent a suitable approximation of the right elasticity of expenditure. The question arises as to whether these results are more general or whether Moschini's results are specific to a particular Monte Carlo design. Considering variations in properties, such as price collinearity and sample size change, could well use to calculate the Stone index.

Furthermore, it is necessary to examine the Stone index's sensitivity to measurement units' changes to obtain some meaningful final range of values that could adequately described by the previously published result. Finally, although the Stone index appears to be generally lower, the question remains whether all invariant indices are as good as Moschini's results suggest. The validity of this question turns out when we note that Moschini's advocacy of average price escalation (variants of the Paasche index) has been uncritically accepted in the literature; they can see Mittelhammer, Shi and Wahl (1996), Asche, Bjorndal, Salvanes (1998), and La France (1998).

The main motivation of the contribution is to point out the sensitivity of price indices calculated using the model of demand systems into the context of different levels of price and quantity. We simulated different prices and quantities, so increasing items causing the reduction of their variability. Besides, the variability we consider as a key factor impact on behavior represented by price indexes. The aim of the contribution is not to compare different models of AIDS. We focus on the partial output of specific expenditure elasticities calculated by price indices.

1 METHODOLOGY

It is necessary to describe methodologically price indices, representing partial calculations to the final output of demand system models (in our case, we computed LA/AIDS-Linearized Approximation of the Almost Ideal Demand System) in software R 3.6.3 (Henningsen and Hamann, 2007). Defectively, three basic price indices are generating. There is confusion about which of the price indices is relevant and suitable for interpreting the given situation's decision-maker. For this reason, we decided to describe them in detail and specify their advantages and disadvantages.

Contributions of LA/AIDS models, it is a standard linearization model implementing through the Stone nonparametric index to obtain approximate income deflator parameters. This approximation has some drawbacks. Pashardes (1993) pointed out that this mistake could be understood based on an approximation because of a missed variable. The resulting estimates of query function parameters

Summary 1 Summary about AIDS price indexes

Paasche price index	Calculates the price change between two periods by comparing the cost of purchasing in each period the bundle of goods purchased in the final period
Laspeyres price index	Calculates the price change between two periods by comparing the cost of purchasing in each period the bundle of goods purchased in the initial period
Tornqvist price index	Is a weighted average of the growth rates of the individual prices, with weights equal to the average of the expenditure shares in the two periods used to compute the growth rates

Source: Diewert (1987)

may be biased. Besides, Moschini (1995) pointed out that the Stone index failed to satisfy ownership "adequacy" because the price index's growth rate is not invariant for the unit of measurement of prices. Moschini proposes three alternative indexes: two normalized forms of the Stone index and the Törnqvist index. The indices in Summary 1 will be described in detail in the following chapter.

1.1 Paasche index (the first modified Stone index)

The Stone index typically used in estimating linear almost ideal demand systems is not invariant to changes in measurement units, which may seriously affect the model's approximation properties. A modification to the Stone index or a regular price index is both desirable practices in estimating linear, almost ideal models.

As Moschini (1995) points out, the Stone index can be adjusted to be invariant to unit rates. It is referred to as a modified Paasche index because it uses the current period of expenditure weights. The resulting price index in logarithms is:

$$\log P_t^P = \sum_{k=1}^{n} w_{kt} \log \frac{p_{kt}}{p_k^0} . \tag{1}$$

Using this price index, the price elasticity is slightly different from the elasticity that formed the Stone index.

Assuming that prices will never be completely collinear, the literature states that a variable called measurement error was introduced when using the Stone index (see Alston, Foster, Green, 1994; Asche and Wessells, 1997; Moschini, 1995). Index Stone does not respect the basic property of index numbers because it changes in measurement units of prices. One solution for correcting units of measurement error is that prices is adjusted by averaging. Moschini (1995) proposes in his paper to use the Laspeyres price index to overcome measurement error because it is a change in units of measurement.

1.2 Laspeyres index (the second modified Stone index)

Historically, the Laspeyres index was created as the first price index. The formula, which formed in 1871 the German economist Etienne Laspeyres:

$$P_L = \frac{\sum(p_t \cdot q_0)}{\sum(p_0 \cdot q_0)} , \tag{2}$$

calculates the price change between two periods by comparing the cost of purchasing in each period the bundle of goods purchased in the initial period, according to Wynne and Sigalla (1994).

Index measures the change in the price of the basket of goods and services concerning the set weights of the base period. The Laspeyres price index is referred to as a method of weighing quantities in a base year.

Moschini (1995) also proposed a Laspeyres modification, P^L. This modification is analogous to the Laspeyres index in logarithms, with weights calculated based on the expenditure period under review. The index is the following:

$$\log P_t^L = \sum_j^n w_j^0 \log \frac{p_{jt}}{p_j^0} . \tag{3}$$

Then the expression holds:

$$\frac{\log \delta p_t^L}{\log \delta p_{jt}} = w_j^0 . \tag{4}$$

Price elasticity of goods i concerning the price of goods j given by expression:

$$\varepsilon_{ij} = -\delta_{ij} + \frac{\gamma_{ij}}{w_{it}} - \frac{\beta_i}{w_{it}} w_j^0. \tag{5}$$

The income elasticity of the goods i is expressed by:

$$\mu_i = 1 + \frac{\beta_i}{w_{it}}, \tag{6}$$

which is the same as the expression for the complete nonlinear model of AIDS. In this case, elasticity patterns are closer to full nonlinear AIDS formula but not identical to the price elasticity.

The index's main disadvantages are that it is upward-biased and tends to overstate price increases (compared to other price indices). It tends to overestimate price levels and inflation.

1.3 Törnqvist price index

The combination (or average) of quantities in both periods is a typical character of the Fisher and Törnqvist price index. This approach seeks to overcome some of the difficulties associated with using a fixed basket at any time. If there are no clear indications that it will be better to use both periods as a basis or benchmark, a combination of both periods appears to be a reasonable compromise.

Paasche, Laspeyres, or any superlative index number can be regarded as discrete approximations to the continuous line integral Divisia index, which has some useful optimality properties from economic theory. These discrete approximations are closer to the Divisia index if the chain principle is used, argued Diewert (1993). The Törnqvist price index is a weighted geometric mean of the price relatives where the weights are the average expenditure shares in the two periods.

Törnqvist indexes are described as symmetrically weighted indexes because they treat the weights from the two periods equally.

The axiomatic approach also referred to as the test approach, consists of formulating 'desirable' properties, which price and quantity indices should satisfy. Prices and quantities of commodities thereby regard as separate variables. A set of functional equations characterizes a price (or quantity) index if it is the unique solution to this set, described by Balk and Diewert (2001).

The economic approach assumes optimization, such as cost minimization or revenue maximization, which implies a relation between prices and quantities. Within the economic approach, Diewert (1976) obtained a characterization of the Törnqvist price index, namely the economical price index corresponding to a linearly homogeneous translog unit cost or revenue function.

This note provides a characterization of the Törnqvist price index from the axiomatic approach. We consider rather broad class of aggregated price relatives. We show that the imposition of two rather natural requirements reduces this class to a single element, the Törnqvist price index. Diewert (1976) has shown that the Törnqvist index is accurate for the function of translogged unit costs:

$$\log c_0(p) = \alpha_0 + \sum_{k=1}^{n} \alpha_k logp_k + \frac{1}{2}\sum_{k=1}^{n} \sum_{j=1}^{n} \gamma_{kj}^{*} logp_k logp_j, \tag{7}$$

where:

$$\sum_{i=1}^{n} \alpha_i = 1, \sum_{k=1}^{n} \alpha_{kj} = 0, , \gamma_{ij}^{\square} = \gamma_{ji}^{\square}, \text{k} = 1, 2, \dots \text{N}.$$

The Törnqvist index is considered a discrete approximation to the already existing Divisia index, which is defined as:

$$\log P_t^T = \frac{1}{2} \sum_j (w_{jt} + w_j^0) \log \frac{p_{jt}}{p_j^0} . \tag{8}$$

And therefore:

$$\frac{\log p_t^T}{\log p_{jt}} = \frac{1}{2} \left[w_{jt} + \sum_{k=1}^{n} \log p_{kt} \frac{\delta w_{kt}}{\delta \log p_{jt}} \right] - \frac{1}{2} \sum_{k=1}^{n} \log p_k^0 + \frac{1}{2} w_j^0 = \frac{1}{2} \Big[(w_{jt} + w_j^0) \\ + \sum_{k=1}^{n} w_{kt} \log \frac{p_{kt}}{p_k^0} \frac{\delta w_{kt}}{\delta \log p_{jt}} \Big] = \left[(w_{jt} + w_j^0) + \sum_{k=1}^{n} w_{kt} \log \frac{p_{kt}}{p_k^0} (\varepsilon_{kj} + \delta_{kj}) \right], \tag{9}$$

where ε_{kj} is Marshall cross price elasticity and δ_{kj} is Kronecker delta.

By replacing the last equation of the AIDS model, an uncompensated cross-price elasticity of the demand for the goods *i* arises concerning the price of the goods *j*.

$$\varepsilon_{ij} = \delta_{ij} + \frac{\gamma_{ij}}{w_i} - \left(\frac{1}{2}\right) \frac{\beta_i}{w_i} \left[(w_{jt} + w_j^0) + \sum_{k=1}^{n} w_{kt} \log \frac{p_{kt}}{p_k^0} (\varepsilon_{kj} + \delta_{kj}) \right]. \tag{10}$$

The derivation of income elasticity is analogous:

$$\frac{\delta \log_p^T}{\delta \log_m} = \frac{1}{2} \sum_{k=1}^{n} \log p_{kt} \frac{p_{kt}}{p_k^0} \frac{\delta w_k}{\delta \log m} = \frac{1}{2} \sum_{k=1}^{n} w_{kt} \log \frac{p_{kt}}{p_k^0} (\mu_k - 1). \tag{11}$$

We consider the Divisia Index called Törnqvist (1936) to be a theoretical construct of a continuously weighted sum of the growth rates of the individual components. Scales represent the share of a component in the total value. The so-called growth rates are determined for the Törnqvist index as the difference in the natural logarithms of the successive observations of the components (their log-change). The Divisia or Törnqvist indices have advantages over constant base year weighted indices because they include changes in both purchased quantities and relative prices when relative input prices change.

The Törnqvist index is excellent; thus, it can approximate a continuous production or cost function. "Smooth" means small changes in the relative prices of goodwill associated with small changes in its quantity. Törnqvist corresponds exactly to the production function of the translog, which means that with a change in prices and an optimal response in quantities, the level of the index changes exactly as well as a change in production or utility.

1.4 The Linearized Almost Ideal Demand System (AIDS)

All the above indexes are included in the calculations generating cross and own-price elasticity of demand models systems. This paper has used the demand system LA/AIDS, the linear approximate almost ideal demand system. Each equation in the AIDS given as:

$$w_i = \alpha_i + \sum_i \gamma_{ij} \, lnP_i + \beta_i \ln\left(\frac{X}{P}\right) + \mu_i , \tag{12}$$

where w_i is share of the *i*th good (because $wi = PiQi/X$), P_j is the price of the *j*th good, *X* is the total expenditure on all goods in the system, *P* si a price index, μt is the residuals, and assumed to have zero mean and constant variance, α_i, β_i, and γ_{ij} are parameters.

The price index (*P*) is a translog index:

$$lnP = \alpha_0 + \sum_i \alpha_i \, lnP_i + \frac{1}{2} \sum_j \square \, \gamma_{ij} \, lnP_i \, lnP_j . \tag{13}$$

The price index from Formula (12) makes Formula (13) a nonlinear estimation, raising estimation difficulties. To avoid nonlinear estimation, many empirical studies used Stone (1953) price index (P^*) instead of P, as suggested by Deaton and Muellbauer (1980):

$$lnp^* = \sum_i w_i lnP_i \,. \tag{14}$$

The model that uses the Stone geometric price index is called the Linear Approximate AIDS (LA/AIDS). It shows that if prices are highly collinear, then the LA/AIDS model can estimate the AIDS model's parameters because the factor of proportionally of P to P* incorporated in the intercept term (Green and Alston, 1990; Hsiao, 1986).

1.5 Characteristics data needed for LA/AIDS model and index calculations

The analysis will be performing for the period 1993–2017. For the analysis, we used data obtained from the Statistical Office of the Slovak Republic, the excel database:

- Consumption of selected kinds of foodstuffs per capita.
- Prices from the consumer prices of the consumer basket of food.

We have selected only four meat types. For a better view, we provide short descriptive characteristics of input data.

Table 1 Descriptive statistics of original input data

	beef_q	pork_q	poultry_q	fish_q	beef_p	pork_p	poultry_p	fish_p
Mean	43 687	183 726	93 565	24 916	6.44	4.98	2.32	0.73
Standard deviation	23 432.12	17 457.23	16 106.65	2 824.04	1.44	1.55	0.17	0.19
Minimum	19 171.00	151 555.00	63 081.00	20 196.00	4.07	3.04	2.04	0.42
Maximum	115 070.00	235 766.00	119 932.00	31 254.00	8.68	8.31	2.71	1.03
Skewness	1.19	0.79	–0.29	0.46	0.00	0.74	0.85	0.14
Kurtosis	1.13	1.07	–1.12	–0.53	–1.32	–0.75	0.01	–0.99

Note: p is price in EUR and q is the quantity in kg.
Source: Own calculation

We provided descriptive statistics for original and adjusted, so simulated data for deeper insight into the analysed data. From Table 1, we can review that data have different characteristics. The range between max and min values is wider. If we look at the mean and standard deviation, we can see the standard

Table 2 Descriptive statistics of adjusted input data

	beef_q	pork_q	poultry_q	fish_q	beef_p	pork_p	poultry_p	fish_p
Mean	46 696	198 912	101 420	27 136	7.0	5.3	2.5	0.8
Standard deviation	25 852.73	47 848.43	31 491.64	7 561.65	2.41	1.67	0.56	0.29
Minimum	17 253.90	126 409.50	56 142.09	16 958.25	4.07	2.51	1.68	0.43
Maximum	126 577.00	262 149.00	179 898.00	41 232.00	10.91	8.93	3.77	1.24
Skewness	1.15	**–0.21**	1.06	0.32	0.38	0.36	0.45	0.41
Kurtosis	1.26	–1.54	0.57	–1.20	–1.46	–0.47	–0.50	–1.40

Note: p is price in EUR and q is the quantity in kg.
Source: Own calculation

deviations presented by small values. Therefore, the time series of values are closer to the mean value, so they are not significantly fluctuating values. The skew and kurtosis values also indicate a descriptive condition. All the variables examined, except beef_q and pork_q, show a more pointed distribution of values.

Table 2 shows that the values of standard deviations have changed and the values of skewness and kurtosis have changed, concurrently. As an example, we mention pork_q where skewness and kurtosis have changed, and the value of kurtosis has changed in the variable poultry_p.

For clarification our simulation, we present the analyzed data in line graphs. The blue line represents the original price or the original value of consumption. The red line represents simulated changes in the original time series.

The gradual increase in beef and fish prices can follow up in Figure 1. We can evaluate that poultry prices did not increase significantly during the period under review, so they fluctuated significantly over twenty years. On the other hand, pork became cheaper during the period under review. The year-on-year change in prices is not significant. The time series of original prices showed low values of year-on-year changes, which caused similarity in the calculated values of indices in the demand model. This reason,

Figure 1 Prices and quantity of meat items-beef, pork, poultry, and fish, 1991–2018 and simulations of shocks in adjusted data of prices and quantities

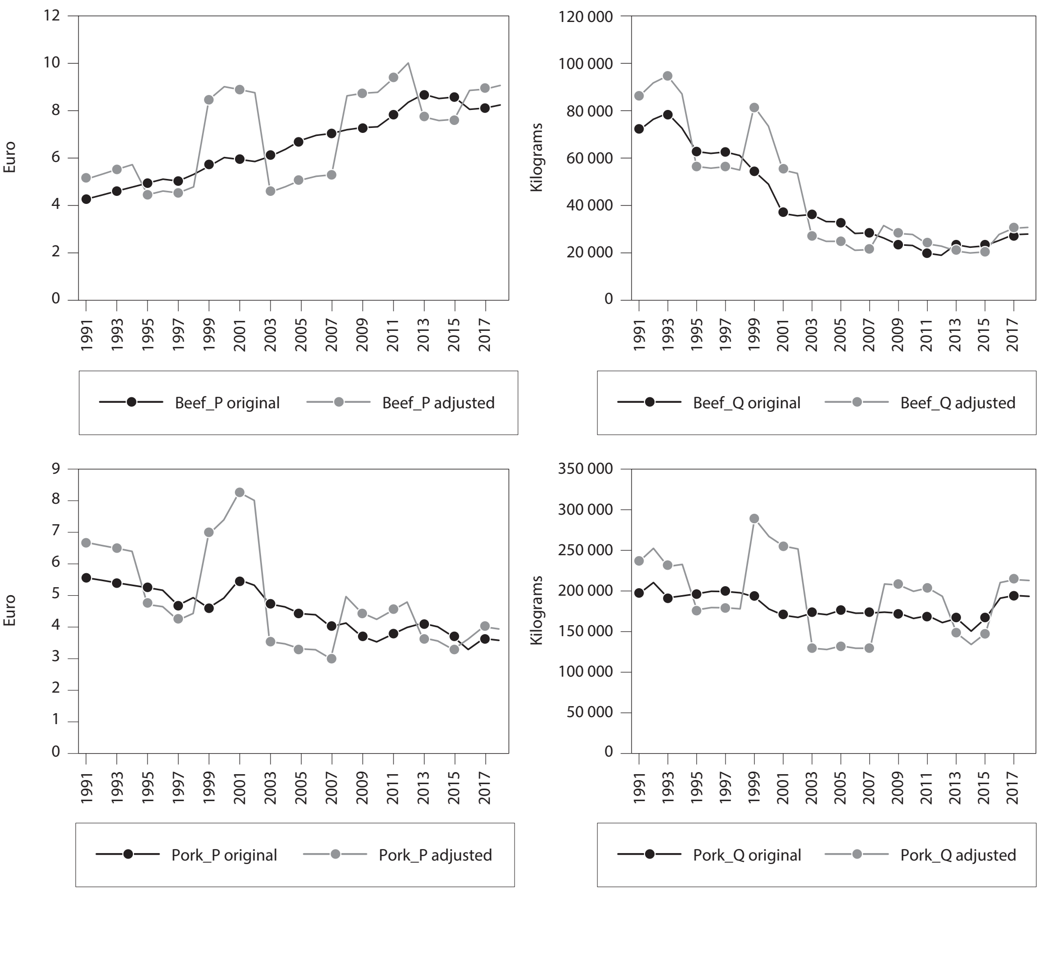
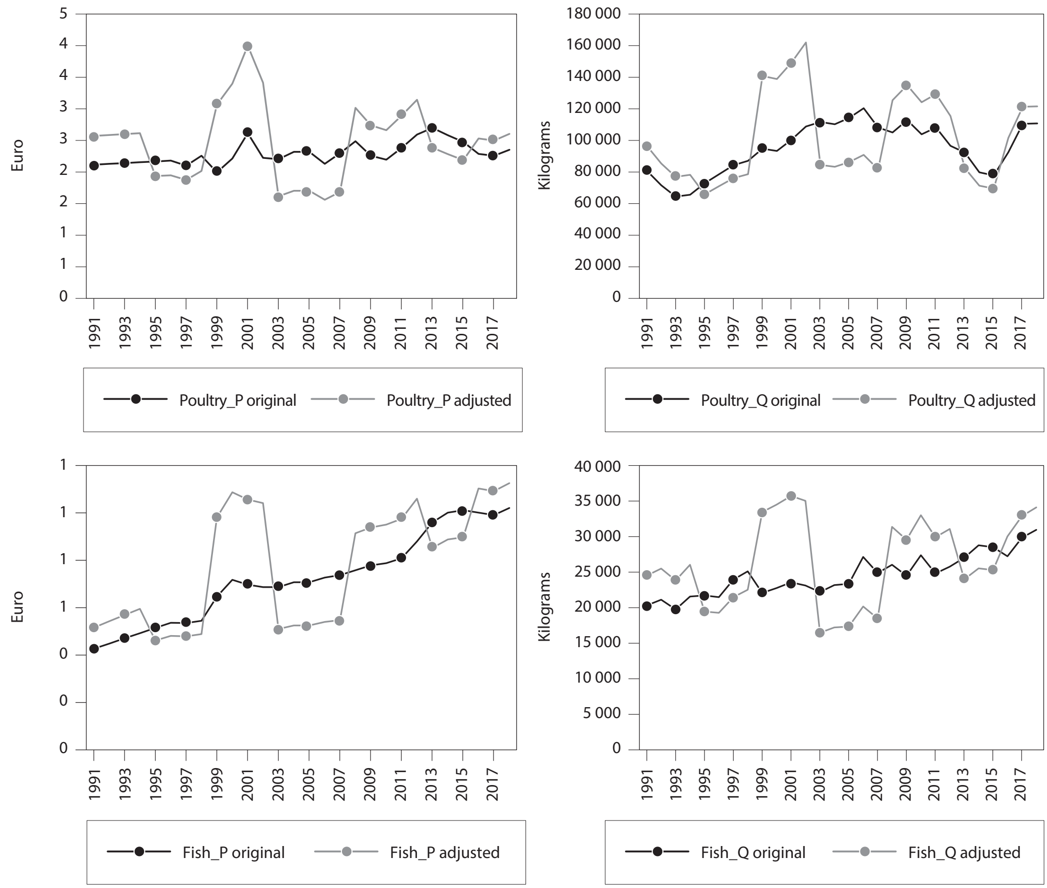


Figure 1 (continuation)



Note: p is price in EUR and q is the quantity in kg.
Source: Data from the Statistical Office of the Slovak Republic (available at: <*www.statistics.sk*>)

we decided to simulate price shocks that would affect the value of the price index. We simulated price random shocks with a 10% decrease and increased compared to the original value.

2 RESULTS AND DISCUSSION

In the results chapter we present the values of price indices of the original data and compare them with simulated data and at the same time we point out the values of price indices of other authors, for comparison and understanding of the presented situation. Following the demand model generates three different price indices and we will examine which of them presents the most relevant picture of the value of income elasticities.

2.1 Comparison of different price approximations through price indices

The analysis aims to point out the issue of variance in the researched data. The already mentioned elasticities calculation is one of the necessary moments in consumer behavior economic analysis. We decided to examine and assess specific price indices' suitability in the output AIDS model based on the studied issues.

The elasticities calculated by the AIDS model themselves indicated a situation that is identical to the real situation. It is worth mentioning that the values of the indices did not differ, although each has been calculated from a different methodological basis. We investigated what does affect this condition. The first table shows the index values of the actual data. We see that no different values were found in the indices.

Income elasticity measures the sensitivity of demand to changes in consumer income and influenced by the time during which elasticity is measured (shorter period shows lower elasticity of demand) and the level of demand for the goods (if there are more essential goods in the model, the model results from income elasticity of demand), Sloman and Norris (2002). Income elasticity can be interpreted as a percentage change in the quantity demanded if other income changes by approximately 1% while factors remain ceteris paribus. Given that the elasticity is independent of measuring units in which demand measured, consumers' most meaningful rate sensitivity demands a change in income or prices described (Benda-Prokeinova, 2016).

Table 3 Income elasticities computed by the AIDS model

	Törnqvist index	Stone index	Laspeyres
Pork	0.908564	0.908564	0.908564
Poultry	0.871678	0.871691	0.871678
Beef	1.583992	1.583982	1.583992
Fish	–0.20235	–0.20234	–0.20235

Note: Data from the Statistical Office of the Slovak Republic (available at: <*www.statistics.sk*>).
Source: Own calculations

Suppose we estimate the average level of beef expenditure for the whole sample, and the income elasticity of demand for beef was 1.58 in Table 3. In that case, a 10% increase in household income could increase beef demand by 15.8%. The lowest value of income elasticity is shown by poultry 0.88. Suppose household income increased by 10%. The demand for poultry would increase by 8.8%. Estimates of intake elasticity for the studied foods were statistically significant, evaluated Benda-Prokeinova (2016).

We can state that pork, beef, and poultry show a positive value of income elasticity of demand, which indicates that these are normal goods. According to the elasticities values, beef shows a luxury commodity, while other types of meat were income-inelastic, indicating its necessity. The model also has inferior food – fish with the value –0.20.

Table 4 Expenditure elasticities by a contribution of Buse and Chan (2000)

Panel A: Meat Data

	Stone index	Laspeyres index	Paasche index	Törnqvist index
$\eta 1$	1.8865 (0.0981)	1.3375 (0.1759)	1.3515 (0.1771)	1.3445 (0.1765)
$\eta 2$	0.8958 (0.1288)	0.4546 (0.1095)	0.4494 (0.1114)	0.4519 (0.1104)
$\eta 3$	0.7677 (0.1551)	1.4236 (0.2584)	1.3819 (0.2692)	1.4038 (0.2639)
$\eta 4$	0.8825 (0.0210)	0.9835 (0.0328)	0.9832 (0.0332)	0.9833 (0.0330)

Source: Buse and Chan (2000)

Buse and Chan (2000) investigated invariance in price indices calculated by the AIDS model. When we compare their results with our results, we find that they have differences between the indices. Our calculations have the same values, in contrast with the output Buse and Chan, where the values of indices in rows are slightly different. That's, in fact, the subject of our interest. How come that the AIDS model computes the indices of the same value? We were considering that a possible answer lies in the invariance of input data. Another reason causing similar values of price indices may lie in the trend of the examined time series. The prices of the examined meat items grew at a similar pace. It can also be one of the reasons for similar values of price indices.

For better understanding, we provided part of the research Buse and Chan paper in Table 4. Subsequently, we focused on increasing the invariance first in beef consumption, while prices remained ceteris paribus. We found that the indices' values did not differ from each other except for poultry and fish in these two commodities. Input`s prices are significantly distorted. For example, we can find the canned fish price only in the consumer basket and not for a real fish piece. A paper provided similar results from Sheng et al. (2008), where the Stone and Laspeyres index acquired different values of elasticity expenditure.

Table 5 Estimated expenditure and own-price elasticities for food items, Malaysia

Food item	LA/AIDS with Stone price index	LA/AIDS with Laspeyres price index
Rice	1.334	0.9091
Bread	0.7536	0.3177
Meat	1.0318	1.4064
Fish	0.9425	1.244
Milk and dairy	0.6284	0.8698
Eggs	1.1429	0.7675
Oils and Fats	1.1158	1.1054
Fruits	1.0325	1.0905
Vegetables	1.1955	1.1729
Sugar	0.9607	0.7458
Other	0.8651	1.4395

Source: Sheng et al. (2008)

The study provides a complete food demand system in Malaysia by analyzing data from the 2004/2005 household expenditure survey using LA/AIDS models to include the Stone price index and the Laspeyres price index. The study results suggest that the Laspeyres price index's application leads to more likely estimates of the elasticity of expenditure and own prices in Malaysia. This discovery is similar to the discovery described by Alston et al. (1994), Asche and Wessells (1997), and Moschini (1995). Therefore, this study's further implications are based on LA / AIDS models that include the Laspeyres price index. Evaluated by the research of Sheng et al. (2008) in Table 5.

For this reason, we focused only on beef, pork, poultry, and fish. Beef has experienced a renaissance in consumption over the last 20 years, and pork is the second national dish of Slovaks after dumplings. Poultry has been considered meat for the poor in Slovakia and we consume woefully low amounts of fish below the recommended doses.

If we follow the values of price indices in both simulations, we should notice one problem. We assumed that the values of the indices would not fluctuate significantly when the quantity changed. We calculated the values of price indices from data in which we simulated an increase and decrease of 10% of the quantity consumed. Thus, a significant change in the variability of meat consumption will not affect the resulting values of the indices. This assumption has been fulfilling. We conclude that

Table 6 Simulations of invariance in price and quantity in the meat items – beef, pork, poultry, and fish

	Original data			Simulation of the Beef quantity			Simulation of the Pork quantity			Simulation of the Poultry quantity			Simulation of the Fish quantity		
	Törnqvist	Stone	Laspeyres	Törnqvist	Stone	Laspeyres	Törnqvist	Stone	Laspeyres	Törnqvist	Stone	Laspeyres	Törnqvist	Stone	Laspeyres
Pork	0.9086	0.9086	0.9086	1.1514	1.1661	1.1598	0.9068	0.9038	0.9059	0.7671	0.7706	0.7681	0.8992	0.8992	0.8992
Poultry	0.8717	0.8717	0.8717	–0.9682	0.0812	–0.9522	1.37	1.12	1.99	–3.1690	1.35	–4.1086	0.7428	0.7428	0.7428
Beef	1.40	1.40	1.40	1.31	1.11	1.09	1.0675	1.0823	1.0720	0.9204	0.6814	0.9306	1.51	1.51	1.51
Fish	–0.2023	–0.2023	–0.2023	0.7266	0.7324	0.7355	0.5044	0.4983	0.5025	0.5618	–1.2179	0.5593	0.8990	0.8990	0.8990
	Original data			**Simulation of the Beef price**			**Simulation of the Pork price**			**Simulation of the Poultry price**			**Simulation of the Fish price**		
	Törnqvist	Stone	Laspeyres	Törnqvist	Stone	Laspeyres	Törnqvist	Stone	Laspeyres	Törnqvist	Stone	Laspeyres	Törnqvist	Stone	Laspeyres
Pork	0.9086	0.9086	0.9086	1.0999	0.9292	0.7754	1.1228	1.1301	0.9171	0.9878	1.0789	0.9525	0.8960	0.8960	0.8960
Poultry	0.8717	0.8717	0.8717	0.4678	0.7768	19.0109	0.8298	0.2337	2.34	0.1711	1.96	0.4808	0.6969	0.6969	0.6969
Beef	1.40	1.40	1.40	1.54	1.10	1.91	0.9230	1.81	1.26	1.33	1.92	1.14	1.63	1.63	1.64
Fish	–0.2023	–0.2023	–0.2023	0.8153	–3.3637	6.82	0.6952	0.8542	0.2910	0.6990	1.1116	0.4254	0.9278	0.9278	0.9278

Note: Data from the Statistical Office of the Slovak Republic.
Source: Own calculations

the calculated indices copy the original values. This result suggests that the resulting price index has been affected by price and not quantity. For this reason, we will not even interpret price indices calculated from the simulated quantity consumed. We will focus only on price indices calculated from simulated prices. In the second assumption, we assumed that the change in price variability would affect the values of price indices. The simulations of meat product prices has adjusted so that price shocks occurred in time series, which increased and decreased by 10%. This assumption has yielded interesting findings. If we look at the values of the beef index, it is considered a luxury item, but the values vary from 1.31 to 1.69. An interesting situation occurred in pork, where according to the outcome of the original values, pork is normal good. If we changed prices of the beef (increasing and decreasing by 10%, see Figure 1), pork has changed from normal to luxury good. Similar situation is at poultry. In the original analysis, we consider poultry as a normal good. After changing variability in beef prices, poultry is presented as inferior. Increasing the variability of original price caused by the price shocks can significantly affect the calculation of income elasticities in the AIDS model.

Subsequently, we performed the same simulation on pork and poultry. The interpretation of the results is similar. There has been a change in the income elasticities of poultry. According to the original analysis, poultry was one of the standard goods, but after the change in pork prices, we have to reclassify poultry into luxury goods. Again, we do not record changes in the values of individual indices. Subsequently, we performed the same price shocks in poultry prices, while the other variables entering the AIDS model are ceteris paribus. More important for us was the finding that the indices take on different values. If we compare the original data indices with the values of simulation data indices, we can conclude that the calculated price indices are different. From an interpretative point of view, the output of the analysis confirms the fact that if there are small changes in time series of prices represented by low variability, it is clear that the calculated price indices representing income elasticity will show a distortion, which represents very similar values in all calculated indices.

However, one peculiarity occurred again in fish. We simulated price shocks. The values of the price indices are identical. They differ from the original values, but Laspeyres and Törnqvist are the same compared to Stone. It is quite likely that the price changes did not bring about such a massive change that would affect the final value of price indices. To clarify the reasons why income elasticities react in this way, it is appropriate to see the price in Figure 1 for beef, pork, poultry and fish. Figure 1 shows prices (blue line); we see that price changes occurred gradually. No price shocks and jumps took place in Slovakia.

In the analysis, we also came to cases (as an example is the output of a beef simulation), that according to the Törnqvist index, pork meat is considered luxury good and according to the Stone and Laspeyres index it is a normal commodity. Several similar situations have arisen in the simulation of individual prices. The ambiguity of the output led us in the search for an answer to the question which of the indices can we consider the most trustworthy? After studying the detailed information about each of the indexes, we came to the answers.

As a typical index used to estimate linear almost ideal demand systems, the Stone index is not constant against changes in measurement units that can affect the model's approximation properties. In other words, if the data changes show a low variability, it is possible to use interpretations of the Stone index or its modified and improved version of the Paasche index.

The values of the Laspeyres index are skewed upwards and, as a result. It overestimates price increases (compared to other price indices) and thus overestimates price levels and inflation.

The Törnqvist index is almost "consistent", which means that the result is almost the same index values formed by combining many prices and quantities or by combining their subgroups and then combining these indices. It follows from the above that the Törnqvist index expresses the most realistic characteristic identifying the consumer's relationship to the goods under investigation (Diewert, 1976).

Balk and Diewert (2001) developed the theorem on the Törnqvist index. The price index is a unique member of the class of aggregate related prices, which has the property that it is linearly homogeneous in prices of the comparable period. At the same time, it satisfies the time-reversal test.

Buse and Chan (2000) present another approach for identification and using an appropriate index. There are some practical suggestions for applied research in the field of price indices. Suppose, for any reason; it is necessary to estimate the LAI model. In that case, examining the price correlation structure it is necessary to decide whether to use the Laspeyres or Tornqvist index. The Laspeyres index performs exceptionally well with strong positive collinearity, with the latter being better below zero or mixed collinearity. In the mostly time-series studies, price data are strongly positively correlated; Theil (l976) provides an example. We also decided to investigate the structure of price correlation as Buse and Chan (2000).

Table 7 Correlation matrix of the meat prices

Pearson Correlation Coefficients, N = 29
Prob > |r| under H0: Rho = 0

	beef_p	pork_p	poultry_p	fish_p
beef_p	1.00000	–0.89984 <.0001	0.70405 <.0001	0.97258 <.0001
pork_p	–0.89984 <.0001	1.00000	–0.40532 0.0292	–0.86770 <.0001
poultry_p	0.70405 <.0001	–0.40532 0.0292	1.00000	0.65786 0.0001
fish_p	0.97258 <.0001	–0.86770 <.0001	0.65786 0.0001	1.00000

Source: Own calculation, computed in SAS 9.4

From the analysis in Table 7 it is clear that all correlation coefficients are significant. Three coefficients acquired a negative value and three a positive value. We can talk about mixed values. According to Buse and Chan (2000) theory, based on the values of price correlations, the Tornqvist index is more suitable for further analysis. Following we verify the collinearity in the price data using the variance inflation factor (VIF) for the theories described above.

Table 8 Variance inflation factors and testing of variabilities of the meat items

	Original data		Adjusted data		
	R Square	**VIF**	**R Square**	**VIF**	**F-test of variability (p-value)**
Pork	0.801534	5.038646	0.2788	1.386578	0.0457861
Poultry	0.101108	1.11248	0.6743	3.07031	5.802E-05
Beef	0.033365	1.034517	0.1516	1.178689	2.896E-09
Fish	0.753128	4.050677	0.5682	2.315887	0.0755607

Source: Own calculations

VIF quantifies how much the scatter inflates. The standard errors – and hence the variances – of the estimated coefficients inflated when multi-collinearity exists. A variance inflation factor exists for each of the predictors in a multiple regression model (Course material, Regression Pitfalls, 2018).

According to Han (2018), the general rule of thumb is that VIFs exceeded 4-warrant further investigation, while VIFs exceeding value 10 are signs of serious multi-collinearity and is requiring correction. In our case, investigated variables do not exceed the value of VIF 10, neither in the original nor the modified data.

Based on the assumption of zero multi-collinearity, we can say that it would be most appropriate to interpret only the Tornqvist index. We evaluated this conclusion by studying the literature and assessing the input values and statistical verification's real state.

Due to the invariance of the analyzed data, we were still interested in whether there is a difference in variability between the original and modified data. It was best to test the agreement of the variables in the researched variables. Using the F-Test Two-Sample for Variances, we found a difference in the variance values between the original and the modified data. This condition seemed quite substantial.

CONCLUSION

The primary motivation of the paper was to point out the sensitivity of price indices calculated by the model of demand systems at different price and quantitative levels. We pointed out the important role of variability of input data, which is the basis for identifying consumer behavior. Followed, we assumed that the change in price variability would affect the values of price indices. This assumption was correct.

More important for us was the finding that the indices take on different values. If we compare the original data indices with the values of simulation data indices, we can conclude that the calculated price indices are different. From an interpretative point of view, the output of the analysis confirms the fact that if there are small changes in time series of prices represented by low variability, it is clear that the calculated price indices representing income elasticity will show a bias, which represents very similar values in all calculated indices (our case).

The main reason of the values of the identical price indices are small changes. The time series of original prices showed low values of year-on-year changes. The variability of price values did not change significantly in the observed period (we mean extreme increases and decreases).

Actual data suggest that all types of meat studied are constantly growing without massive price shocks. If the price of one type of meat increases, other types of meat will increase as well. The rise in price is negligible for the consumer, because the changes are in the tenths or hundreds. The observed similarity of expenditure elastics indicates that all studied types of meat have an upward trend in time series, not only in consumption but also in price.

From the simulations we found out even if the price of one product increases, it will not affect the consumer's behavior towards other goods. Extreme conditions would have to exist, such as the complete failure of the production one or two types of meat, and then a significant change in behavior could occur.

Several similar situations arose in the simulation of individual prices. We used the theory of Buse and Chan (2000) to find out which of the indices we can consider the most trustworthy. We examined the price correlation structure which is necessary for decide whether to use the Laspeyres or Tornqvist index.

All correlation coefficients are significant. Three coefficients acquired a negative value and three a positive value. We can talk about mixed values. According to Buse and Chan (2000) theory, based on the values of price correlations, the Tornqvist index is more suitable for further analysis. Similar results we gained by testing multi-collinearity. Based on the assumption of zero multi-collinearity, we can say that it would be most appropriate to interpret only the Tornqvist index.

References

ALSTON, J. M., GRAY, R., SUMNER, D. A. The wheat war of 1994. *Canadian Journal of Agricultural Economics/Revue Canadienne d'agroeconomie*, 1994, 42(3), pp. 231–251.

ALSTON, J. M., FOSTER, K. A., GREEN, R. D. Estimating elasticities with the linear approximate almost ideal demand systems. *Economic Letters*, 1994, 71, pp. 351–356.

ASCHE, F., BJØRNDAL, T., SALVANES, K. G. The demand for salmon in the European Union: the importance of product form and origin. *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie*, 1998, 46(1), pp. 69–81.

ASCHE, F. AND WESSELLS, C. R. On price indices in the almost ideal demand system. *American Journal of Agricultural Economics*, 1997, 79, pp. 1182–1185.

BALK. B. M. AND DIEWERT. W. E. A characterization of the Törnqvist price index. *Economics Letters*, 2001, 72(3), pp. 279–281.

BENDA-PROKEINOVA. R. AND HANOVÁ, M. Consumer's behavior of the foodstuff consumption in Slovakia. In: *Procedia-Social and Behavioral Sciences*, 2016, 220, pp. 21–29.

BLUNDELL. R. Econometric approaches to the specification of life cycle labor supply and commodity demand behavior. *Econometric reviews*, 1987, 6(1), pp. 103–165.

BUSE, A. Evaluating the linearized almost ideal demand system. *American Journal of Agricultural Economics*, 1994, 76, pp. 78l–793.

BUSE, A. Testing homogeneity in the linearized almost ideal demand system. *American Journal of Agricultural Economics*, 1998, 80, pp. 208–220.

BUSE, A. AND CHAN, W. H. Invariance price indices and estimation in almost ideal demand systems. *Empirical Economics*, 2000, 25(3), pp. 519–539.

DEATON, A. AND MUELLBAUER, J. An almost ideal demand system. In: *The American economic review*, 1980, 70(3), pp. 312–326.

DEATON, A. AND MUELLBAUER, J. *Economics and consumer behavior*. Cambridge university press, 1980.

Detecting Multicollinearity Using Variance Inflation Factors [online]. 2020. [cit. 8.9.2020]. <https://online.stat.psu.edu/stat462/node/180>.

DIEWERT, W. E. Exact and superlative index numbers. *Journal of econometrics*, 1976, 4(2), pp. 115–145.

HAN, K. How do you perceive this author? Understanding and modeling authors' communication quality in social media. *PloS one*, 2018, 13(2), e0192061.

HENNINGSEN, A. AND HAMANN, J. D. Systemfit: A package for estimating systems of simultaneous equations in R. *Journal of statistical software*, 2007, 23(4), pp. 1–40.

LA FRANCE, JT. The silence bleating! of the lambdas: Comment. *American Journal of Agricultural Economics*, 1998, 80, pp. 221–230.

MITTELHAMMER, R. C., SHI, H., WAHL, T. I. Accounting for aggregation bias in almost ideal demand systems. *Journal of agricultural and Resource Economics*, 1996, pp. 247–262.

MOSCHINI, G. Units of measurement and the Stone index in-demand system estimation. *American Journal of Agricultural Economics*, 1995, 77, pp. 63–68.
PASHARDES, P. Bias estimates the almost ideal demand system with the Stone index approximation. *Economic Journal*, 1993, l03, pp. 908–915.
SHENG, T. Y., SHAMSUDIN, M. N., MOHAMED, Z., ABDULLAH, A. M., RADAM, A. Complete demand systems of food in Malaysia. *Agric. Econ. – Czech*, 2008, 54, pp. 467–475.
STATISTICAL OFFICE OF THE SLOVAK REPUBLIC. *Consumption of selected kinds of foodstuffs per capita* [online]. Bratislava, 2020. [cit. 8.1.2020]. <http://datacube.statistics.sk/#!/view/en/VBD_SLOVSTAT/ps2041rs/v_ps2041rs_00_00_00_en>.
TÖRNQVIST, L. *The Bank of Finland's consumer price index. What is the Laspeyres Price Index?* [online]. 1936. [cit. 25.5.2020]. <https://corporatefinanceinstitute.com/resources/knowledge/economics/laspeyres-price-index>.
WYNNE, M. A. AND SIGALLA, F. D. The consumer price index. *Federal Reserve Bank of Dallas Economic Review*, 1994, 2, pp. 1–22.

System of Consumer Price Statistics: the First Thirty Years of Russian Experience

Maria Kozlova[1] | *Ural State University of Economics, Yekaterinburg, Russia*

Abstract

The article deals with the analysis of Russian consumer price statistics as a system of absolute and relative indicators. The sampling method and design used by the national statistical office Rosstat are considered as Russian experience and adaptation of the CPI international standard. The survey of the system of consumer price statistics is realized on the basis of the indicator representativeness, the assessment of time and spatial differentiation and similarity in consumer prices. The main indicators of this system are the price indices. These include the consumer price index, fundamental consumer price index and cost of living index which are intended for dynamics analysis with additive seasonal decomposition of time series and to spatial differentiation of price level in Russian regions and cities.

Keywords	***JEL code***
Consumer price index, cost-of-living index, Russian consumer price statistics	*C43, C83*

INTRODUCTION

Consumer price statistics is one of the most needed in the sphere of macroeconomic regulations and wage adjustment by the inflation rate. It is one of the dynamically developing statistics as well. These facts determine the multitasking, representativeness, and adaptiveness of price statistics indicators. Therefore, there is a need to construct a system of statistical indicators which complement each other. This system should include time and spacial indicators, mainly index numbers. It corresponds to the multitasking of the price index number.

Russian experience in constructing price statistics indicators is not longer: consumer price data has been collected and consumer price index (CPI) compiled since 1991. These first thirty years of Russian consumer price statistics were subject to the changing conditions, namely the dramatic increase of prices, significant spatial differentiation in regional economies and imbalance of social and transport infrastructure development. This research offers a survey of Russian consumer price statistics with the assessment of indicator representativeness and analysis of consumer price indicator relationship.

[1] Assistant professor, Information technology and statistics department, Ural State University of Economics. Also Assistant professor, Econometrics and statistics department, Ural Federal University, Yekaterinburg, Russia. E-mail: kozlova_mary@mail.ru.

The consumer price statistics research focuses mainly on the CPI which has many limitations and disadvantages. The main disadvantage which is a bias is minimized by the improvement of sample design and the use of new formulae. This article focuses on a different way of consumer price statistics development and national adaptation. There is the use of the auxiliary statistical indicators constructed on the common base of samples and methods and eliminating the CPI limitations. As a result, the system of consumer price statistics indicators is formed. It becomes a means of representation of regional differences, absolute price levels, seasonal variations, and its elimination, notably meets the requirements of multitasking. In this research the consolidation of indicators of consumer price statistics is considered as a systematic phenomenon.

1 HISTORICAL BACKGROUND

The CPI as the first and main indicator of consumer price statistics has a long history, almost over a century. It was needed initially to adjust the wage of clerical workers and earners in the big industrial cities in US when the price of goods in the First World War time changed very quickly and unpredictably. The basis of the CPI construction was very simple, notably it needed to form the goods basket with constant quantities and then calculate its value every month. The result is the division of two basket values. This principle was based on the CPI and is used today. In the 1920s the basket consisted of forty nutrition goods. Later, the bundle of basket has increased by adding the new type of goods and services and extending the existing categories of food. Now it numbers more than 300 items in US, more than 500 in Russia.

In the 1920s and 1930s the consumer price index was called the cost of living index (COLI) but in 1945 the COLI was transformed into the consumer price index after publishing of Alexander Könus' paper and as a result of many discussions about the limits of calculated indicator and its name (Könus, 1939). This differentiation has played very important role in the CPI development: it was a driver to the theoretical basis development for the CPI and expansion of Könus' conception into the price indices in general. Könus' idea about the COLI had allowed to show the substance of the calculated CPI and to determine what the CPI represents. As a result of this theoretical development there was the understanding of the CPI bias from the real change of prices in the consumer market because the Laspeyres formula used for CPI calculating doesn't take into account the consumer reaction to the price changes.

The extension of Könus' conception was realized by the generation of "new" formulae which include the components numerically determining the consumer reaction (Wald, 1939; Klein and Rubin, 1947–1948; Chetty, 1971; Galatin, 1973; Fry and Pashardes, 1989; Balk, 1990). But this idea of new form of formula couldn't develop in the practice of national statistical offices because of the difficult valuation of these components on the macroeconomic level. This extension showed a new wave of the CPI realization and theoretical research of the COLI. It was the assessment of the bias between the CPI and COLI calculated by the econometric model. J. Ulmer (1946) was the first who wrote about the exact assessment of bias but it was without proof. The bias was valuated as 1.5%. Braithwait (1980), Balk (1990), Beatty and Larsen (2005) calculated the bias on the basis of the national statistics.

R. Pollak had published the series of articles about the COLI. This index number was included in the context of economics and relation with the household economy. In Pollak's research there are some propositions about the classification of price index numbers and its integration in the theory of production and consumption. Firstly, price index numbers are divided into the subindices of two types such as partial and conditional index numbers (Pollak, 1975). Its difference subsists in the relation with the complete index. It is not constructed from the conditional index numbers which include the part of bundle of goods and reflect the consumer behavior in situation of this part of bundle. On the contrary, the partial index numbers in generalization are the upper bound of complete index because there is week separability of utility function and consumer behavior is similar in the cases of the complete bundle and its part. Furthermore, Pollak includes the new variables in the cost-of-living formulation. He uses the household production of goods as a notion of technology of production in the COLI. As a result, there are the variable and constant technology COLIs, which are related to the subindices because the household technology may affect the utility function as well as the price change (Pollak, 1978). The idea of subindex was adjusted in the question of aggregation process (Blackorby and Russell, 1978). Pollak's conception of subindex has undergone the development in the group COLI as well (Pollak, 1980). These COLIs are concerned the households of other types and thereafter the consumer patterns.

The development of the CPI and COLI conception was realized on the basis of consumer theory and the next step was a finding of formula which reflects the consumer behavior and does not need the additional statistical information and complex assessment. The class of superlative indices met these requirements. The scientific explanation of superlative index was related to the form of utility function and consumer reaction on the price change. The superlative index corresponds with the translog cost function which assesses the substitution effect connected with the price change better than Laspeyres and Paasche indices (Diewert, 1976; Reinsdorf, Diewert, Ehemann, 2002).

In the 1990s and 2000s there was a time of theoretical and practical deduction in the relationship of the CPI and COLI. There were many researches generalizing the experience of the COLI conceptions and describing the framework of two indices. Triplett (2001) presents two points of view on the problem of the CPI and COLI: first, the CPI is the same as COLI, but the latter is a result of economics, that is a theoretical notion; second, the CPI is a part of the COLI, its difference is the number of goods and services, which are included in the index calculation. More significant development was recorded in the practical assessment of the CPI. At that time the theoretical results in the COLI conception were applied in the CPI methodology change. In the history of the CPI politics there were the moments, when the special commission was organized for the deduction of theory achievements of the CPI and its estimation of practical using: the Stigler Commission report in 1961 (Rippy, 2014) and the Boskin Commission report in 1996 (Boskin et al., 1998). If the results of the first-mentioned report came to the improvement of samples of households (for weights) and goods. The second-mentioned report gave the information about the need of the additional CPI construction to estimate more accurately the inflation, namely it suggested the use of the Törnqvist or other superlative formula for the new complementary index number. The recommendations of these reports were used in the methodological development of the U.S. CPI. In addition to the change of sampling design the new CPIs were introduced in the statistical practice. Its difference concerns the other household sample for weighing, the other formula on the high level of aggregation (Abraham, 2003). Furthermore, there were the implicit transformations, namely the replacement of arithmetic mean by the geometric mean on the low level of price information aggregation (Dalton et al., 1998), the change of assessment process for several types of goods (Stewart et al., 1999). The above-mentioned introductions were later used by other countries.

The other way of CPI development concerns the use of digital technologies for data collecting. From the 2000s to the present there are research of scanner data using for CPI compiling (Leclaire et al., 2019; Richardson, 2003).

2 FROM INTERNATIONAL STANDARD TO NATIONAL ADAPTATION

The international document about the CPI is *Consumer price index manual: concepts and methods* (2020) which is based on *Consumer price index manual: Theory and practice* published in 2004. It doesn't reflect the exact assessment process of the CPI by the national statistical offices. It is only the way which needs to create and develop the national methodology for the CPI. National statistical offices themselves decide how to generate the methodology of CPI compiling and to sample the cities, the goods and the retail shops. The choosing of a strategy relates to the aim which must be achieved by means of the CPI.

In the text of *Consumer price index manual: concepts and methods* there are the most common motivation of CPI compiling: "the indexation of wages, rents, contracts, and social security payments; the deflation of household final consumption expenditure in the national accounts; and the use as a general macroeconomic indicator, especially for inflation targeting and for setting interest rates" (*Consumer price index manual: concepts and method…s*, 2020, p. 1). It is evident that this purpose list couldn't include all possible CPI aims because of economic system variety of countries. So the national statistical offices add the new indicators to the consumer price statistics, change the consumer coverage (all population or its part).

Russia had no time to develop evolutionarily and profoundly the CPI compiling process in compliance with the standards generated by the International Labor Office (ILO), International Monetary Fund (IMF) and other organizations together during the 20th century. In the USSR despite of the active development of price statistics theory in the 1920s, the consumer price statistics did not develop. Since the beginning of the 1990s it has needed to construct the CPI and other basic macroeconomic indicators as measures of comparability of economies.

At the end of 1991 the start of the CPI methodology, data collection and assessment process was announced in Russia. The CPI methodology must have been formed during one month since the publishing of law concerning the indexation. It is a very interesting fact that in science literature, especially in the articles of the economic and statistical journal, there is not a single sample of research about CPI elaboration in Russia. There are the series of researches about the price and inflation situation in the first half of 1990s after the price liberalization. The fundamental idea of these investigations was a description of the large-scale growth rate, a comparison of Russian price statistics data with the indicators of other countries (Masi and Koen, 1996). The other significant research of the Russian CPI concerned a problem of assessment process, especially the CPI bias. This investigation was based on the data of the second half of 1990s, which had two parts: price information collected by the national statistical office Rosstat and the household expenditure accumulated by the group of researchers from the University of North Carolina. The result was that the CPI bias for 1994–2001 varies from 0.64 to 0.87 percentage points as cumulative indicator for this time (Gibson et al., 2008).

The main stage of the CPI compiling process is considered as a practical assessment in Russia. The first stage of the CPI compiling is a sampling, which is the most interesting because it supposes the consideration of the national features as far as possible. The city sample where the consumer prices are collected was formed on the basis of some criteria. According to the Rosstat methodology (*Official statistical methodology…*, 2014) it does not use the probability sampling techniques within each region, the use of present country division in the first stage of city sampling is a typical example of stratification mentioned as "a common sampling technique used in the CPI" (*Consumer price index manual: concepts and methods…*, p. 68).

The significant features for city inclusion are the geographic position within the region in the present administrative and territorial division, the size of consumer market, which correlates with the number of people in the city, the number of cities in the region. The most common image of standard region from the position of city sample is that (1) there are from 2 to 4 of cities, (2) one of these is a region capital city, (3) sum of population share in these cities is not less 35% and (4) cities should be located in different parts of the region but an exception concerns the city which distinguishes by the special consumer market. The sample numbers are 282 cities and urban villages. The change of its list is not significant: during the last ten years 10 cities were added and 3 cities were excluded in 2019, in other years there were single cases of replacement or exception.

The formation of the *outlet sample* is realized on the basis of the sampling techniques combination. On the first stage there is the stratification of outlets in four populations which include all organizations operating in the city consumer market. The first population consists in the outlet selling the foodstuff, the second is the nonfood outlet population, there are the service outlets in the third population and the last outlet population consists in the small organizations which sell good, foodstuff or nonfood, or/and services and do not include in the first three population. On the second stage of sampling the ranking of outlets in each population is realized by the gross sales variable. The number of outlets depends on the number of populations: a lower rate of organizations relates to bigger size of population. The rate of outlet sample is 30%, if the population is less 300 organizations. For the population from 300 to 1 000 item the rate reduces by 10% and more by 10% when the sampling is realized on the basis of more than 1 000 organizations.

On the third stage the organizations are selected in the outlet sample in according to the above-mentioned rate and at regular interval in the ranking of outlets. After this simple random sampling the organizations are evaluated by the criteria of spatial diversity within the city, regular sales of goods and focus on the mass consumers.

The i*tem sample* is formed centrally by Rosstat for all regions, but on the regional level and for the regional CPI compiling the other special goods may be included. The bundle of goods and services in the CPI basket is revised continuously because of the changing of consumer pattern and emerging of new goods. The main formation of the item sample was in the 1990s. Since the beginning of XXI century the basket numbers have included more than 500 items. In 2005 the Rosstat methodology embraced 511 items but 48 items were not used for CPI compiling, only were collected. Today the sample numbers about 520 items, the sample size may differ insignificantly from year to year.

Despite the non-significant change in the item quantity there were many substitutions. In foodstuffs there were the integration of other quality goods in one item (for boiled sausage and vodka), the inclusion of many fresh vegetables and fruits, coffee beans, turkey meat and some additional seafood items. In the nonfood item sample the new goods were added, there are such as drugs, building materials, household appliances, mobile phone, jeans' clothes. The many types of textile and perfumery were excluded or united. The services are replaced according to the housing legislation: the addition of the household refuse services and the payment of capital repairs. Many types of financial services were included as well. The sewing services and house appliance repair were excluded because these services ceased to be important in the consumer pattern.

The next steps of the CPI compiling process are less dependent on the national features. The choice of many goods sold in the sampled outlets is realized on the base of the constant presence in outlet, popularity among consumers. These goods correspond to the item from sample. It numbers 5 or more to ensure the CPI time series continuity. The quantity of item price in each city depends on the three factors – the price dispersion, the share in the consumption and its quality.

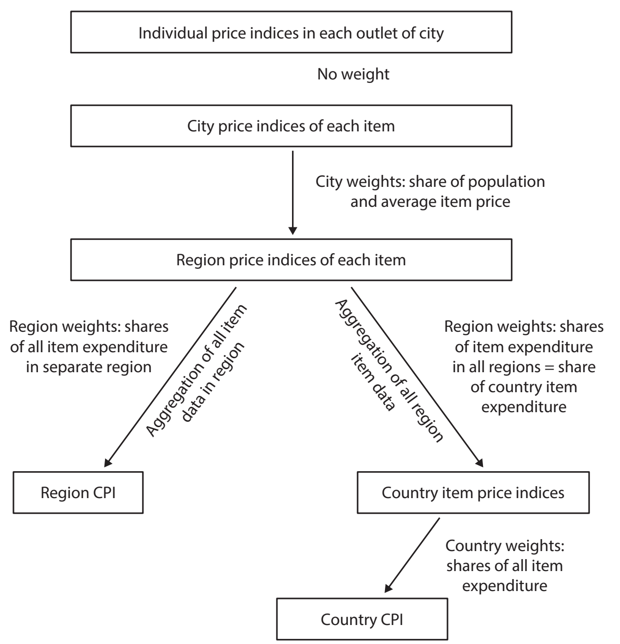
Before the CPI calculation the weights for the price indices are formed. The weights are the indicators of two types for other levels of price aggregation. There are the levels of city, region, and country. Because of the use of three levels in the Russian CPI compiling it is not enough only the item consumer expenditure or its share to construct the intermediate of the aggregates. The regional CPI and intermediate aggregates do not exist in the CPI system of all countries, so the weights of low-level aggregations are formed by the national statistical offices without strong requirements and based on the local features. The Russian CPI compiling is realized in consecutive order from city to regional level, from regional to country level.

The calculation of regional level index numbers is fulfilled by the weight of two complex indicators. There are the shares of population in the cities used for the price data collection and the average level of item price in these cities. In the *Consumer price index manual: concepts and methods* the judgement about population is following: "population statistics are sometimes used to split household expenditure across regions; however, this approach is not preferred as it assumes that expenditures per capita or per household are the same in all regions" (2020, p. 56, para. 3.28). The Russian methodology of the CPI includes the shares of population as an index price weights on the regional level. It assumes that each city in the sample has a consumer market of the size corresponding to the number of people. The share of population is one of two weights used simultaneously. The second weight in price index is the average price in the previous period. Furthermore in 2017 there was a renovation of city weight methodology. The aggregate shares of population in cities included in the sample were not always 100%. The new methodology assumes the recalculation of the initial shares by the addition of other regional localities to sampled cities. This addition is realized on the basis of two factors: first, the proximity of the minimum subsistence level which is calculated for each locality in terms of the age structure of population; second, there is the geographical proximity of localities.

The basic weights for the high level of aggregation are the shares of item expenditure obtained by the household budget survey which is organized quarterly. The weights are calculated on the basis of the average item expenditure of two years which are previous with a lag of one years to compiling period. This corresponds to the common trends of weights for price indices. Additionally, the adjustment of expenditure structure concerns the data replacement of the share of alcoholic beverages. The expenditure of alcoholic beverages is changed by the data of retail trade statistics because the respondents of the household budget survey understate the consumption of these goods.

The brief description of the aggregation process is shown in Figure 1. The main simplification concerns the representation of one index number for the bundle of goods and services because it does not need to description.

Figure 1 Process of Russian CPI compiling with the used weights



Source: Own construction based on the Rosstat methodology

The CPI on the high-level aggregation is calculated as a Laspeyres-type index number that corresponds to the practice of the national statistical offices in other countries. The Russian CPI has two types of assessment. There are the monthly and weekly price indices. The compiling process and weights are identical. Its difference consists in the item sample size: for the weekly CPI 83 items are used, as a result, the multiplication of four weekly CPI is not equal to the monthly CPI. The weekly CPI was very important indicator in the 1990s, today it is the interim assessment of inflation. The monthly CPI remains the basic inflation indicator in the Russian national economy and one of consumer price statistics indicators.

3 SYSTEM OF CONSUMER PRICE STATISTICS INDICATORS

The Russian consumer price statistics numbers 6 indicators, half of which are index numbers, 2 indicators are the absolute values and one of this is an average price. This system is formed thanks to the common basis of collection data principles and methods of the low-level indicator calculation.

The main element of system is the CPI which is compiled on the ground of the consumption pattern of all people. This is without the division by type of profession or the place of living. The U.S. Bureau of Labor Statistics compiles the CPI for all urban consumers and for urban wage earners and clerical workers, for example. There are two (weekly and monthly) assessments of the Russian CPI, about this fact is described in the previous chapter of this article.

The monthly CPI traditionally has a disadvantage for the forecasting and other related aims. This is a seasonal variation. Russian statistical practice has an experience of seasonality elimination. Rosstat does not apply the seasonal adjustments for the monthly CPI, neither does it correct this published price index. Instead of correction it compiles the other consumer price index named as a fundamental (or basic) CPI (F-CPI) which is *per se* the part of CPI. F-CPI is compiled on the basis of the same samples and assessment process that is for CPI but there is one difference. The item sample numbers less goods and services. Due to the F-CPI Rosstat proposes the additive decomposition of the consumer price changes. It publishes monthly the three indicators: CPI, F-CPI and residual CPI which is compiled on the basis of the item non-included in F-CPI. There indicators and its relations may represent in the following way:

$$CPI = \frac{\sum_{j=1}^{n} i_{pj} d_j}{\sum_{j=1}^{n} d_j}, \tag{1}$$

$$F - CPI = \frac{\sum_{j=1}^{k} i_{pj} d_j}{\sum_{j=1}^{k} d_j}, \tag{2}$$

$$residual\ CPI = \frac{\sum_{j=k}^{n} i_{pj} d_j}{\sum_{j=k}^{n} d_j}, \tag{3}$$

where j is item of sample, all item number is n, the items from 1 to k are used for the F-CPI compiling, the rest of items from k to n need to assess the residual CPI. Also i_{pj} is the consumer price index for j-item, d_j is the corresponding weight:

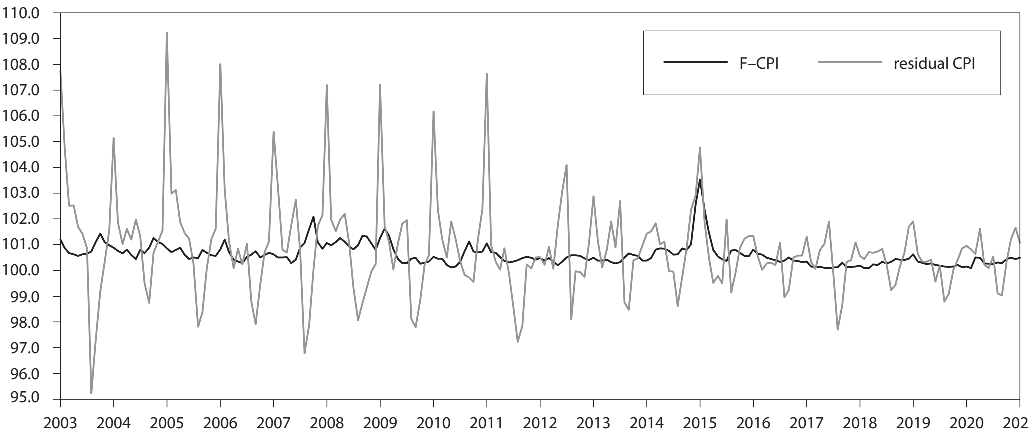
$$\sum_{j=1}^{n} d_j = \sum_{j=1}^{k} d_j + \sum_{j=k}^{n} d_j = 100\%. \tag{4}$$

As a result, the additive decomposition of the CPI change is realized by the F-CPI and residual CPI:

$$\Delta_p = \Delta_{non-seasonal}^{F-CPI} \times \sum_{j=1}^{k} d_j + \Delta_{seasonal}^{residual\ CPI} \times \sum_{j=k}^{n} d_j\,. \tag{5}$$

The ratio of non-seasonal and seasonal changes is correlated with the summary share of expenditure. The residual CPI includes the items which are seasonal-determined goods and services, also the items for which non-market pricing is typical (sometimes with the seasonal variation). Due to the exclusion of two types of items from F-CPI the seasonal variation is minimized, that it is apparently in Figure 2.

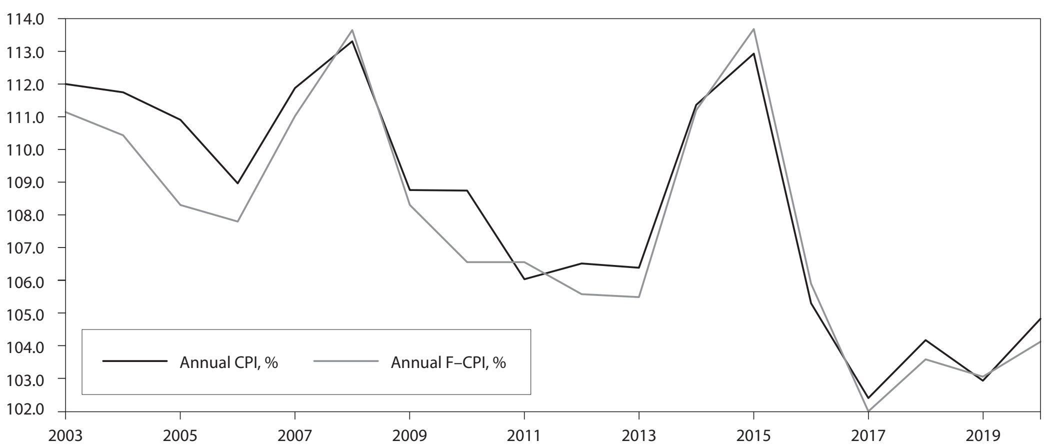
Figure 2 Russian monthly F-CPI and residual CPI in 2003–2021



Source: Own construction based on the Rosstat monthly data

The annual CPI and F-CPI have similar values but the F-CPI is not significant to the prediction aim and macroeconomic regulation. The CPI and F-CPI trends are shown in the Figure 3 where the annual data is used. Thereby the F-CPI is actual in the monthly or quarterly data analysis, in other cases it is less representative than the CPI because of the size of sample for the F-CPI compiling.

Figure 3 Russian annual CPI and F-CPI in 2003–2021



Source: Own construction based on the Rosstat annual data

The use of the monthly F-CPI is more appropriate in the macroeconomic models because of the low variance and standard error. The summary share of all excluded items from F-CPI is about 20–30%

by the expenditure indicator and about 20–22% by the number items.

Besides the seasonality the CPI as a one of the basic macroeconomic indicators attracts the society attention by its representativeness. The samples and sampling for the CPI compiling are considered above. This representative feature of the CPI relates to the quality of indicator on the base of variation. The CPI is an arithmetic mean on the high level of aggregation, as a result the test of its representativeness may be realized by the dispersion and standard deviation. For Russia this testing is very important because of the dramatic price change and its spatial and item heterogeneity. According to the Russian CPI compiling method (Figure 1) the testing must be concerned with the item price variation; it may use the regional price variation (spatial dispersion of prices) to estimate of CPI representativeness.

The standard deviation of item price indices is evaluated by the weight method. The weights are consumer expenditure shares which are used in the CPI compiling process. In Table 1 there are the monthly and annual indicators for 2020.

Table 1 Russian CPI in 2020 and its standard deviation by item price indices

Month	CPI in 2020, % to the previous month	Maximum item price index, %	Minimum item price index, %	Standard deviation, percentage point
January	100.40	123.29	89.56	2.16
February	100.33	119.49	94.14	1.71
March	100.54	114.91	83.64	1.99
April	100.83	250.95	79.35	5.44
May	100.27	118.22	66.44	2.21
June	100.22	121.11	70.55	2.97
July	100.35	110.49	83.20	2.14
August	99.96	114.09	80.39	2.54
September	99.93	112.34	83.77	2.27
October	100.43	118.29	90.35	1.76
November	100.71	139.20	94.66	3.15
December	100.87	147.47	95.60	3.71
Year	104.91	168.64	89.83	6.18

Source: Own construction based on Rosstat data, standard deviation is compiled

Besides the significant difference of item price indices, the monthly and annual CPIs have a low standard deviation which characterizes CPI as a representative indicator. For the estimation of the spatial dispersion it uses the regional CPI and number of people which live in region. The last indicator is the weight for the deviation estimation. This is more relevant weight that corresponds to the size of region, size of consumer market and consumer expenditure which are used to the regional CPI compiling.

In Table 2 the standard deviation by the regional CPI differentiation is not significant as for item price index deviation. It is much lower than the standard deviation for the item price index differentiation. It should be noted that it is determined by the lower difference between the regional CPIs. Furthermore for two types of standard deviation there is a cumulative effect for the annual value of standard deviation. It is higher than the monthly average.

The other relative indicator in the consumer price statistics is so-called cost-of-living index number. This price index has nothing in common with the COLI from the CPI research in the 20th century. In Russian statistical practice COLI (*Methodological recommendations…, 2012*) is a spatial index number.

Table 2 Regional CPIs in 2020 and country CPI standard deviation			
Month	**Maximum regional CPI, %**	**Minimum regional CPI, %**	**Standard deviation, percentage point**
January	101.86	99.50	0,.3
February	101.81	99.79	0.16
March	101.75	99.90	0.29
April	101.62	100.20	0.28
May	100.75	98.83	0.17
June	100.83	99.41	0.21
July	100.34	99.51	0.22
August	100.43	99.51	0.21
September	100.73	99.41	0.20
October	101.00	99.90	0.20
November	102.00	100.04	0.27
December	101.69	100.06	0.25
Year	107.28	101.91	0.80

Source: Own construction based on Rosstat data, standard deviation is compiled

Its methodology is not special. Rosstat has used the COLI compiling methods used by the American Chamber of Commerce Researchers Association (ACCRA) and the Council for Community and Economic Research (C2ER).

The Russian COLI is the newest indicator in the consumer price statistics. The CPI and COLI form the time-space coordinate system due to the common statistical data. In the case of U.S. CPI and COLI this condition is not fulfilled: price indices are compiled by other organizations with different price sample and the same weights, the frequency of price sampling and index publication is different also.

For the Russian COLI there are two features which do not correspond completely with the CPI. Firstly, this is the number of sampled items included in the COLI. It is less than for the CPI and slightly more than half of CPI item sample. Secondly, the COLI is compiled only for cities used for the consumer price collecting. On the one hand, it is relevant to the price difference in the other cities. On the other hand, the COLI and CPI do not correlate directly because COLI is compiled for cities, CPI is published for regions and country. The Russian COLI interpretation concerns the identification of "expensive" and "cheap" cities in relation to the country average level of consumer prices.

The absolute indicators of consumer price statistics are the value of constant set of goods and services and value of foodstuff minimum set. Both sets play an independent role and add to price index. The constant set of goods and services is more significant and more used. It includes 83 items from the sample. It is compiled by the multiplication of recent consumer prices and the constant quantities which do not change since the beginning of the 2000s. As a result the value of constant set of goods and services has not the bias related to the item sample structure. The small sample for constant set of goods and services provides the indicator value perspicuous for statistical data consumers and clear for interpretation.

As an absolute indicator the constant set of goods and services defines only the consumer price level in a country and regions where it may have a significant difference. In Figure 4 and Figure 5 there is an illustration of other patterns concerning the absolute price level and price change by CPI.

For the absolute price level there is one feature. The north regions and remote regions with underdeveloped transport infrastructure have a high price level. The regions with the high price level are

Figure 4 Regional differentiation of value of constant set of goods and services in 2020, rubles

Source: Own construction based on Rosstat data

Figure 5 Regional differentiation of CPI in 2020 (in %)

Source: Own construction based on Rosstat data

not the regions with the high change price (Figure 4 and Figure 5). On the contrary, the north territories show a low-price change. The consumer prices by the CPI increase faster in European and south regions.

The value of constant set of goods and services has the advantage in comparison with the published consumer price for each item from sample. It is one number for the item set that allows to analyze the common consumer price level in country or region. The average consumer prices differ and one of them does not characterize the situation on the consumer market.

The last indicator of consumer price statistics which is described in this article is the value of the foodstuff minimum set. The foodstuff minimum set includes 33 foods. Their quantities are enough to support life in the biological sense. The structure of this set and constant quantities of included items are formed in 1999 and remain the same at this moment. The significance of the foodstuff minimum set for the macroeconomic researchers and analysts of consumer market is not evident. The use of this set in the econometric model is not suitable because it does not include the consumer goods and foodstuff to form a representative indicator. The food quantities differ from similar indicators in the constant set of goods and services. As a result, the comparison of sets does not make any sense. The one special significance of the value of foodstuff minimum set is that it is a part of the value of subsistence minimum. On the base of it there is formation of the main minimum level of social payments. Until 2021 the value of subsistence minimum and the value of foodstuff minimum set as its part had been used as the base of the minimum wage setting. Today the last indicator is defined by the average and median income. So the value of foodstuff minimum set has not the former significance and may disappear in the consumer price statistics.

CONCLUSION

Formation of the consumer price statistics in Russia intends for its systemic nature and the multitasking. There are the balances of absolute (in rubles) and relative (in persent) indicators, time and spatial indicators. The first thirty years of Russian consumer price statistics was a time of finding an indicator set for the inflation assessment, inter-city and interregional differences in the consumer price level. Russian CPI has a weekly estimation, today it is not so important as in the 1990s, but there is a need for the short-run forecast. In research it was shown that the monthly CPI is a representative index number in terms of item change price and regional change price besides of dramatic dynamics in some periods. The representative characteristics reached are the sample design which is shown as well.

The CPI corresponds to the spatial price index COLI which supports the idea about time-space system of index numbers. This is a very important way of price statistics development, especially for a country with strong interregional differentiation. For this reason, the absolute indicators for price level estimation should be used. The value of constant set of goods and services intends to perform this task. One region may have the high rate of growth with absolute low-price level and vice-versa. As a result the time and space price indices represent the adequate statistical image in connection with the absolute value of goods bundle. This aim cannot be reached by the average price for each item.

The system of consumer price statistics in Russia is changing in accordance to the conditions in national economy and society. Furthermore, the CPI as a central element of this system is changing as well. These transformations concern mainly the item and outlet sample to maintain the representativeness of the CPI. In addition, the new technology of data collecting allows to use more adequate information about prices and quantities simultaneously. In 2016 the new technology of retail prices in the outlets was installed in Russia, so this possibility of price data collecting is not only for the taxation aim, for the CPI compiling as well. The Director of Rosstat announces that the scanner data is one additional way of data collecting and the prices from scanner data will be used for in the next year. This will only increase the price sample in the cities. The replacement of the data collected traditionally by the scanner data is not planned to be realized in the short run.

References

ABRAHAM, K. G. Toward a cost-of-living index: progress and prospects. *The Journal of Economic Perspectives*, 2003, 17(1), pp. 45–58.

BALK, B. M. On calculating cost-of-living index numbers for arbitrary income level. *Econometrica*, 1990, 58(1), pp. 75–92.

BEATTY, N. K. AND LARSEN, E. R. Using Engel curves to estimate bias in the Canadian CPI as a cost of living index. *The Canadian Journal of Economic*s, 2005, 38(2), pp. 482–499.

BLACKORBY, C. AND RUSSEL, R. R. Indices and subindices of the cost of living and the standards of living. *International Economic Review*, 1978, 19(1), pp. 229–240.

BOSKIN, M. J., DULBERGER, E. R., GORDON, R. J., GRILICHES, Z., JORGENSON, D. W. Consumer prices, the consumer price index, and the cost of living. *The Journal of Economic Perspectives*, 1998, 12(1), pp. 3–26.

BRAITHWAIT, S. D. The substitution bias of the Laspeyres price index: an analysis using estimated cost-of-living indexes. *The American Economic Review*, 1980, 70(1), pp. 64–77.

CHETTY, V. K. On the construction of cost of living and productivity indices. *International Economic Review*, 1971, 12(1), pp. 144–146.

Consumer price index manual: concepts and methods. Washington, D.C.: International Monetary Fund, 2020.

Consumer price index manual: Theory and practice. Geneva: International Labor Office, 2004.

DALTON, K. V., GREENLEES, J. S., STEWART, K. J. Incorporating a geometric mean formula into the CPI. *Monthly Labor Review*, 1998, October, pp. 3–7.

DIEWERT, W. E. Exact and superlative index numbers. *Journal of Econometrics*, 1976, 4, pp. 115–145.

FRY, V. AND PASHRDES, R. Construction the true cost of living index from the Engel curves of the Piglog model. *Journal of Applied Econometrics*, 1989, 4(1), pp. 41–56.

GALATIN, M. A true price index when the consumer saves. *The American Economic Review*, 1973, 63(1), pp. 185–194.

GIBSON, J., STILLMAN, S., LE, T. CPI bias and real living of standards in Russia during the transition. *Journal of Development Economics*, 2008, 87, pp. 140–160.

KLEIN, L. R. AND RUBIN, H. A constant-utility index of the cost of living. *The Review of Economic Studies*, 1947–1948, 15(2), pp. 84–87.

KONÜS, A. A. The problem of the true index of the cost of living. *Econometrica*, 1939, 7(1), pp. 10–29.

LECLAIRE, M., LÉONARD, I., RATEAU, G., SILLARD, P., GAËTAN, V., VERNÉDAL, P. Scanner data: Advances in methodology and new challenges for computing consumer price indices. *Economics and Statistics*, 2019, 509, pp. 13–29.

MASI, P. AND KOEN, V. Relative price convergence in Russia. *Staff Papers* (International Monetary Fund), 1996, 43(1), pp. 97–122.

Methodological recommendations of cost of living index compiling in selected cities (in Russian). Approved by Rosstat, Moscow, 8 August 2012.

Official statistical methodology of statistical observation of the consumer prices of goods and services and compiling of consumer price indices (in Russian). Approved by Rosstat, Moscow, 30 December 2014.

POLLAK, R. Welfare evaluation and the cost-of-living index in the household production model. *The American Economic Review*, 1978, 68(3), pp. 285–299.

POLLAK, R. Group cost-of-living indexes. *The American Economic Review*, 1980, 70(2), pp. 273–278.

POLLAK, R. Subindexes in the cost of living index. *International Economic Review*, 1975, 16(1), pp. 135–150.

REINSDORF, M. B., DIEWERT, W. E., EHEMANN, C. Additive decompositions for Fisher, Törnqvist and geometric mean indexes. *Journal of Economic and Social Measurement*, 2002, 28, pp. 51–61.

RICHARDSON, D.V. Scanner indexes for the consumer price index. In: FEENSTRA, R. C. AND SHAPIRO, M. D., (eds.) *Scanner Data and Price Indexes.* University of Chicago Press, 2003.

RIPPY, D. The first hundred years of the Consumer Price Index: a methodological and political history [online]. *Monthly Labor Review*, 2014, April. [28.1.2021]. <https://www.bls.gov/opub/mlr/2014/article/the-first-hundred-years-of-the-consumer-price-index.htm>.

STEWART, K. J. AND REED, S. B. Consumer price index research series using current methods, 1978–98. *Monthly Labor Review*, 1999, June, pp. 29–38.

TRIPLETT, J. R. Should the cost-of living provide the conceptual framework for a consumer price index? *The Economic Journal*, 2001, 111(472), pp. F311–F334.

ULMER, M. J. On the economic theory of cost of living index numbers. *Journal of the American Statistical Association*, 1946, 41(236), pp. 530–542.

WALD, A. A new formula for the index of cost of living. *Econometrica*, 1939, 7(4), pp. 319–331.

Recursive Estimation of Volatility for High Frequency Financial Data

Petr Vejmělka[1] | *Charles University, Prague, Czech Republic*
Tomáš Cipra[2] | *Charles University, Prague, Czech Republic*

Abstract

The paper deals with recursive estimation of financial time series with conditional volatility. It surveys the recursive methodology suggested in Hendrych and Cipra (2018) and adjusts it for various alternatives of GARCH models which are usual in financial practice. Such a recursive approach seems to be suitable for the dynamic estimation with high-frequency data. The paper verifies the applicability of recursive algorithms of particular models to high-frequency data from the Czech environment, particularly in the context of risk prediction.

Keywords	***JEL code***
GARCH, high-frequency financial time series, recursive estimation, risk prediction, volatility	*C32, C51, C58*

INTRODUCTION

In the case of financial time series modeling, models with conditional heteroscedasticity GARCH are currently preferred in practice. They present the most powerful tool for routine modeling of financial time series. In practice, these models are commonly estimated using static (off-line or batch) methods, e.g., the maximum likelihood estimation. However, the application of the static methods to high-frequency data, such as stock market data, is problematic or even impossible. As an example, in the case of stock prices, where minute or even more frequent data are encountered, the use of static methods would be computationally impossible. For this reason, recursive methods are preferred for high-frequency data.

In literature, there have already been proposed recursive algorithms for GARCH model estimation, e.g. Kierkegaard et al. (2000), Aknouche and Guerbyenne (2006) or Hendrych and Cipra (2018, 2019). The recursive methodology suggested in Hendrych and Cipra (2018) can be adjusted for various types of GARCH models (see the recursive algorithms for models GJR-GARCH, IGARCH and EGARCH in Section 2). The aim of this paper is to verify the applicability of these recursive algorithms to real high-frequency data from the Czech environment. In particular, the risk prediction potential of this recursive methodology is investigated using specific methods (MAPE criterion, realized volatility).

[1] Department of Probability and Mathematical Statistics, Charles University, Sokolovská 83, Prague 8, Czech Republic. Corresponding author: e-mail: vejmelp@karlin.mff.cuni.cz.
[2] Department of Probability and Mathematical Statistics, Charles University, Sokolovská 83, Prague 8, Czech Republic.

The paper is organized as follows. Section 1 presents models with conditional heteroskedasticity that can be used for modeling in finance, namely the modifications of the GARCH model. At the same time, this section presents the algorithm for recursive estimation of volatility for the GARCH model. Section 2 focuses on the description of recursive formulas for modifications of the GARCH model, such as GJR-GARCH, IGARCH and EGARCH, and briefly comments results of a simulation study. In Section 3, the application of the proposed algorithms to real high-frequency data is presented including the risk prediction analysis. Finally, the last section summarizes conclusions.

1 GARCH MODELS AND RECURSIVE ESTIMATION OF VOLATILITY

When working with time series, there exist several ways how to model them. However, when dealing with financial time series, the usual data generating mechanism depends on the first and second conditional moments, see Cipra (2020). Thus, these time series are assumed in the following form:

$$y_t = \mu_t + e_t = \mu_t + \sigma_t \varepsilon_t, \tag{1}$$

where μ_t represents the conditional mean, σ_t is the square root of the conditional variance and ε_t's are independent, identically distributed random variables with zero mean and unit variance. Our primary aim is to model the conditional variance and, in particular, to find recursive algorithms for its estimation in time.

1.1 GARCH models in financial practice

In literature, many different approaches to modeling the conditional variance have been considered so far. However, the strongest tool for financial time series modeling, which has not yet been overcome, are GARCH models.

1.1.1 GARCH model

The most important model from this class of models is the GARCH model proposed by Bollerslev (1986), where the equation for the conditional variance has the following form:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i e_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2, \tag{2}$$

where $\alpha_0 > 0$, $\alpha_i \geq 0$ for $i > 0$, $\beta_j \geq 0$ for $j > 0$. These are the conditions to ensure positivity of the conditional variance. The stationarity is provided by fulfillment of an additional condition $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1$. By adding the lagged values of the conditional variance into the equation, the model can more successfully capture volatility clustering, which is typical for financial time series.

Although the GARCH model is undoubtedly the most popular of the models with conditional heteroscedasticity, we encounter many modifications of this basic model in financial practice. These modifications aim to eliminate some of the drawbacks of the GARCH model and improve its properties so that it is as close as possible to the real behavior of the data (see below).

1.1.2 IGARCH model

One of the simplest extension is the integrated GARCH model with orders p, q, usually denoted as the IGARCH(p, q) model (see Engle and Bollerslev, 1986). The only difference consists in a stricter parameter constraint:

$$\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j = 1. \tag{3}$$

Formula (3) causes the non-existence of the unconditional variance. The impact of current information persists in conditional volatility forecasts for long horizons. For instance, the popular EWMA model is a special case of the IGARCH(1,1) model.

1.1.3 GJR-GARCH model

In order to capture the leverage effect, another modification was proposed by Glosten, Jagannathan and Runkle (1993). The volatility equation of GJR-GARCH(p, q) model has the following form:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} (\alpha_i + \gamma_i I_{t-i}^{-}) e_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2, \tag{4}$$

where I_{t-i}^{-} denotes an indicator, which is equal to 1 if $e_{t-i} < 0$ and equal to 0 otherwise.

The sufficient conditions for σ_t^2 being positive are $\alpha_0 > 0$, $\alpha_i \geq 0$ and $\alpha_i + \gamma_i \geq 0$ for $i > 0$ and $\beta_j \geq 0$ for $j > 0$. There is no general set of conditions to ensure that the time series is stationary. The new parameter γ_i, which regulates the different effect of e_{t-i} according to its sign. If e_{t-i} is negative, the impact is higher and the leverage effect is present.

1.1.4 EGARCH model

Another model including the leverage effect is the exponential GARCH model (EGARCH(p, q)) proposed by Nelson (1991). We apply it in the form:

$$\ln(\sigma_t^2) = \alpha_0 + \sum_{i=1}^{p} \alpha_i \ln(\sigma_{t-i}^2) + \sum_{j=0}^{q} \delta_j \frac{y_{t-1-j}}{\sqrt{\sigma_{t-1-j}^2}} + \sum_{k=0}^{q} \gamma_k \left(\frac{|y_{t-1-k}|}{\sqrt{\sigma_{t-1-k}^2}} - \sqrt{\frac{2}{\pi}} \right), \tag{5}$$

where α_0, α_i, δ_j and γ_k are parameters. Due to the logarithmic transformations in (5), the positivity of volatility is fulfilled without any conditions imposed on the parameters.

1.2 State-space representation of GARCH models

For some of these models, it is possible to use algorithms implemented in various software systems such as EViews or R. However, this approach cannot be applied, for example, to high-frequency data such as stock market prices or index levels since the volume of such data may be enormous and real-time parameter estimation is not possible. For this reason, a recursive approach is more appropriate.

Several articles already dealt with a derivation of recursive algorithms for GARCH models, e.g., Kierkegaard et al. (2000), Aknouche and Guerbyenne (2006), Gerencsér et al. (2010), and Hendrych and Cipra (2018). These articles primarily focused on the GARCH model. In this paper, we will follow the procedure proposed by Hendrych and Cipra (2018), which is based on the general recursive algorithms, see also Ljung and Söderström (1983) or Ljung (1999). The procedures of this type are called the recursive pseudo-linear regression or the prediction error method.

In order to obtain a recursive algorithm, which could be used to estimate parameters in the basic GARCH model (2), it is necessary to transform the volatility equation into a vector form. Furthermore, the conditional mean will be considered equal to zero for simplicity. The modified form of the conditional volatility Formula (2) is:

$$\boldsymbol{\varphi}_t^T(\boldsymbol{\theta})\, \boldsymbol{\theta} = \alpha_0 + \sum_{i=1}^{p} \alpha_i y_{t-i}^2 + \sum_{j=1}^{q} \beta_j \boldsymbol{\varphi}_{t-j}^T(\boldsymbol{\theta})\, \boldsymbol{\theta}, \tag{6}$$

where $\boldsymbol{\theta} = (\alpha_0, \alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q)^T$ is the vector of model parameters, and $\boldsymbol{\varphi}_t(\boldsymbol{\theta}) = (1, y_{t-1}^2, \dots, y_{t-p}^2, \boldsymbol{\varphi}_{t-1}^T(\boldsymbol{\theta})\, \boldsymbol{\theta}, \dots, \boldsymbol{\varphi}_{t-q}^T(\boldsymbol{\theta})\, \boldsymbol{\theta})^T$ is the vector constructed so that the volatility equation holds. Since the conditional mean is zero, the terms e_{t-i}^2 were replaced by y_{t-i}^2.

The most important target is to construct the estimates $\hat{\boldsymbol{\theta}}$ recursively in time. Hendrych and Cipra (2018) suggest a self-weighted approach based on the minimization of a loss function corresponding to the weighted log-likelihood approach. The final algorithm has the following form:

$$\hat{\boldsymbol{\theta}}_t = \hat{\boldsymbol{\theta}}_{t-1} + \frac{\hat{\mathbf{P}}_{t-1}\hat{\boldsymbol{\psi}}_t(y_t^2 - \hat{\boldsymbol{\varphi}}_t^T\hat{\boldsymbol{\theta}}_{t-1})}{\hat{\boldsymbol{\psi}}_t^T\hat{\mathbf{P}}_{t-1}\hat{\boldsymbol{\psi}}_t + \lambda_t(\hat{\boldsymbol{\varphi}}_t^T\hat{\boldsymbol{\theta}}_{t-1})^2}, \tag{7}$$

$$\hat{\mathbf{P}}_t = \frac{1}{\lambda_t}\left[\hat{\mathbf{P}}_{t-1} - \frac{\hat{\mathbf{P}}_{t-1}\hat{\boldsymbol{\psi}}_t\hat{\boldsymbol{\psi}}_t^T\hat{\mathbf{P}}_{t-1}}{\hat{\boldsymbol{\psi}}_t^T\hat{\mathbf{P}}_{t-1}\hat{\boldsymbol{\psi}}_t + \lambda_t(\hat{\boldsymbol{\varphi}}_t^T\hat{\boldsymbol{\theta}}_{t-1})^2}\right], \tag{8}$$

$$\hat{\boldsymbol{\varphi}}_{t+1} = (1, y_t^2, \dots, y_{t+1-p}^2, \hat{\boldsymbol{\varphi}}_t^T\hat{\boldsymbol{\theta}}_t, \dots, \hat{\boldsymbol{\varphi}}_{t+1-q}\hat{\boldsymbol{\theta}}_{t+1-q})^T, \tag{9}$$

$$\hat{\boldsymbol{\psi}}_{t+1} = \hat{\boldsymbol{\varphi}}_{t+1} + \sum_{j=1}^{q}\hat{\beta}_{j,t}\hat{\boldsymbol{\psi}}_{t+1-j} \tag{10}$$

for $t \in \mathbb{N}$, where $\hat{\mathbf{P}}_t$ is a $(1 + p + q) \times (1 + p + q)$ square matrix.

Several issues need to be addressed. The first of them is the choice of weights $\{\lambda_t\}$. One of the possible options is the application of a recursive formula:

$$\lambda_t = \tilde{\lambda}\lambda_{t-1} + (1 - \tilde{\lambda}),\ t \in \mathbb{N}, \tag{11}$$

as suggested in Ljung and Söderström (1983). A recommended choice of constants $\tilde{\lambda}$ and λ_0 is $\tilde{\lambda} = 0.99$ and $\lambda_0 = 0.95$. Another important choice is setting the initial estimates of the vector of parameters and some other quantities. Different options may be appropriate for each situation. One of the possibilities is to set $\hat{\boldsymbol{\theta}}_0 = (\frac{1}{n}\sum_{i=1}^{n} y_{1-i}^2[1 - (p + q)\eta], \eta, \dots, \eta)^T$, where η is a small positive constant satisfying $(p + q)\eta < 1$ for a suitable n, $\hat{\mathbf{P}}_0 = c\mathbf{I}$, where c is a suitable positive constant (e.g., c = 10^2 for this model), which ensures that the initial estimates are less influential and there is a faster convergence to the actual vector of parameters, $\hat{\boldsymbol{\varphi}}_1 = (1, y_{1-p}^2, \dots, y_0^2, k, \dots, k)^T$ with k equal to a small positive constant, and finally $\hat{\boldsymbol{\psi}}_1 = \hat{\boldsymbol{\varphi}}_1$ and $\hat{\boldsymbol{\psi}}_i = \mathbf{0}$ for $i = -q + 2, \dots, 0$. If the values $y_{1-p}^2, \dots, y_0^2$ are not known, one can assume them to be equal to zero.

In addition to these choices, one can extend the proposed algorithm with a mechanism how to ensure the positivity of the conditional variance and the stationarity. This is achieved by taking the estimate at time t, according to its obtained values, as $\hat{\boldsymbol{\theta}}_t$ if $\hat{\boldsymbol{\theta}}_t \in \mathbf{D}_S$ and as $\hat{\boldsymbol{\theta}}_{t-1}$ if $\hat{\boldsymbol{\theta}}_t \notin \mathbf{D}_S$, where $\mathbf{D}_S$ is the set of vectors $\boldsymbol{\theta}$ satisfying the conditions imposed on the parameters to ensure positivity and stationarity. In the case of the GARCH model, $\mathbf{D}_S = \{\boldsymbol{\theta} \in \mathbb{R}^{p+q+1} \mid \alpha_0 > 0, \alpha_1, \dots, \alpha_p \geq 0, \beta_1, \dots, \beta_q \geq 0, \sum_{i=1}^{p}\alpha_1 + \sum_{j=1}^{q}\beta_j < 1\}$. If the estimate lies outside this set, it is ignored and the previous estimate is considered instead.

2 RECURSIVE ESTIMATION OF SELECTED GARCH MODELS

In this section, recursive algorithms for estimating the parameters of various modifications of the GARCH model from Section 1 will be presented. For the GJR-GARCH model, just a simple modification of the basic algorithm is needed. In other cases, major changes are necessary.

2.1 Recursive estimation of GJR-GARCH model

In this case the vectors $\boldsymbol{\theta}$ and $\boldsymbol{\varphi}_t(\boldsymbol{\theta})$ are modified into:

$$\boldsymbol{\theta} = (\alpha_0, \alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q, \gamma_1, \dots, \gamma_p)^T, \tag{12}$$

$$\boldsymbol{\varphi}_t(\boldsymbol{\theta}) = (1, y_{t-1}^2, \dots, y_{t-p}^2, \boldsymbol{\varphi}_{t-1}^T(\boldsymbol{\theta})\,\boldsymbol{\theta}, \dots, \boldsymbol{\varphi}_{t-q}^T(\boldsymbol{\theta})\,\boldsymbol{\theta}, y_{t-1}^2 I_{t-1}^-, \dots, y_{t-p}^2 I_{t-p}^-)^T, \tag{13}$$

Similarly as in GARCH:

$$\psi_t(\boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}} [\boldsymbol{\varphi}_t^T(\boldsymbol{\theta})\, \boldsymbol{\theta}] = \boldsymbol{\varphi}_t(\boldsymbol{\theta}) + \sum_{j=1}^{q} \beta_j \psi_{t-1}(\boldsymbol{\theta}). \tag{14}$$

Hence the corresponding estimation algorithm coincides with the one for the GARCH model except for the equation for $\hat{\boldsymbol{\varphi}}_{t+1}$. Namely,

$$\hat{\boldsymbol{\theta}}_t = \hat{\boldsymbol{\theta}}_{t-1} + \frac{\hat{\mathbf{P}}_{t-1}\hat{\psi}_t(y_t^2 - \hat{\boldsymbol{\varphi}}_t^T\hat{\boldsymbol{\theta}}_{t-1})}{\hat{\psi}_t^T\hat{\mathbf{P}}_{t-1}\hat{\psi}_t + \lambda_t(\hat{\boldsymbol{\varphi}}_t^T\hat{\boldsymbol{\theta}}_{t-1})^2}, \tag{15}$$

$$\hat{\mathbf{P}}_t = \frac{1}{\lambda_t}\left[\hat{\mathbf{P}}_{t-1} - \frac{\hat{\mathbf{P}}_{t-1}\hat{\psi}_t\hat{\psi}_t^T\hat{\mathbf{P}}_{t-1}}{\hat{\psi}_t^T\hat{\mathbf{P}}_{t-1}\hat{\psi}_t + \lambda_t(\hat{\boldsymbol{\varphi}}_t^T\hat{\boldsymbol{\theta}}_{t-1})^2}\right], \tag{16}$$

$$\hat{\boldsymbol{\varphi}}_{t+1} = (1, y_t^2, \dots, y_{t+1-p}^2, \hat{\boldsymbol{\varphi}}_t^T\hat{\boldsymbol{\theta}}_t, \dots, \hat{\boldsymbol{\varphi}}_{t+1-q}^T\hat{\boldsymbol{\theta}}_{t+1-q}, y_t^2 I_t^-, \dots, y_{t+1-p}^2 I_{t+1-p}^-)^T, \tag{17}$$

$$\hat{\psi}_{t+1} = \hat{\boldsymbol{\varphi}}_{t+1} + \sum_{j=1}^{q} \hat{\beta}_{j,t}\hat{\psi}_{t+1-j} \tag{18}$$

for $t \in \mathbb{N}$.

Also the initial estimates may be constructed in the similar way as in the case of the GARCH model. One can take $\hat{\boldsymbol{\theta}}_0 = (\frac{1}{n}\sum_{i=1}^{n} y_{1-i}^2[1-(p+q)\eta], \eta, \dots, \eta, 0, \dots, 0)^T$, where η is a small positive constant satisfying $(p+q)\eta < 1$ for a suitable n, $\hat{\mathbf{P}}_0 = c\mathbf{I}$, where c is a suitable positive constant, $\hat{\boldsymbol{\varphi}}_1 = (1, y_{1-p}^2, \dots, y_0^2, k, \dots, k, 0, \dots, 0)^T$, with k equal to a small positive constant, $\hat{\psi}_1 = \hat{\boldsymbol{\varphi}}_1$ and $\hat{\psi}_i = \mathbf{0}$ for $i = -q+2, \dots, 0$.

2.2 Recursive estimation of IGARCH model

As stated above, this model differs from the GARCH model by the condition $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j = 1$. In order to include this condition directly into the volatility equation, one can rewrite it to:

$$\sigma_t^2 = y_{t-p}^2 + \alpha_0 + \alpha_1(y_{t-1}^2 - y_{t-p}^2) + \dots + \alpha_{p-1}(y_{t-p+1}^2 - y_{t-p}^2) + \beta_1(\sigma_{t-1}^2 - y_{t-p}^2) + \dots + \beta_q(\sigma_{t-q}^2 - y_{t-p}^2), \tag{19}$$

i.e., in the vector form:

$$\sigma_t^2(\boldsymbol{\theta}) = y_{t-p}^2 + \boldsymbol{\varphi}_t^T(\boldsymbol{\theta})\boldsymbol{\theta}, \tag{20}$$

where the vectors $\boldsymbol{\theta}$ and $\boldsymbol{\varphi}_t(\boldsymbol{\theta})$ are such that (20) holds. Since the expression for $\sigma_t^2(\boldsymbol{\theta})$ was changed, it is necessary to derive the recursive algorithm newly. The derivation runs in a similar way as for the GARCH model. The final recursive algorithm can be written as:

$$\hat{\boldsymbol{\theta}}_t = \hat{\boldsymbol{\theta}}_{t-1} + \frac{\hat{\mathbf{P}}_{t-1}\hat{\psi}_t(y_t^2 - y_{t-p}^2 - \hat{\boldsymbol{\varphi}}_t^T\hat{\boldsymbol{\theta}}_{t-1})}{\hat{\psi}_t^T\hat{\mathbf{P}}_{t-1}\hat{\psi}_t + \lambda_t(y_{t-p}^2 + \hat{\boldsymbol{\varphi}}_t^T\hat{\boldsymbol{\theta}}_{t-1})^2}, \tag{21}$$

$$\hat{\mathbf{P}}_t = \frac{1}{\lambda_t}\left[\hat{\mathbf{P}}_{t-1} - \frac{\hat{\mathbf{P}}_{t-1}\hat{\psi}_t\hat{\psi}_t^T\hat{\mathbf{P}}_{t-1}}{\hat{\psi}_t^T\hat{\mathbf{P}}_{t-1}\hat{\psi}_t + \lambda_t(y_{t-p}^2 + \hat{\boldsymbol{\varphi}}_t^T\hat{\boldsymbol{\theta}}_{t-1})^2}\right], \tag{22}$$

$$\hat{\boldsymbol{\varphi}}_{t+1} = (1, y_t^2 - y_{t-p+1}^2, \dots, y_{t+2-p}^2 - y_{t-p+1}^2, \hat{\boldsymbol{\varphi}}_t^T\hat{\boldsymbol{\theta}}_t + y_{t-p}^2 - y_{t-p+1}^2, \dots, \hat{\boldsymbol{\varphi}}_{t+1-q}^T\hat{\boldsymbol{\theta}}_{t+1-q} + y_{t+1-q-p}^2 - y_{t-p+1}^2)^T, \tag{23}$$

$$\hat{\psi}_{t+1} = \hat{\varphi}_{t+1} + \sum_{j=1}^{q} \hat{\beta}_{j,t} \hat{\psi}_{t+1-j} \tag{24}$$

for $t \in \mathbb{N}$.

Again the initial can be set analogously to the case of the GARCH model. That means $\hat{\theta}_0 = (\frac{1}{n}\sum_{i=1}^{n} y_{1-i}^2[1-(p-1+q)\eta], \eta, \dots, \eta, \eta, \dots, \eta)^T$, where η is a small positive constant satisfying $(p-1+q)\eta < 1$ for a suitable n, $\hat{\mathbf{P}}_0 = c\mathbf{I}$, where c is a suitable positive constant, $\hat{\varphi}_1 = (1, y_{2-p}^2 - y_{1-p}^2, \dots, y_0^2 - y_{1-p}^2, k, \dots, k)^T$, where k equals to a small positive constant, $\hat{\psi}_1 = \hat{\varphi}_1$ and $\hat{\psi}_i = \mathbf{0}$ for $i = -q+2, \dots, 0$.

2.3 Recursive estimation of EGARCH model

In the previous section, the specific form of the EGARCH model suitable for recursive estimation was introduced. One should remind that the conditional variance is assumed in the logarithmic form (5). The corresponding vector notation looks as follows:

$$\sigma_t^2(\boldsymbol{\theta}) = \exp(\boldsymbol{\varphi}_t^T(\boldsymbol{\theta})\boldsymbol{\theta}) , \tag{25}$$

where

$$\boldsymbol{\theta} = (\alpha_0, \alpha_1, \dots, \alpha_p, \delta_0, \dots, \delta_q, \gamma_0, \dots, \gamma_q)^T, \tag{26}$$

and $\boldsymbol{\varphi}_t(\boldsymbol{\theta})$ is such that (25) holds. The derivation provides the corresponding recursive algorithm in the form:

$$\hat{\boldsymbol{\theta}}_t = \hat{\boldsymbol{\theta}}_{t-1} + \frac{\hat{\mathbf{P}}_{t-1}\hat{\psi}_t(y_t^2 - \exp(\hat{\varphi}_t^T\hat{\boldsymbol{\theta}}_{t-1}))}{(\hat{\psi}_t^T\hat{\mathbf{P}}_{t-1}\hat{\psi}_t + \lambda_t)\exp(\hat{\varphi}_t^T\hat{\boldsymbol{\theta}}_{t-1})} , \tag{27}$$

$$\hat{\mathbf{P}}_t = \frac{1}{\lambda_t}\left[\hat{\mathbf{P}}_{t-1} - \frac{\hat{\mathbf{P}}_{t-1}\hat{\psi}_t\hat{\psi}_t^T\hat{\mathbf{P}}_{t-1}}{\hat{\psi}_t^T\hat{\mathbf{P}}_{t-1}\hat{\psi}_t + \lambda_t}\right], \tag{28}$$

$$\begin{aligned}\hat{\varphi}_{t+1} = {} & \Bigg(1, \hat{\varphi}_t^T\hat{\boldsymbol{\theta}}_t, \dots, \hat{\varphi}_{t+1-p}^T\hat{\boldsymbol{\theta}}_{t+1-p}, \frac{y_t}{\sqrt{\exp(\hat{\varphi}_t^T\hat{\boldsymbol{\theta}}_t)}}, \dots, \frac{y_{t-q}}{\sqrt{\exp(\hat{\varphi}_{t-q}^T\hat{\boldsymbol{\theta}}_{t-q})}} , \\ & \frac{|y_t|}{\sqrt{\exp(\hat{\varphi}_t^T\hat{\boldsymbol{\theta}}_t)}} - \sqrt{2/\pi}, \dots, \frac{|y_{t-q}|}{\sqrt{\exp(\hat{\varphi}_{t-q}^T\hat{\boldsymbol{\theta}}_{t-q})}} - \sqrt{2/\pi}\Bigg)^T,\end{aligned} \tag{29}$$

$$\hat{\psi}_{t+1} = \hat{\varphi}_{t+1} + \sum_{i=1}^{p}\hat{\alpha}_{i,t}\hat{\psi}_{t+1-i} + \sum_{j=1}^{q}\frac{\hat{\delta}_{j,t}y_{t-j} + \hat{\gamma}_{j,t}|y_{t-j}|}{2\sqrt{\exp(\hat{\varphi}_{t-j}^T\hat{\boldsymbol{\theta}}_{t-j})}}\hat{\psi}_{t-j} \tag{30}$$

for $t \in \mathbb{N}$.

One can supplement the algorithm with the following initial estimates: $\hat{\boldsymbol{\theta}}_0 = (\frac{1}{n}\sum_{i=1}^{n} y_{1-i}^2[1-[p+2(q+1)]\eta], \eta, \dots, \eta, \eta, \dots, \eta, \eta, \dots, \eta)^T$, where η is a small positive constant satisfying $[p+2(q+1)]\eta < 1$ for a suitable n, $\hat{\mathbf{P}}_0 = c\mathbf{I}$, where c is a suitable constant, $\hat{\varphi}_1 = (1, k, \dots, k, \frac{y_0}{\sqrt{\exp(k)}},$

$\dots, \frac{y_{-q}}{\sqrt{\exp(k)}}, \frac{|y_0|}{\sqrt{\exp(k)}} - \sqrt{2/\pi}, \dots, \frac{|y_{-q}|}{\sqrt{\exp(k)}} - \sqrt{2/\pi})^T$ with k equal to a small positive constant, $\hat{\boldsymbol{\psi}}_1 = \hat{\boldsymbol{\varphi}}_1$ and $\hat{\boldsymbol{\psi}}_i = \mathbf{0}$ for $i = \min\{-p + 2, -q + 1\}, \dots, 0$.

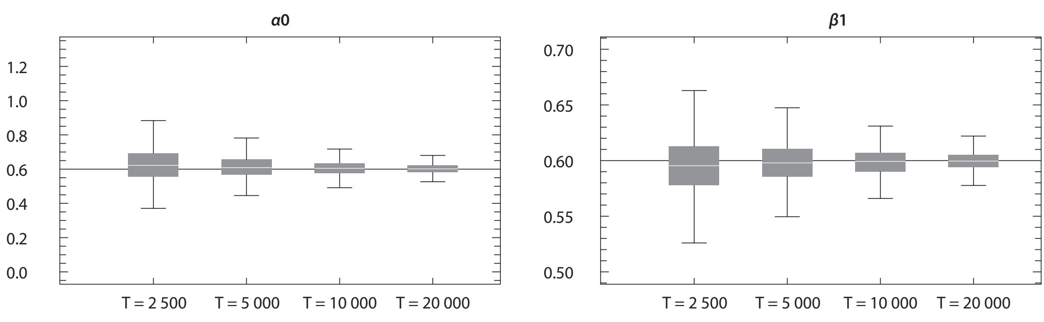
Remark: Other models were considered, e.g., the matrix extension of GARCH model respecting the interactions of model components. However, due to an extensive number of parameters, the numerical outputs (mainly volatility predictions) were not satisfactory. It is a well-known fact that the quality of GARCH modeling decreases with increasing number of model parameters. For the same reason, only the lowest orders of models were applied numerically in this paper (mostly $p = q = 1$). The identification criteria (mainly AIC) mostly confirmed that such order choices do not differ significantly from the optimal ones. As the estimation of μ_t is concerned, in the context of high-frequency financial data its approximation by zero level is realistic. Other alternatives consist in the application of various econometric methods (see, e.g., Cipra, 2020). Finally, the impact of distribution of residuals ε_t's is covered approximatively by using the quasi log-likelihood approach.

2.4 Simulation study

An extensive simulation study was performed to evaluate the proposed recursive algorithms. In particular cases 1 000 time series of length 20 060 were simulated for particular models applying $\mu_t = 0$ and $\varepsilon_t \sim iid$ N(0,1) in (1). The first 60 of the 20 060 observations were used in order to determine the initial estimates, as suggested in previous sub-sections. The remaining 20 000 observations are used for the subsequent on-line estimation.

To compare recursive algorithms, figures with boxplots were produced for each model. In this subsection we present only the case of IGARCH(1,1) model (3) with parameters $\alpha_0 = \beta_1 = 0.6$ (hence $\alpha_1 = 0.4$), see Figure 1. For each parameter, boxplots of estimates at times T = 2 500, T = 5 000, T = 10 000 and T = 20 000 are shown. Every box shows the range between the first and third quartile of obtained estimates, the white bar represents the median and the long line indicates the true value of the parameter.

Figure 1 Boxplots of the IGARCH(1,1) process estimates with $(\alpha_0, \beta_1) = (0.6, 0.6)$



Source: Own construction

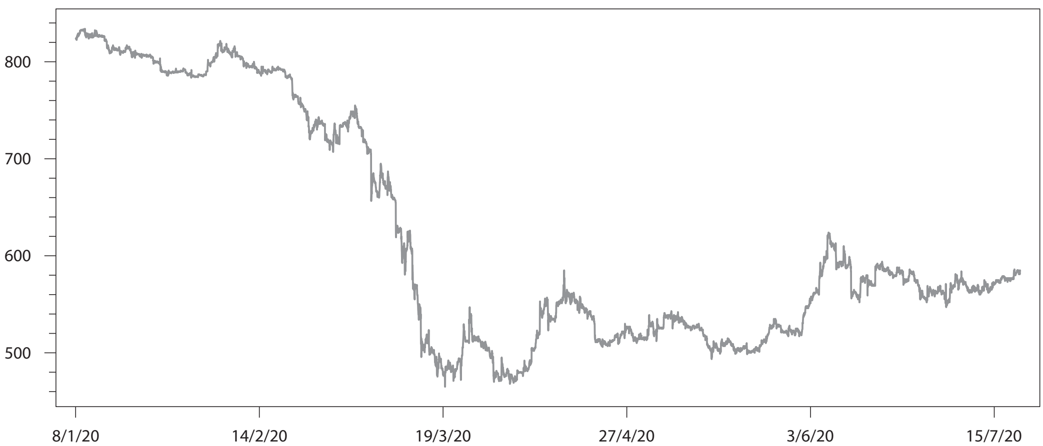
It can happen (particularly, when the true parameters are close to the borders of corresponding parameter constraints, e.g., for IGARCH(1,1) with $\alpha_0 = 0.2$ and $\beta_1 = 0.9$ that the convergence of recursive algorithms is slower when some parameters are overestimated and remaining parameters underestimated with a mutual compensation effect. Fortunately, such behavior does not distort the volatility estimation being the target output of particular recursive algorithms.

3 REAL DATA EXAMPLES

In the previous section, the ability of models to estimate parameters was verified. The primary role of the proposed recursive algorithms is their use for modeling high frequency time series. For example, for stock traded assets, one can encounter one-minute and even tick data. With such a high frequency, the volume of data is great even in a short period. In this case study, the given algorithms will be applied to a real-time series and will be also compared mutually in order to select the most suitable model for the observed time series. Even more important than the estimated parameters in a given model is the estimation of volatility, which plays a key role when trading the given asset.

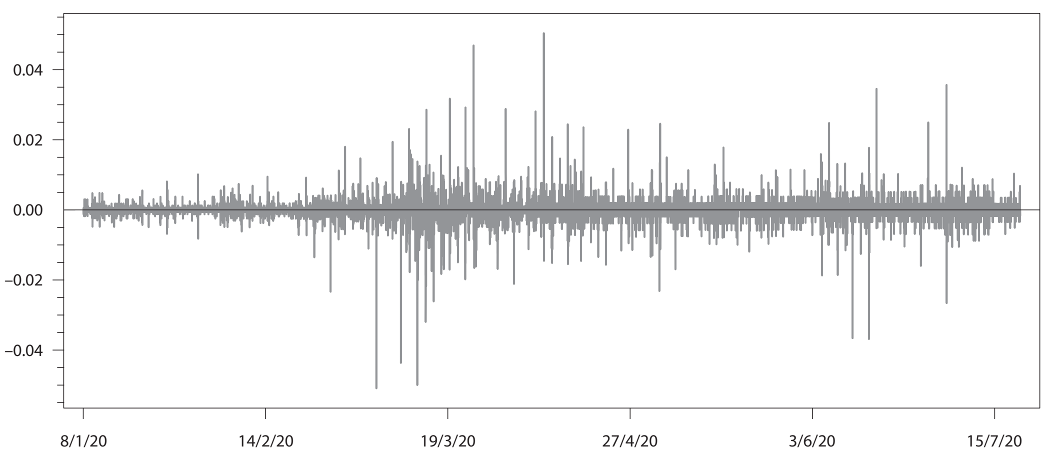
The time series consists of the stock prices of the company Komerční banka (KB) from January to July 2020. The benefit of this choice is the fact that KB operates on the Czech capital market, its stocks

Figure 2 Komerční banka stock prices from January 8th, 2020 to July 22nd, 2020, five-minute data



Source: Bloomberg

Figure 3 Logarithmic returns of Komerční banka from January 8th, 2020 to July 22nd, 2020, five-minute data



Source: Own construction

are traded on the Prague Stock Exchange (PX), and therefore, the data from Czech environment are used in the study. The second advantage consists in the investigated data period. In the given time interval, the coronavirus epidemic started in the Czech Republic, which significantly affected stock prices. Thus, we can verify how the algorithms cope with possible crises.

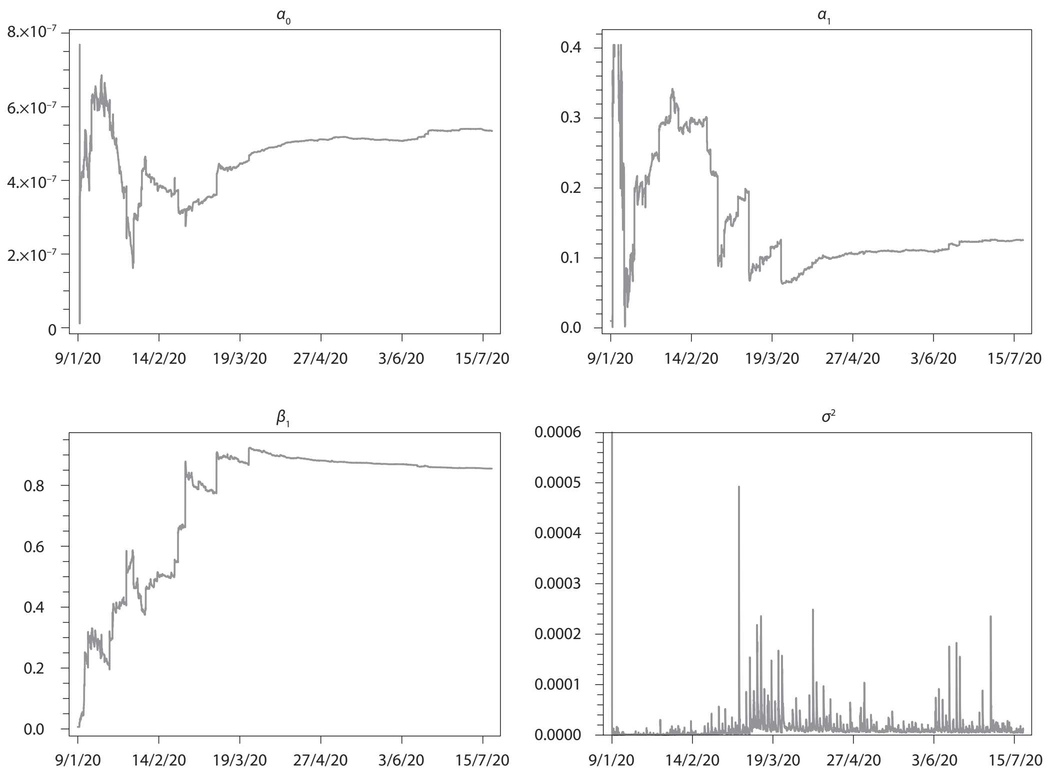
Five-minute data for the period from January 8th, 2020 to July 22nd, 2020 are available using the Bloomberg database (10 282 observations in total). These stock prices are plotted in Figure 2.

In practice, logarithmic returns are usually considered for modeling, the aim of which is, among other consequences, to make the time series stationary. Generally, logarithmic returns r_t are calculated as $r_t = \ln \frac{P_t}{P_{t-1}}$, where P_t and P_{t-1} are prices of an asset at times t and $t-1$ (in our case stock prices). The time series of the logarithmic returns of KB in the given time period is shown in Figure 3.

3.1 Estimation

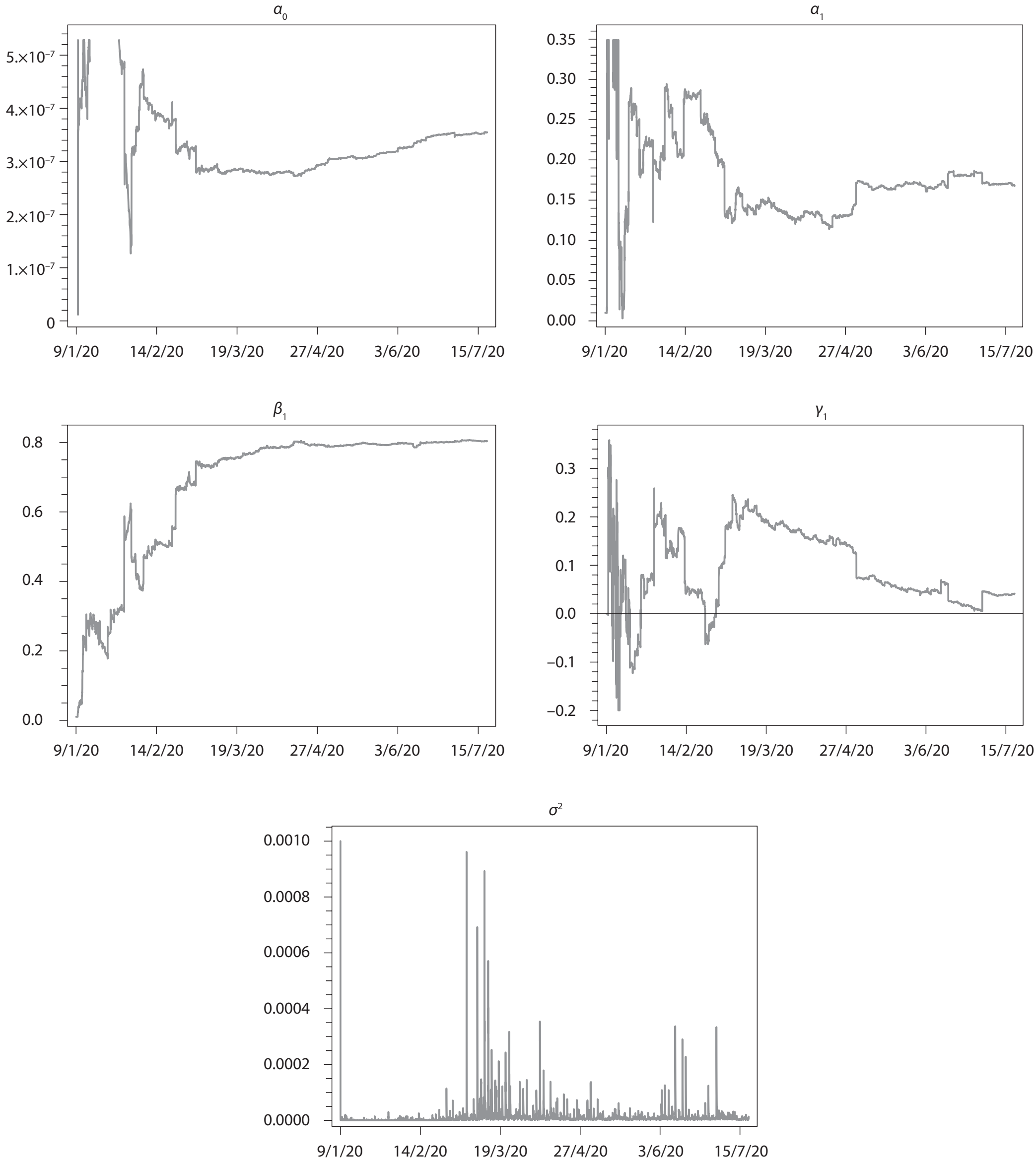
We can proceed now to the estimation of the presented models ($p = 1$, $q = 1$) for the given time series of the logarithmic returns. For each model, the graphs of the development of parameters over time and the estimate of the conditional variance are given.

Figure 4 KB: Parameter estimates and conditional volatility – GARCH(1,1)



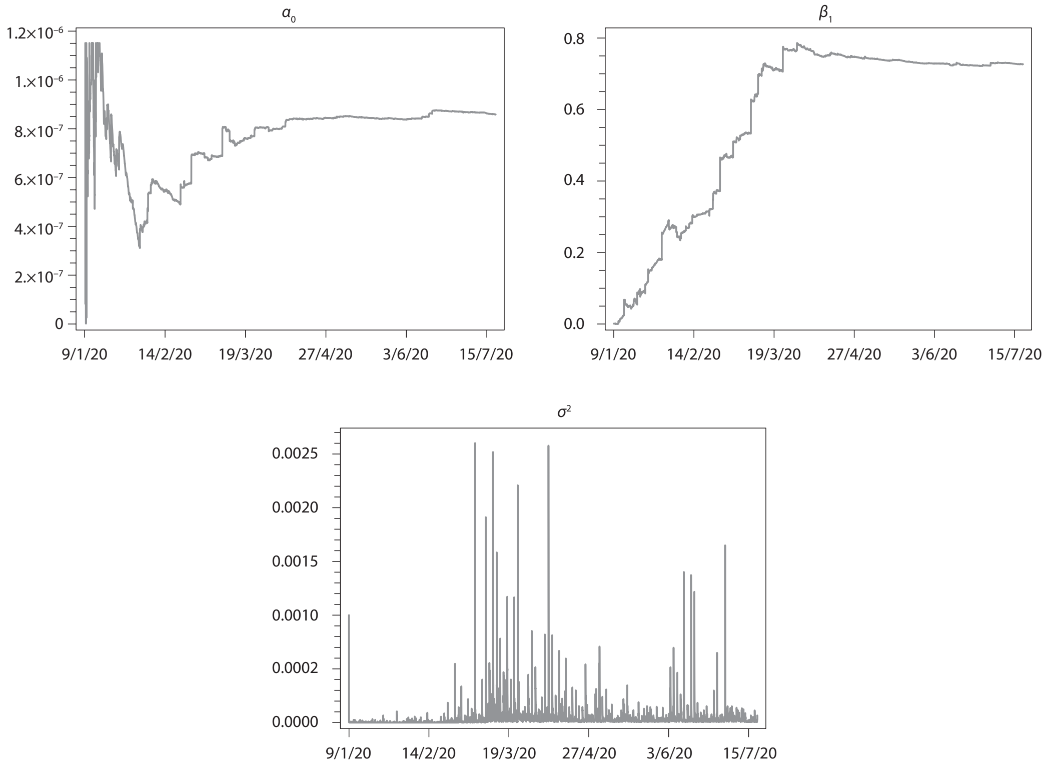
Source: Own construction

Figure 5 KB: Parameter estimates and conditional volatility – GJR-GARCH(1,1)



Source: Own construction

Figure 6 KB: Parameter estimates and conditional volatility – IGARCH(1,1)



Source: Own construction

Figure 7 KB: Parameter estimates and conditional volatility – EGARCH(1,1)

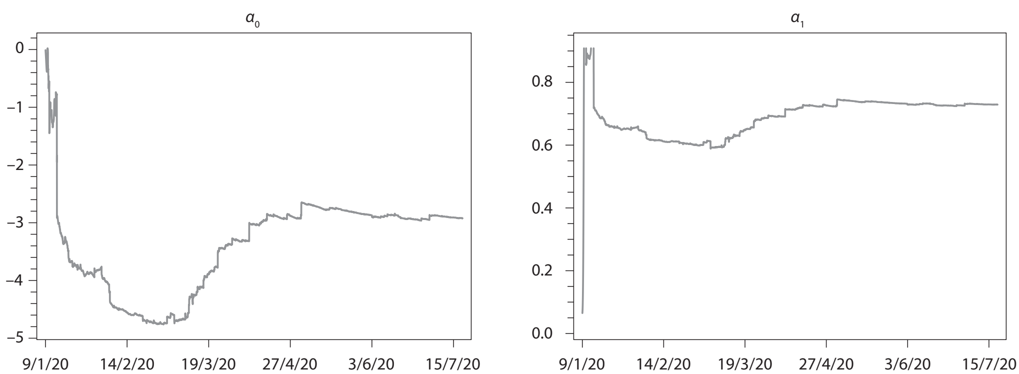
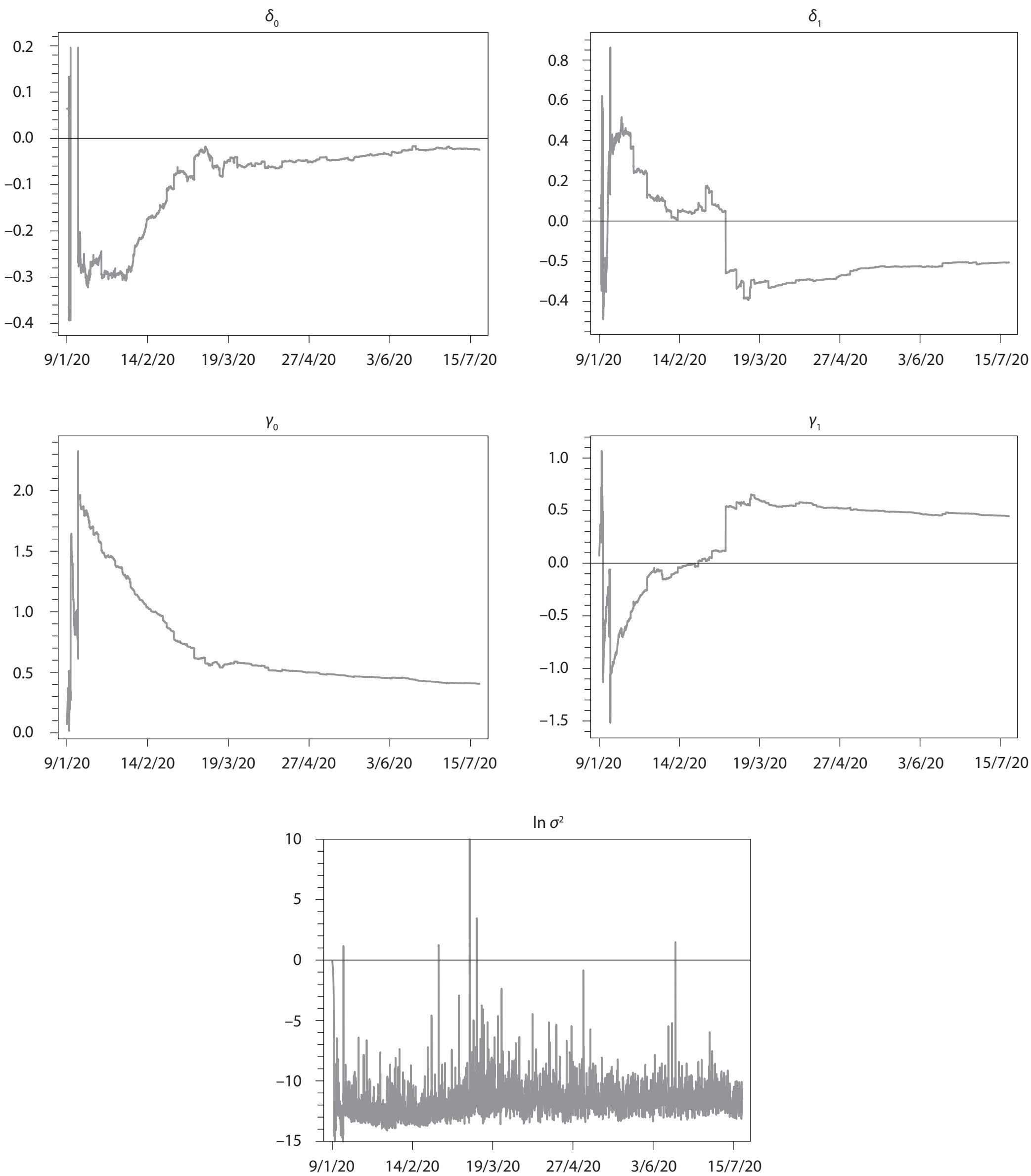


Figure 7 (continuation)



Source: Own construction

For all models except for the EGARCH model, one can observe a similar shape of the conditional volatility graph. The graphs differ only in the scale. In the case of the EGARCH model, the logarithm of the conditional variance is presented which is the output of the model.

The onset of the coronavirus crisis can be clearly identified in Figures 4–7. Significant changes in parameters are visible in this period. Moreover, a considerable increase is also evident in the estimated volatility, which was caused by a significant drop in KB stock prices. The second period of increased volatility occurred in June 2020, when the stock price gradually increased. For the GJR-GARCH and the EGARCH models taking into account the leverage effect, it is possible to verify that the given financial time series really has this characteristic. In the GJR-GARCH(1,1) model, the leverage effect is indicated by positive values of the parameter γ_1. In the EGARCH(1,1) model, the leverage effect is present when the parameters δ_0 and δ_1 are negative (the significance of positive or negative values of estimated parameters gamma and delta can be tested statistically). In both models, this is true for major parts of the time series.

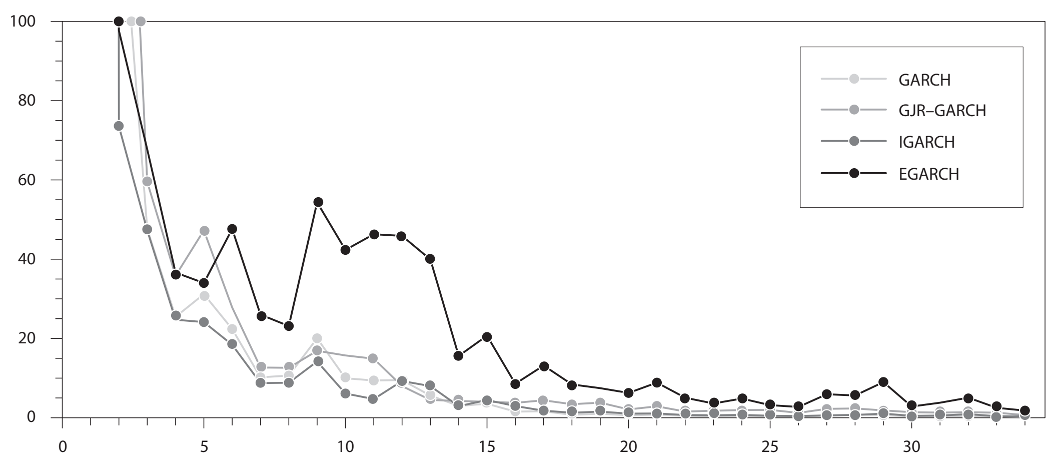
3.2 Risk prediction

Since in the case of financial time series, the ability of risk prediction is very important, we decided to use the measure for the accuracy of volatility predictions by particular models, which is inspired by MAPE (Mean Absolute Percentage Error). To decide on the best model in different periods, we divided the time series into segments with length of three hundred observations, and the following percentage quantities were calculated for each segment:

$$\widetilde{MAPE}_{i+1} = \frac{100}{300} \sum_{t=i*300+1}^{(i+1)*300} \left| \frac{\hat{\sigma}_{t+1}^{2}(t+1) - \hat{\sigma}_{t+1}^{2}(t)}{\hat{\sigma}_{t+1}^{2}(t+1)} \right|, \tag{31}$$

where i is taken as $i = 0, \ldots, 33$, $\hat{\sigma}_{t+1}^{2}(t+1)$ is the estimated volatility at time $t+1$ and $\hat{\sigma}_{t+1}^{2}(t)$ is the one-step ahead prediction of the conditional variance value at time $t+1$ with the information available till time t. Thus, the proposed measure assesses, how the given model predicts volatility one step ahead. The lower the value, the better the predictions are. The advantage of this approach is that one can model the given time series using more models parallelly and choose the best model on-line. Figure 8 and Table 1 show a comparison of $\widetilde{MAPE}$ for the given time series and the particular segments, as introduced above.

Figure 8 KB: Graph of computed $\widetilde{MAPE}$ (in percentage)



Source: Own construction

Table 1 KB: Computed $\widetilde{MAPE}$ (in percentage)

	2	5	10	15	20	25	30	34
GARCH	139.84	31.18	9.97	3.53	0.61	0.40	0.36	0.22
GJR-GARCH	211.87	47.28	15.72	4.03	2.11	1.98	1.40	0.68
IGARCH	73.52	24.14	6.12	4.47	1.05	0.51	0.48	0.35
EGARCH	97.67	33.99	42.34	20.71	6.26	3.18	2.84	1.81

Source: Own construction

Predictions are not very accurate at the beginning of the time series. However, this can be expected due to the initial calibration of the models. Later, the forecasts noticeably improved. One can notice that for the GARCH(1,1), the GJR-GARCH(1,1) and the IGARCH(1,1) models, $\widetilde{MAPE}$ decreased faster. This is could be explained by a higher number of parameters in the EGARCH(1,1) model, which takes a longer time to calibrate itself. However, in the second half of the time series, the values of $\widetilde{MAPE}$ are already very similar for all models. Figure 8 also shows that the IGARCH(1,1) model was the best in the first third of the time series, while the GARCH(1,1) model was the best in the rest. The conclusions of the residual analysis, e.g., by calculating the AIC, correspond to the findings obtained from the $\widetilde{MAPE}$ comparison of the particular models.

Figure 9 KB: Realized daily volatility calculated by intra-day data and daily volatility predictions using particular models

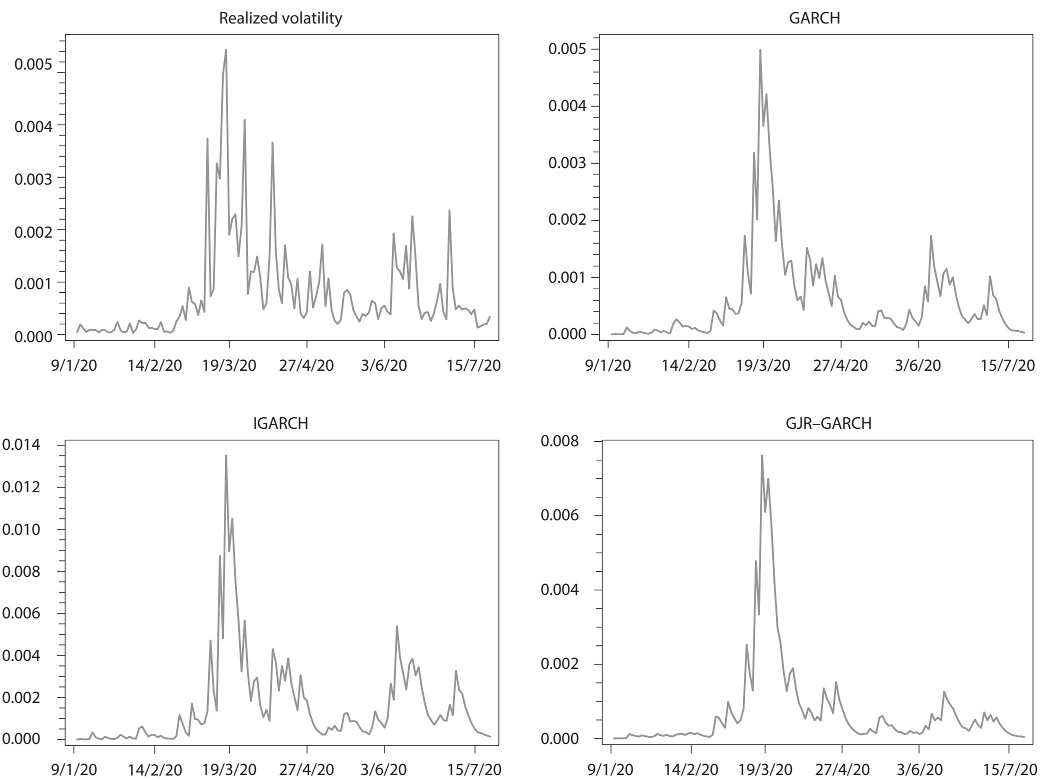
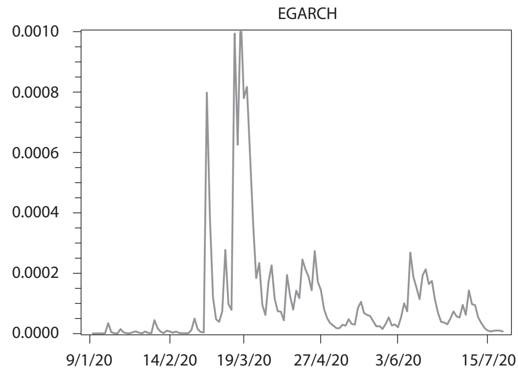


Figure 9 (continuation)



Source: Own construction

Since the five-minute predictions need not be relevant for financial practice, we have tried to evaluate the risk prediction potential of the recursive methodology also by using daily data which enables to compare the outputs of particular models with the realized volatility calculated by means of the original intra-day data (see, e.g., Patton and Sheppard, 2009). For instance, Figure 9 plots the realized daily volatilities calculated by intra-day data and the corresponding daily volatility predictions provided by particular models which are estimated recursively using daily data in particular models (for the same stocks, the same period and the same model orders $p = 1$ and $q = 1$). The consequent analysis shows that the outputs by the model GARCH(1,1) are closest to the realized volatility.

Other datasets have been used to verify the behavior of recursive estimates, e.g., the ČEZ stock prices from January to July 2020. As with KB, the stock prices were strongly influenced by the pandemic.

CONCLUSION

This article focuses on recursive algorithms for GARCH model modifications and their use for on-line estimation. The main advantages of the recursive estimation in the context of high-frequency time series are low memory requirements and overall speed. Thus, the proposed recursive algorithms can be applied to financial time series, which are typical representatives of high-frequency time series. In addition to the survey of recursive algorithms for GARCH model modifications, a nonnegligible benefit of this article consists in the numerical case study. A high-frequency time series of logarithmic returns of Komerční banka was investigated using the recursive algorithms from Section 2. The considered data are also of interest due to the fact that the observations are recorded in the period when the coronavirus pandemic started in the Czech Republic. The presented outputs certify that the algorithms have clearly identified the pandemic. Finally, we suggested an efficient methodology to compare recursive risk predictions among particular models.

ACKNOWLEDGEMENT

The paper is supported by the Czech Science Foundation 19-28231X. The authors thank to Dr. Radek Hendrych for an efficient help when preparing this paper and to anonymous reviewers for helpful comments and improvements.

References

AKNOUCHE, A. AND GUERBYENNE, H. Recursive Estimation of GARCH Models. *Communications in Statistics – Simulation and Computation*, 2006, 35(4), pp. 925–938.

BOLLERSLEV, T. Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 1986, 31(3), pp. 307–327.

BROCKWELL, P. J. AND DAVIS, R. A. *Time Series: Theory and Methods.* 2nd Ed. New York: Springer-Verlag, 1991.

CIPRA, T. *Time Series in Finance and Economics.* Cham: Springer, 2020.

CIPRA, T. AND HENDRYCH, R. Modelling of Currency Covolatilies. *Statistika: Statistics and Economy Journal,* 2019, 99(3), pp. 259–271.

ENGLE, R. F. Autoregressive Conditional heteroskedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, 1982, 50(4), pp. 987–1008.

ENGLE, R., F. AND BOLLERSLEV, T. Modelling the Persistence of Conditional Variance. *Econometric Reviews*, 1986, 5(1), pp. 1–50.

GERENCSÉR, L., ORLOVITS, Z., TORMA, B. Recursive estimation of GARCH processes. *Proceedings of the 19th International Symposium on Mathematical Theory of Networks and Systems*, 5–9 July, 2010, Budapest, pp. 2415–2422.

GLOSTEN, L. R., JAGANNATHAN, R., RUNKLE, D. E. On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. *Journal of Finance*, 1993, 48(5), pp. 1779–1801.

HAMILTON J. D. *Time Series Analysis.* Princeton: Princeton University Press, 1994.

HENDRYCH, R. AND CIPRA, T. Self-weighted Recursive Estimation of GARCH Models. *Communications in Statistics – Simulation and Computation*, 2018, 47(2), pp. 315–328.

HENDRYCH, T. AND CIPRA, T. Recursive Estimation of EWMA Model. *Journal of Risk*, 2019, 21, pp. 43–67.

KIERKEGAARD, J., NIELSEN, J., JENSEN, L., MADSEN, H. *Estimating GARCH Models Using Recursive Methods.* Lyndgby: Working Paper, Technical University of Denmark, 2000.

LJUNG, L. *System Identification: Theory for the User.* 2nd Ed. Upper Saddle River (NJ): Prentice Hall, 2000.

LJUNG, L. AND SÖDERSTRÖM, T. *Theory and Practice of Recursive Identification.* Cambridge (MA): MIT Press, 1983.

NELSON, D. B. Conditional Heteroskedasticity in Asset Returns: a New Approach. *Econometrica*, 1991, 59(2), pp. 347–370.

PATTON, A. J. AND SHEPPARD, K. Optimal Combinations of Realised Volatility Estimators. *International Journal of Forecasting*, 2009, 25, pp. 218–238.

TSAY R. S. *Analysis of Financial Time Series.* 2nd Ed. Hoboken: John Wiley & Sons, 2005.

Drivers of Food Prices: New Evidence from Turkey

Mustafa Ozan Yıldırım[1] | *Pamukkale University, Denizli, Turkey*

Abstract

This study intends to determine the drivers of high food prices in Turkey by employing the Structural Vector Auto Regression (SVAR) model for the January 2011 and March 2021 periods. The study has used external and domestic factors such as oil prices, world food prices, interest rate, exchange rate, money supply growth rate, producer price in agricultural goods. The findings indicate that all determinants show a significant positive contribution to the explanation of food prices except oil prices. The most substantial explanatory factor of food price is the price inertia shock in food prices. Domestic factors such as producer prices, interest rate, money supply, and exchange rate have also contributed to high food prices, while oil prices and world food prices have not played any substantial role. The results are robust compared to a different SVAR model identified by Cholesky decomposition. It is inferred that both exchange rate and monetary expansion have been quite effective in variations of food price in recent years. Overall, the findings indicate that controlling the food price movements is critical to ensuring overall price stability in the Turkish economy.

Keywords

Food prices, monetary policy, Turkey, SVAR

JEL code

C32, E52, Q18

INTRODUCTION

In recent years, the increase in food prices has become one of the key problems of many developing countries like Turkey. This upswing exerts pressure over any country's social and economic conditions since food is an important part of mandatory consumption of households, especially on the impoverished ones who spend significant money on food (Abdullahi, 2015; Eştürk and Albayrak, 2018). On the other hand, since food prices widely determine headline inflation, the rapid increases and volatility in food prices hamper inflation targeting, which most central banks officially announced (Chadwick and Bastan, 2017; Iddrisu and Alagidede, 2021). Increases in food prices in many developing countries distort inflation forecasting, which may have a detrimental impact on inflationary perceptions and public morale (i.e., the Central Bank's credibility), all of which are critical for the effectiveness of inflation targeting. When food prices rise due to exogenous shocks, overall inflation eventually follows, and citizens' quality of life suffers (Bhattacharya and Gupta, 2018; Wu and Wu, 2021).

Food price increases might affect headline inflation in direct and indirect ways (Rangasamy, 2011). The weight of food in the consumer basket determines the direct effect of rising food prices on overall inflation. When the food price rises are higher than those of the other items in the basket, food inflation contributes more to the overall inflation than the food weight in the consumer basket. This is more

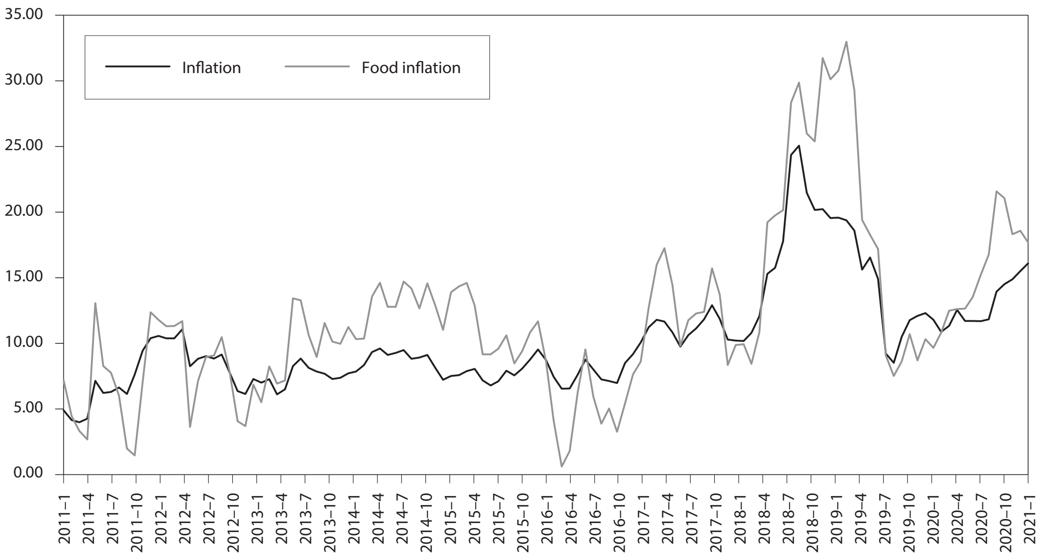
[1] Assistant Professor, Department of Economics, Kınıklı Campus, 20070, Denizli, Turkey. E-mail: moyildirim@pau.edu.tr. ORCID 0000-0001-7982-4181.

appropriate for low-income and developing countries that usually have a higher weight in their Consumer Price Index (CPI) for food products (Iddrisu and Alagidede, 2021). The effect of food prices on headline inflation may also be "indirect" to influence inflationary expectations, incomes, and prices in the CPI. Indirect effects are generally called 'second round' inflation effects in empirical research (Rangasamy, 2011).

Food consumption constitutes a significant share of household spending in low-income and developing economies. Hence, changes in food prices lead to significant fluctuations in these leading inflation-targeting countries. As a major developing economy, the share of food consumption in Turkey has a significant impact on household spending. According to the Household Budget Survey of 2019, it is reported that the proportion of the household's total expenditures on food and non-alcoholic beverages increased from 20.3% in 2018 to 20.8% in 2019, which is the highest share in recent ten years. It is also revealed that the weight in the food and non-alcoholic beverages reached 25.94% in the CPI basket in 2021, which is the highest weight after 26.22% in 2012.

Turkey has faced a long time of high inflation, primarily determined by continuing high food inflation during the past decade. Figure 1 shows the tendency in food inflation[2] and overall inflation in Turkey over the 2011–2021 period. Two features worth mentioning are shown in Figure 1. Firstly, the food inflation was mostly higher than the CPI inflation over the 2011–2021 period. The average inflation is 10.36%, while the food prices inflation average is 12.08% over the 2011–2021 period. Therefore, annual food price inflation is approximately four percentage points above inflation both in 2014 and 2019. Second, there exists a strong correlation between food prices and headline inflation. Though correlation does not involve causality, the view that food inflation can be a significant cause of inflationary pressures in Turkey appears to be an intuitive support. Increases in food prices put pressure on inflation and make it impossible to achieve inflation targets, set as 5% per year since 2012. Ganioglu (2017) reveals that the main reason for the deviation in headline inflation from core inflation was the increasing food prices in Turkey. Accordingly, higher food prices also make it difficult to anchor consumer inflation expectations.

Figure 1 Food and overall inflation in Turkey

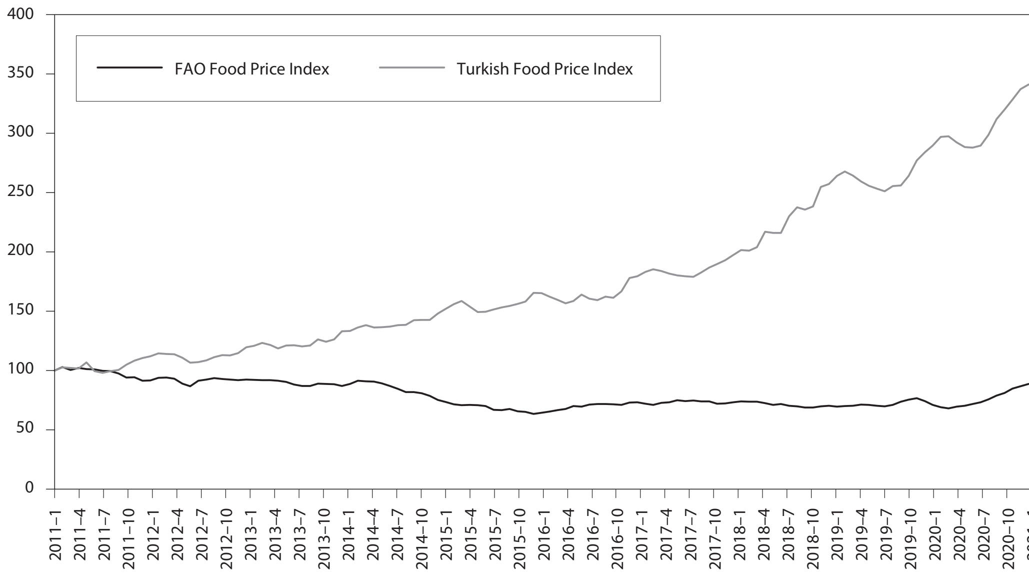


Source: Author's compilation

[2] Food inflation displays the boost in the food part of the Consumer Price Index.

As well as the difference between the food price inflation and CPI inflation, the food price index in Turkey and the world food price index have been seen to diverge further. Figure 2 shows that a co-movement existed between the world food prices and food prices in Turkey until 2012. Since then, there has been a divergence between these two price indexes. However, while world food prices decreased after 2012, food prices in Turkey continued to increase and became one of the main determinants of inflation. Akçelik and Yücel (2016) indicate that food prices in Turkey in the period after 2010 are higher than those of other developing countries in terms of both levels of food price and level of food price volatility. Işık and Özbuğday (2021) report that the food price inflation was 18% and 19.5% in Turkey while these were 1.9% and 2% in the European Union countries in 2018 and 2019.

Figure 2 Turkey Food Price Index and World Food Price Index



Source: Author's compilation

Accordingly, it is vital and useful for policymakers and households to spend a considerable amount of money on food in their budget to consider the external and domestic driving factors in explaining the food price increase. In this respect, analyzing the factors that cause huge food price increases could ensure a fruitful understanding of food price inflation and developing food policies. Nonetheless, studies on the driving factors of food price increases in Turkey are still scarce. Hence, this study investigates the external and domestic variables that reveal the increases in food prices in Turkey. The variables that are fundamental for food price inflation are the oil price, the world food prices, monetary factors, the exchange rate, production prices in the agricultural sector.

This study examines the following questions: (1) which type of external and domestic shocks give a superior interpretation of food price inflation? (2) what are the other factors that specify food inflation in Turkey in addition to those factors? (3) to what extent does the expansionary monetary policy, which is implemented after 2015, impact food prices? (4) which types of policies can be proposed? So, this study provides answers to these questions using a Structural Vector Autoregression (SVAR) model, allowing for the dynamic interlinkages among these variables. This study adds to the existing literature by empirically examined role of external and domestic factors in explaining food prices by utilizing monthly data from January 2011 to March 2021.

The rest part of the study is formulated as follows. Section 1 shortly reports the empirical studies that explain the important factors driving the food price inflation in different countries. Section 2 introduces the empirical methodology to gauge the contribution of several factors to food prices and the data used. Results of the estimation are reported in Section 3. Last section provides a summary of the findings and conclusion of the study.

1 LITERATURE REVIEW

Considerable studies have investigated the factors influencing food prices in different economies, especially with recent sharp upswings in food prices globally. Since the proportion of food in the consumption spending is higher in developing countries, the number of studies for these countries in the literature is relatively big. In one of the previous studies from African countries, Kargbo (2000) examines the role of monetary and macroeconomic aspects in explaining significant increases in food prices in Eastern and Southern Africa countries such as Ethiopia, Tanzania, Sudan, Malawi, Kenya, South Africa, Zambia by using cointegration method over the 1960–1996 period. The findings indicate that food production, income, trade policy restrictions, real exchange rate, and monetary policies are mainly responsible for driving food prices. A similar study by Kargbo (2005) employed the VECM model to examine the monetary and macroeconomic factors' impact on food prices in some West African countries such as Senegal, Ghana, Nigeria, Cote d'Ivoire. The study indicates that trade policy, real exchange rates, and monetary policy innovations substantially affect food prices.

Recently, a couple of studies that examine the relationship between monetary policy, interest rates, exchange rate, and food prices have increased by considering both advanced and developing countries' experience. However, in the existing studies based on the literature in both countries, diverse findings are obtained with different techniques utilized by the researchers. Akram (2009) indicates that shocks to interest rate and real exchange rate positively affect explaining substantial shares of fluctuations in commodity prices by employing the Structural VAR approach for the US. In a similar vein, Hammoudeh et al. (2015) examine the role of interest rate shock on commodity prices in the United States by employing the SVAR model. The study shows that a positive interest rate innovation causes a positive and insistent increase in the variation of food prices. In contrasting evidence, Abdullah and Kalim (2012) reveal that monetary shock does not contribute to food prices in Pakistan, while supply-side factors have a dominant role in explaining the food prices. Ahsan et al. (2012) examine the macroeconomic determinants that trigger the food prices in Pakistan by using the ARDL cointegration method. Their finding points to the importance of the money supply in explaining food inflation in both the long and short term. The study also indicates that agricultural subsidies have a slight impact on reducing food prices. In another study, Awan and Imran (2015) investigate the cost-push and demand-pull factors that affect food inflation in Pakistan. Their result shows that per capita GDP, fertilizer prices, money supply, fuel prices, and foreign aid have a positive impact on food prices, while the exchange rate is negatively associated with food prices. Concerning studies in Nigeria, Abdullahi (2015) examines the driving factors in food price inflation in Nigeria by employing the cointegration test and the VECM model. The study finds that GDP and energy price plays a significant role in food price inflation, whereas money supply and exchange rate lower food price in the long-term period. In a related study, Egwuma et al. (2017) investigate the link between food inflation and different macroeconomic indicators such as output, food import, and crude oil price for Nigeria by using the cointegration method and realize that all these factors are positively related to food price inflation. Recently, Bhattacharya and Jain (2020) investigated whether monetary policy is an effective policy tool to control the food price inflation in developed and developing economies for the 2006–2016 period. They found that monetary policy shocks created a positive and important impact on food prices in both countries. Iddrisu and Alagidede (2020) examine the drivers of food prices in South Africa by applying a quantile regression approach. The study shows that monetary policy has a positive and substantial effect on food prices. In addition, they reveal that exchange rate fluctuations, transport cost, and world food prices are significant determinants of food prices in South Africa, while GDP does not have a significant effect. Iddrisu and Alagidede (2021) investigate the nexus between monetary policy and food price using the quantile regression approach in Ghana, where the proportion of food consumption is 43.9% in the CPI basket. The study shows that a contractionary monetary policy to control the increase in general inflation destabilizes food prices. They also demonstrate that output and transportation cost contributed significantly to the explanation of food inflation while fluctuations in world food prices and exchange rate do not play an important role. Fasanya and Olawepo (2018) show the effect of lending rate, oil price, and exchange rate shocks on Nigeria's variation of food prices by using multivariate GARCH models.

Previous studies investigating international oil and world food prices impact on domestic food prices have been somehow mixed in different countries. Holtemöller and Mallick (2016) reveal that inflationary supply shocks arising from global food prices play a substantial role in food prices in India. Norazman et al. (2018) reveal that real effective exchange rate and world food commodity prices are the most important factors that clarify the food price fluctuations in Malaysia. On the other hand, Baltzer (2013) indicates that international prices do not play a significant role in domestic food price fluctuations in China and India. Similarly, Bhattacharya and Gupta (2018) study the explaining factors of rising food prices in India in the recent decade by considering SVAR and SVECM approaches. The study reveals that agricultural wage inflation is largely responsible for the rapid food price increase. However, international prices play a limited role in food price inflation even if they have a significant pass-through impact on, especially tradable goods. El-Karimi and El-Ghini (2020) study the pass-through of world commodity prices to food prices in Morocco across different commodities by employing the SVAR method. The study reveals that world food prices' effect on domestic food inflation is positive, and there is a powerful imported component in the food consumption basket.

There are also studies in which variables related to the agricultural sector, especially agricultural production, are used. Rangasamy (2011) explores the determinants of food price fluctuations in South Africa using the VAR model. The study reveals the dominating role of domestic factors such as nominal exchange rate, household expenditures on food, and food production price in explaining food price inflation. Irz et al. (2013) investigate both short and long-term food price dynamics by estimating a vector error-correction (VEC) model in a cointegration framework for Finland. Their findings point out that agricultural commodities, labor, and energy substantially affect the food price inflation. Joiya and Shahzad (2013) investigate the driving factors of food price increases in Pakistan by employing the ARDL model. The study reveals that GDP and food export play a positive role in food prices while food imports and credit to the agriculture sector reduce food prices. Their finding also highlights agricultural loans as an effective tool to control the increase in food prices. Ismaya and Anugrah (2018) examine the driving factor of food inflation in Indonesia by applying GMM estimation. The study points out that agriculture sector output, food import, food production, infrastructure, demand level, agriculture sector credit play an important role in explaining food prices. In one of the most recent studies that examine the determinants of food price, Wu and Xu (2021) investigate the impacts of shocks in agricultural output, production material price, and production price on food price for 26 provinces in China by applying a heterogeneous panel structural vector autoregressive (SVAR) approach. According to the findings of the study, price inertia shock (food price shock) is the main responsible for the driving force of food price.

Bayramoğlu and Yurtkur (2015) investigate the determinants of food prices in Turkey by employing the VAR model over the 1999–2014 period. Their findings indicate that the US dollar and Euro exchange rates play an important role in determining short-term innovations in food prices. The study reveals that there has been limited contribution of oil prices, agricultural producer prices, and world food prices to the variation of food prices. Altıntaş (2016) examines the impact of oil prices on food prices by employing an asymmetric framework over the 2000–2013 period. The study indicates that a rise of 1% in oil prices brings about 0.47% increases in food prices, whereas a 1% decline in oil prices leads to a 0.19% lessen in food prices. It is inferred that the positive oil price shock has a more significant effect on food prices than negative oil price innovations in Turkey. Işık and Özbuğday (2021) utilize the cointegration approach to consider the role of agricultural input prices in explaining the recent rapid increases in food prices in Turkey and confirm the existence of positive contribution of agricultural input prices on food prices. Ertugrul and Seven (2021) explore the differences between world food prices and Turkish food prices. Their findings indicate that the exchange rate plays an important role in the increasing differences between both, whereas oil prices contribute to lessening those differentiations.

2 METHODOLOGY AND DATA

This section introduces the methodology, identification framework, variables in the model, and the data. The SVAR model allows one to investigate the response of food prices to unanticipated shocks by considering the dynamic relationship between food prices and macroeconomic variables.

2.1 The SVAR methodology

The SVAR model gives a facility to identify restrictions in harmony with economic reasoning and preceding expertise. The structural identification of a VAR model is given as the following:

$$\mathrm{A}\mathrm{y}_t = \mathrm{C(L)}\mathrm{y}_t + \mathrm{B}\mathrm{u}_t\,, \tag{1}$$

where A is the matrix of contemporaneous interactions between variables, y_t is an (n × 1) vector of the endogenous macroeconomic variables, C(L) is an (n × n) matrix of lag operator L, representing impulse-response functions of the shocks to the elements of y_t, B is an (n × n) matrix which captures the linear relations between structural shocks and those in the reduced form; finally, u_t presents an (n × 1) vector of structural shocks which are uncorrelated and identically distributed in a normal manner.

Unfortunately, Formula (1) cannot be estimated directly because of identification problems; the reduced form is determined by multiplying Formula (1) by an inverse matrix A^{-1} to estimate the SVAR model:

$$y_t = D(L)y_t + u_t\,, \tag{2}$$

where $(L) = A^{-1}\,C(L)y_t$, $u_t = A^{-1}\,Bu_t$. u_t is an $(nx1)$ vector of shocks in a reduced form that are uncorrelated and normally distributed yet contemporaneously correlated with each other. The relation between structural shocks and reduced-form shocks is the following:

$$Au_t = B\varepsilon_t\,. \tag{3}$$

Formula (3) is also known as the short run AB model. To obtain the SVAR parameters in Formula (1), one can easily impose a constraint on matrix A and B. To identify structural parameters given a $(kx1)$ dimensional VAR, one would require general $k^2 + \frac{k(k-1)}{2}$ restrictions in the short run AB model on the SVAR (see Amisano and Giannini, 2012). The identifying restrictions are assumed on the structural parameters as follows:

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ a_{21} & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & a_{43} & 1 & 0 & 0 & 0 \\ a_{51} & 0 & a_{53} & a_{54} & 1 & 0 & 0 \\ a_{61} & a_{62} & a_{63} & a_{64} & a_{65} & 1 & 0 \\ a_{71} & a_{72} & a_{73} & a_{74} & a_{75} & a_{76} & 1 \end{bmatrix} \begin{bmatrix} u^{oil} \\ u^{wfp} \\ u^{interest} \\ u^{money} \\ u^{exch} \\ u^{ppi} \\ u^{fprice} \end{bmatrix} = \begin{bmatrix} \beta_{11} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \beta_{22} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \beta_{33} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \beta_{44} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \beta_{55} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \beta_{66} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \beta_{77} \end{bmatrix} \begin{bmatrix} \varepsilon^{oil} \\ \varepsilon^{wfp} \\ \varepsilon^{interest} \\ \varepsilon^{money} \\ \varepsilon^{exch} \\ \varepsilon^{ppi} \\ \varepsilon^{fprice} \end{bmatrix},$$

where a_{ij}s are the coefficients to be estimated. u^{oil}, u^{wfp}, $u^{interest}$, u^{money}, u^{exch}, u^{ppi}, u^{fprice} are structural shocks, while ε^{oil}, ε^{wfp}, $\varepsilon^{interest}$, ε^{money}, ε^{exch}, ε^{ppi}, ε^{fprice} are the reduced form residuals.

In monetary policy literature, the oil price is a widely used indicator and represents the inflationary and negative supply shock (Kim and Roubini, 2000). According to the structural identification above, since the oil price (*oil*) is considered as the external factor for Turkey, it does not react simultaneously to shocks caused by other endogenous model variables. World food price (*wfp*) does not react simultaneously to macroeconomic factors in Turkey, while it is contemporaneously impressed by the oil price shock. The interest rate (*interest*) is a factor that affects food price (*fprice*) but is not simultaneously influenced by other variables' shocks. It starts to react only one period after a financial or exchange rate shock. The money supply (*money*) only responds contemporaneously to interest rate shock but is not contemporaneously influenced by other variables' shocks. The exchange rate (*exch*) is contemporaneously affected by oil price, interest rate, and money supply. The producer price in agriculture (*ppi*) is affected contemporaneously by all other variables' shocks. Finally, the food price is permitted to respond simultaneously to all other variables.

2.2 Data

This study uses monthly frequency data ranging from January 2011 to March 2021 to examine the effect of driving factors on food prices in a SVAR approach. The variables selected for investigating the food price dynamics are consistent with the current literature:[3] oil prices, world food prices, interest rate, money supply growth rate, exchange rate, producer prices of agriculture products, and food price. Since Turkey is situated between Europe and Asia, it mainly relies on countries closer to Europe to maintain its economic ties. For the purpose of this study the oil price for Brent (Europe) is chosen based on US dollars per gallon. The world food price variable reflects the international pressure on food prices. A new monetary policy framework in CBRT has been created to monitor financial stability as well as price stability since the end of 2010. Within this framework, the CBRT has started to use more than one interest rate as a policy tool, such as the BIST overnight repo rate, policy rate, and overnight borrowing/lending rate. Since the new monetary policy framework consists of a combination of various policy instruments, the Weighted Average Funding Cost (WAFC) data is chosen as the interest rate variable in the study to reflects the new monetary policy stance. However, recent studies use the WAFC as an interest rate variable in the Turkish economy (Bastav, 2020; Tümtürk, 2020). Monetary policy is represented by the WAFC and measured in percentage. The WAFC is the official monetary policy tool in Turkey after 2011. The growth rate of M2 is utilized to reflect the level of economic activity and aggregate demand (Kargbo, 2005). Since the Turkish economy trades with many different countries, mainly European Union countries, it is used the mixed basket of US dollar and Euro currency. A basket of 0.5 USD + 0.5 EUR representing the nominal exchange rate is used as the nominal exchange rate. The CPI for food represents the food price inflation in Turkey, as in several studies (Irz et al., 2013; Abdullahi, 2015; Bhattacharya and Jain, 2020). The producer price index of agricultural products reflects the cost effect on overall

[3] See the studies of Rangasamy (2011), Bhattacharya and Jain (2020), Iddrisu and Alagidede (2021).

inflation and food prices. Table 1 presents the definition of variables and data sources. All variables are converted into logarithm form and are seasonally adjusted except the interest rate.

Table 1 Definition of variables

Variable	Definition	Source
oil	Europe Brent Spot Price	Central Bank of Republic of Turkey
wfp	World food price index	Food and Agriculture Organization (FAO
interest	Weighted Average Cost of Funding	Central Bank of Republic of Turkey
money	M2 money supply	Central Bank of Republic of Turkey
ppi	Producer Price Index of Agricultural Products	Turkish Statistical Institute
exch	Basket of USD Dollar and Euro	Central Bank of Republic of Turkey
fprices	Food Price Index	Central Bank of Republic of Turkey

Source: Author's compilation

The findings of the summary statistics of the selected macroeconomic variables and food prices are presented in Table 2. The food price index averaged 344.06 over the period, ranging from 189.91 in July 2011 to 658.96 in March 2021. The oil price index ranged from 14.85 to 126.59 with an average of 75.336 and a standard deviation of 28.64. The average world food price index is 106.41 over the period with a standard deviation of 14.41. Monetary policy rated averaged 10.84% during the period, from a minimum of 4.52% in May 2013 to a maximum of 25.5% in March 2019. The growth rate of the money supply showed considerable fluctuations over the period, dropping to as low as 5.6% in August 2012 and rising to as high as 44.69% in October 2020. The exchange rate of Turkish Lira to the mixed basket of US Dollar and Euro ranged between 1.813 and 8.730. The average of producer price index is 113.815 during the period investigated.

Table 2 Descriptive statistics (level series)

Statistics/variables	*oil*	*wfp*	*interest*	*money*	*exch*	*ppi*	*fprice*
Mean	75.336	106.413	10.840	19.482	3.856	113.815	344.068
Max.	126.59	137.612	25.500	44.693	8.730	216.380	658.960
Min.	14.850	84.866	4.520	5.681	1.813	71.440	189.910
Std. Dev.	28.647	14.419	5.254	8.201	1.950	36.933	126.724
Skewness	0.183	0.515	1.387	1.083	0.954	0.989	0.779
Kurtosis	1.679	1.875	4.017	4.436	2.686	2.930	2.467
Jarque-Bera	9.627	11.924	44.792	34.668	19.166	20.086	13.909
Obs.	123	123	123	123	123	123	123

Source: Author's compilation

3 EMPIRICAL ANALYSIS

3.1 Unit Root Tests

This part performs the stationary properties of the variables via the Augmented Dickey-Fuller (ADF) and Phillips Perron (PP) tests. Table 3 summarizes the results of both tests for the investigated variables. All the variables have a unit root in their levels, except the interest rate for the ADF test, and become stationary when they are first differenced at the 1% level. Since all the series are non-stationary in their differences, the model is estimated in the first differences.

Table 3 ADF and PP test results

	ADF Test		PP Test	
Variable	**Level**	**First difference**	**Level**	**First difference**
oil	−2.703 (0.237)	−9.136 (0.000) ***	−2.772 (0.210)	−11.329 (0.000) ***
wfp	−0.014 (0.995)	−9.136 (0.000) ***	0.257 (0.998)	−9.141 (0.000) ***
interest	−3.528 (0.040) **	−3.565 (0.037) **	−2.512 (0.321)	−9.034 (0.000) ***
money	−2.825 (0.191)	−10.098 (0.000) ***	−2.825 (0.191)	−10.081 (0.000) ***
exch	−1.932 (0.631)	−8.762 (0.000) ***	−2.132 (0.522)	−7.418 (0.000) ***
ppi	−1.309 (0.880)	−9.457 (0.000) ***	−0.882 (0.953)	−13.665 (0.000) ***
fprice	−1.564 (0.801)	−9.942 (0.000) ***	−1.432 (0.846)	−10.940 (0.000) ***

Notes: ***, ** and * present the significance at 1%, 5%, and 10% levels, respectively. All tests are conducted for the trend and intercept models. The Schwarz Information Criterion for the selection of lag length is determined when employing the ADF test. The estimate of PP test is based on the Bartlett-Kernel with the aid of the Newer-West bandwidth.
Source: Author's compilation

When the estimate is carried out, the information criteria for optimal lag length selection are determined. The order of the unrestricted VAR has been selected as one according to the Akaike information criterion (AIC), Hannan-Quinn information criteria (HQ), and Schwarz information criteria (SBC), and the stability condition is satisfied.[4] Before obtaining the structural shocks of the SVAR model, it is also necessary to verify the stability of the underlying VAR structure. It is estimated a VAR model and found out that all the eigenvalues lie within the unit circle (see Table A2). This means that the VAR meets the stabilization criterion, and it is safe to proceed with the structural model's Impulse Response Function analysis. Since the SVAR model is over-identified, according to the contemporaneous relation matrix defined in the previous part, it is desirable to control the over-identifying restriction test to prove the validity of the identifying restrictions imposed in the model. The likelihood ratio (LR) test is 5.18 [0.3941], which is higher than 0.05, showing that over-identification is valid. Diagnostic tests of the underlying VAR process is conducted. The LM test provides (see Table A3) the absence of serial correlation in residuals from the VAR model.

3.2 Impulse Response Functions

This section indicates the findings obtained from the impulse response functions (IRFs) to reveal the variables affecting food prices. The IRFs demonstrate the impact of a one standard deviation shock to each of the variables in the model in a certain period. Since the primary interest is to figure out the effect of macroeconomic variables shocks on food prices, Figure 3 only represents the response of food prices to oil price, world food price, interest rate, money supply growth, exchange rate, and producer price shocks. The IRFs illustrate the dynamic path of adjustment to shocks on endogenous variables up to 10 months.

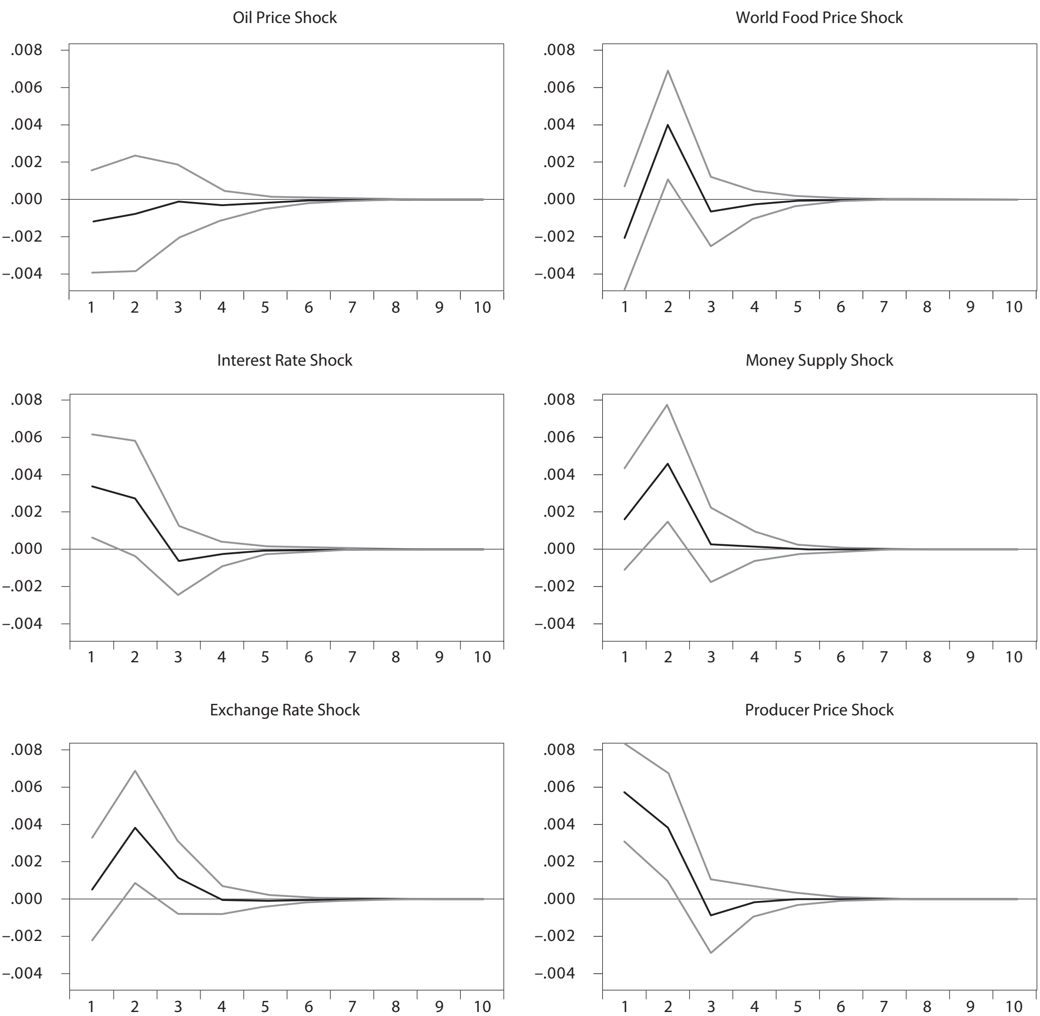
Figure 3 indicates that all the shocks that affect the dynamics of food prices have small and short effects. The response of food prices to oil prices is not statistically significant and does not change over the months. This finding is in line with the study of Ahsan (2012) for Pakistan. Following an unexpected

[4] The lag selection criteria are presented in Table A1. According to AIC and FPE selection criteria, a one lag vector autoregressive model is estimated.

world food price shock, food prices react positively after an initial downturn. An increase in world food prices, on the other hand, has a temporal contribution to food prices. It increases food prices after two months of the shock, but the effect lasts insignificantly after that. This finding is in line with the El-Karimi and El-Ghini (2020) for Morocco.

Similarly, the response of food prices to interest rate shock is positive and significant. Rising the interest rate reflects contractionary monetary policy in Turkey. A rise in monetary policy tightening instantaneously increases food inflation, but the effect lasts for two months and then dies out. The contribution of monetary policy to food prices is in line with the previous findings in the literature Hammoudeh et al. (2015), Bhattacharya and Jain (2020), and Iddrisu and Alagidede (2021). The contractionary monetary policy gives rise to an increase in interest rates that is borrowed for using capital in production.

Figure 3 Response of food prices to other variables, sample: 2011M1–2021M3



Source: Author's compilation

Hence, production becomes expensive, and the rising production costs cause upswings in food prices. The positive money supply shock shows a statistically significant but short-term contribution to food prices. An increase in the growth rate of the M2 money supply reflects an easing of the monetary policy, which turns to an increase in the amount of credit. When more money is demanded for food consumption, the food prices go up, which shows a demand-driven inflationary pressure. The finding is compatible with the results of Awan and Imran (2015) for Pakistan. As may be expected, the effect of the exchange rate shock on food price is positive and continues for nearly four months, but insignificant after the third month. This finding is in line with Başkaya et al. (2008) regarding the short-term effect on food prices. The depreciation of the Turkish Lira against foreign currency (Dollar and Euro) brings about a rising food price inflation in the short period. Food prices are very sensitive to depreciation in the exchange rate. Food prices respond immediately to producer prices in agriculture shocks. A positive shock to the producer price index in agriculture brings about an expected rise in food prices. The impact of this shock on food prices appears to completely disappear in the third month. This result is supported by Irz et al. (2013) for Finland.

3.3 Variance decomposition analysis

The variance decomposition presented in Table 4 demonstrates the information about the relative importance of each random innovation to variables in the model. Using variance decomposition, it can be seen how much of the shocks occurring in the variables are accounted for by the own shock and the shocks of other variables. The variance decomposition of variables in the model is reported by considering the 1, 5, 10, and 20 months prediction horizons for food prices. In the first month, two important dynamics explain the variation in food prices: producer price (13.49%) and interest rate (4.74%). The effects of oil prices (0.58%), world food prices (1.82%), money supply growth (1.07%), and exchange rate (0.14%) shocks are barely apparent in the first month. The main part (78%) explaining the variance in food prices is accounted for by its shocks in the first month. After 20 months, around 15% of the variation in food price is accounted for by producer prices in the agriculture sector, followed by the growth of money supply (7.28%), world food price (6.34%), interest rate (5.89%), and exchange rate (5.04%) respectively. A tiny part of the forecast error variance of food prices can be related to shocks in oil prices (0.66%). The variable for food price has more than 60% of its variance accounted for by own-innovations for the entire forecasting horizon in Turkey. This finding shows that the food price has an important impact on itself, and the current high food price generates expectations of future high food price inflation (i.e., food price inertia). The second highest contribution to the variation of food price is producer price shocks.

The variance decomposition analysis suggests that supply-side variables play a role in explaining a moderate proportion of the variation in food prices in Turkey over the investigated period. Furthermore, even though external shocks play some role, food price variation in Turkey is explained mainly by domestic

Table 4 Forecast error variance decomposition analysis, sample: 2011M1–2021M3

Horizon	Oil price	World food price	Interest rate	Money growth	Exchange rate	Producer price	Food price
1	0.582	1.823	4.745	1.074	0.143	13.490	78.096
5	0.652	6.340	5.898	7.286	5.043	14.746	60.191
10	0.660	6.340	5.898	7.286	5.044	14.745	60.188
20	0.660	6.340	5.898	7.286	5.044	14.745	60.188

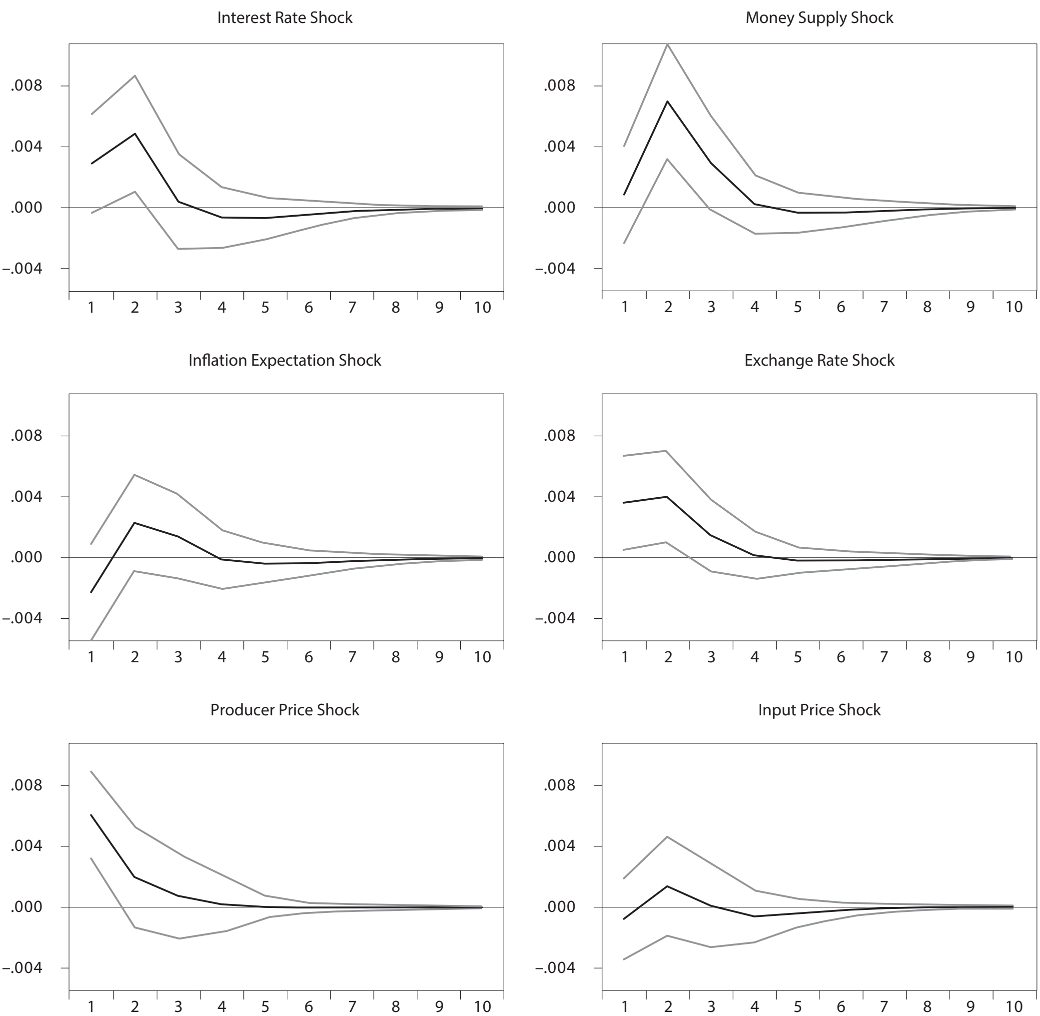
Source: Author's compilation

factors. These results are supported by the existing studies such as Rangasamy (2011), and Bhattacharya and Gupta (2018). The estimated findings are consistent with the impulse response function analysis.

3.4 Robustness checks and further evidence

This section provides a robustness check and re-estimates the SVAR model by varying identification, selecting variables, and the sample period. Instead of structural identification, it is considered the traditional Cholesky identification with all the domestic macroeconomic variables. Following the empirical literature, the variables in Cholesky specification are ordered as follows: interest rate → money supply growth rate → inflation expectation → exchange rate → producer price → agricultural input price → food prices. The variables in the Cholesky scheme are ordered from the most external to the most internal, affecting

Figure 4 Impulse response for food price, sample: 2015M1–2021M2



Source: Author's compilation

each other in one direction most time. In this direction, it is shown that in period T, the top variable is not simultaneously affected by any variable, yet it influences all other variables in the model. The second variable that comes after it reacts only to the first variable before it and affects all other variables, and the process continues in the same way. It is also considered a different sample that begins from January 2015 to February 2021. The choice of a new sample relies on incorporating two possible factors of food prices and other variables examined above. These are inflation expectation (revealed in 2013) and agricultural input price index (released in 2015). Inflation expectation data are represented by the expectation of 12 months ahead annual CPI in percentage and modified mean and obtained from CBRT. The agricultural input price index, which is released by TurkStat, monitors the variability of the inputs purchased by the producers or farmers both in the current production year and for investment purposes as a cost factor.

The empirical findings obtained both IRFs and FEVDs are like previous findings, and the signs of the responses of food prices to different shocks are close, which indicates the robustness of the link between food price and driving factors (see Figure 4).[5] However, the contribution of both inflation expectation and input prices in agriculture to food prices are in line with economic expectations but are not statistically significant.

Comparing the FEVD findings for the post-2015 period with that for the whole sample indicates that some variables in explaining food prices have strengthened in the post-2015 period. The effect of money supply on food prices has doubled after 2015. After 20 months, the money supply explains approximately 18% variation in food price, compared to 7.3% in the full-sample analysis. Once more, it confirms the perception that expansionary monetary policy has led to boost food prices because of the monetary policies implemented by CBRT in Turkey. Similarly, after 20 months of a shock, more than 13% variation in the food prices is accounted for by interest rate shocks. Thus, the contribution of interest rate changes doubles in the post-2015 period. In the post-2015 period, there is no substantial change in producer prices on food prices. After 20 months of a shock, 13.68% of the variation in food price is accounted for by producer price shock. The 8.19% variation in food prices is attributed to exchange rate shocks, which is slightly higher than the entire sample period. Furthermore, the contribution of inflation expectation (3.48%) and agricultural input price (2.42%) shocks in explaining food price volatility remain with a negligible amount after 2015. Finally, Table 5 shows that the variation in food prices comes mainly from its own shocks (50.9%) rather than from the shocks of the other variables at the end of 20 months. It indicates that the food price inertia shock continues to one of the powerful determinants for food prices, which explains approximately 50.9–71.5% in variation in food prices. Overall, these findings reveal that monetary policy shocks have a strong effect on food prices in Turkey.

Based on the above findings that are obtained by two different identification methods (i.e., structural, and recursive) and samples, it is clear that the effect of monetary policy, exchange rate, and producer

Table 5 Variance decomposition of food prices, sample: 2015M1–2021M2

Horizon	Interest rate	Money supply	Inflation expectation	Exchange rate	Producer price	Agricultural input price	Food price
1	7.324	0.020	3.250	4.042	21.605	0.230	71.504
5	13.016	17.835	3.467	8.189	13.694	2.420	51.094
10	13.022	17.839	3.482	8.190	13.686	2.420	50.972
20	13.022	17.839	3.482	8.190	13.686	2.420	50.972

Source: Author's compilation

5 It has not been interpreted in detail due to space constraints.

prices have a substantial role in determining the food prices in Turkey. The findings related to exchange rate and producer prices are consistent with those reported for Turkey by Bayramoğlu and Yurtkur (2015), Ulusoy and Şahingöz (2020), and Ertuğrul and Seven (2021).

CONCLUSION AND POLICY IMPLICATIONS

The findings presented in this study are analyzed from the SVAR approach that investigates the driving external and domestic factors in determining the food prices in Turkey by employing monthly data ranging between January 2011 and March 2021. The external determinants include the oil price and world food prices, while the domestic variables consist of interest rate, money supply growth rate, exchange rate, producer prices in agriculture. Also, to consider the recent dynamics of the Turkish economy that is characterized under expansionary monetary policy and exchange rate depreciation, a different SVAR model that is included only domestic variables by employing for January 2015 and 2021 February. In that model, there are added two new variables, such as inflation expectation and agricultural input prices, to capture the current macroeconomic environment in Turkey.

The impacts of the variables on food price are analyzed by impulse response function (IRF) and forecast error variance decomposition (FEVD). The IRFs and FEVDs analysis reveal that the response of food price to different macroeconomic variables shocks is close to the empirical predictability exhibited in developing countries. Also, the response of food prices to all shocks in the model is short-lived. The study indicates that domestic components are more significant drivers than external factors in determining food prices. According to both IRFs and FEVDs findings for the entire sample period, all variables except oil prices have a significant effect on food prices. The impact of oil prices on food prices in Turkey is not as strong as expected. However, when both oil prices and world food prices are considered, it is found that there is a limited role of international prices on food prices. Furthermore, together with the effect of the exchange rate, there are significant pass-through effects that cause a rise in domestic food prices. The most important factor that largely explains food prices is the producer price in the agricultural sector in that period. Secondly, the contributions of interest rate and money supply growth rate also impact the food price fluctuations to some extent.

Overall, the findings remain robust to the alternate SVAR model that consists of both different identification schemes (i.e., Cholesky identification) and a shorter different sample (January 2015 to February 2021). The food prices in Turkey are greatly affected by the monetary policy shocks (both interest rate and money supply growth rate). This finding suggests that the food price level tends to increase rapidly while the economy is overheating. However, the food prices respond quickly to the easing of the monetary policy. The producer prices continue to create a considerable positive influence on food prices in the new period. Moreover, the role of the exchange rate in determining food prices has increased significantly. The result implies that a pass-through of depreciation from exchange rate to food price inflation. Ertuğrul and Seven (2021) point that massive depreciation in the exchange rate and increases in import share of agricultural food reinforced with the easing of monetary policy has led to high and volatile food prices after 2013. Since the food industry depends mainly on foreign raw materials, the fluctuations in exchange rates affect the food prices in Turkey. Surprisingly, inflation expectation and agricultural input prices do not substantially explain food prices in the near past.

The findings of this study indicate that controlling food price movements is critical to ensuring overall price stability in the Turkish economy. As a result, the adverse outcome of monetary policy shocks on food prices is an additional significant reason to focus on policies capable of mitigating their impact. Monetary policy can stabilize food inflation by controlling aggregate demand in the economy. Furthermore, as food prices are economically and socially important and the share of food spending for the impoverished is high, it is crucial to monitor food prices and agricultural product markets. In addition, the implementation of policies that will stabilize the exchange rate fluctuations will reduce the exchange rate pass-through effect, and the increase in food prices will be partially controlled.

Some additional issues remain open for further research. It is possible to investigate the food price dynamics according to the different sub-sectors in Turkey. As a result, the disaggregation of the analysis across different food sectors will offer further insights. Additionally, a comparative study of driving factors in food prices in different major developing countries would provide general conclusions regarding the effectiveness of economic policies.

References

ABDULLAHI, S. A. Investigating the determinants of food price inflation in Nigeria. *Abuja Journal of Business and Management*, 2015, 1(4), pp. 12–20.

ABDULLAH, M. AND KALIM, R. Empirical analysis of food price inflation in Pakistan. *World Applied Sciences Journal*, 2012, 16(7), pp. 933–939.

AHSAN, H., IFTIKHAR, Z., KEMAL, M. A. The determinants of food prices in Pakistan. *The Lahore Journal of Economics*, 2012, 17(1), pp. 101–128.

AKCELIK, F. AND YUCEL, C. Y. Türkiye'de gıda fiyatları: uluslararası bir karşılaştırma. *CRBT Research Notes in Economics: 2016–23.* Research and Monetary Policy Department, Central Bank of the Republic of Turkey, 2016, July, pp. 1–13.

AKRAM, Q. F. Commodity prices, interest rates and the dollar. *Energy economics*, 2009, 31(6), pp. 838–851.

ALTINTAŞ, H. The asymmetric effects of oil prices on food prices: the case of NARDL modeling for Turkey. *Journal of Management and Economic Research*, 2016, 14(4), pp.1–24.

AMISANO, G. AND GIANNINI, C. Topics in structural VAR econometrics. *Springer Science & Business Media*, 2012.

AWAN, A. G. AND IMMRAN, M. Factors affecting food price inflation in Pakistan. *ABC Journal of Advanced Research*, 2015, 4(1), pp. 75–90.

BALTZER, K. International to domestic price transmission in fourteen developing countries during the 2007–08 food crisis. *WIDER working paper*, No. 31/2013.

BAŞTAV, L. Monetary policy interest rate channel in Turkey: Toda-Yamamoto method (2011–2018). *Fiscaoeconomia*, 2020, 4(2), pp. 311–331.

BAŞKAYA, Y. S., GURGUR, T., ÖĞÜNÇ, F. *İşlenmiş gıda fiyatlarını belirleyen faktörler.* TCMB Çalışma Tebliği, No 08/09, 2008.

BAYRAMOĞLU, A. T. AND YURTKUR, A. K. International factors on food and agricultural price determinations in Turkey. *Anadolu University Journal of Social Sciences*, 2015, 15(2), 63–73.

BHATTACHARYA, R. AND SEN GUPTA, A. Drivers and impact of food inflation in India. *Macroeconomics and Finance in Emerging Market Economies*, 2018, 11(2), pp. 146–168.

BHATTACHARYA, R. AND JAIN, R. Can monetary policy stabilise food inflation? Evidence from advanced and emerging economies. *Economic Modelling*, 2020, 89, pp. 122–141.

CHADWICK, M. AND BASTAN, M. News impact for Turkish food prices. *Central Bank Review*, 2017, 17(2), pp. 55–76.

EGWUMA, H., OJELEYE, O. A., ADEOLA, S. S. What determines food price inflation? Evidence From Nigeria. *FUOYE Journal of Agriculture and Human Ecology*, 2017, 1(2), pp. 48–61.

EL-KARIMI, M. AND EL-GHINI, A. Pass-through of global food commodity prices to food inflation in Morocco: a structural VAR approach. *New Medit*, 2020, 19(3), pp. 57–76.

ERTUĞRUL, H. M. AND SEVEN, Ü. Dynamic spillover analysis of international and Turkish food prices. *International Journal of Finance & Economics*, 2021. DOI: 10.1002/ijfe.2517

FASANYA, I. O. AND OLAWEPO, F. Determinants of food price volatility in Nigeria. *Agricultura Tropica et Subtropica*, 2018, 51(4), pp. 165–174.

GANIOĞLU, A. *Evidence for the explosive behavior of food and energy prices: Implications in terms of inflation expectations.* Working Papers No. 17/2017.

HAMMOUDEH, S., NGUYEN, D. K., SOUSA, R. M. US monetary policy and sectoral commodity prices. *Journal of International Money and Finance*, 2015, 57, pp. 61–85.

HOLTEMÖLLER, O. AND MALLICK, S. Global food prices and monetary policy in an emerging market economy: the case of India. *Journal of Asian Economics*, 2016, 46, pp. 56–70.

IDDRISU, A. A. AND ALAGIDDEDE, I. P. Monetary policy and food inflation in South Africa: a quantile regression analysis. *Food Policy*, 2020, 91, 101816, pp. 1–13.

IDDRISU, A. A. AND ALAGIDEDE, I. P. Asymmetry in food price responses to monetary policy: a quantile regression approach. *SN Business & Economics*, 2021, 1(3), pp. 1–25.

IRZ, X., NIEMI, J., LIU, X. Determinants of food price inflation in Finland – the role of energy. *Energy Policy*, 2013, 63, pp. 656–663.
IŞIK, S. AND ÖZBUĞDAY, F. C. The impact of agricultural input costs on food prices in Turkey: a case study. *Agricultural Economics*, 2021, 67(3), pp. 101–110.
ISMAYA, B. I. AND ANUGRAH, D. F. Determinant of food inflation: the case of Indonesia. *Bulletin of Monetary Economics and Banking*, 2018, 21(1), pp. 1–14.
JOIYA, S. A. AND SHAHZAD, A. A. Determinants of high food prices: the case of Pakistan. *Pakistan Economic and Social Review*, 2013, 51(1), pp. 93–107.
KARGBO, J. M. Impacts of monetary and macroeconomic factors on food prices in eastern and southern Africa. *Applied Economics*, 2000, 32(11), pp. 1373–1389.
KARGBO, J. M. Impacts of monetary and macroeconomic factors on food prices in West Africa. *Agrekon*, 2005, 44(2), pp. 205–224.
NORAZMAN, U. Z., KHALID, H., GHANI, G. M. Food Inflation: a study on key determinants and price transmission processes for Malaysia. *International Journal of Business & Society*, 2018, 19(1), pp. 117–138.
RANGASAMY, L. Food inflation in South Africa: Some implications for economic policy. *South African Journal of Economics*, 2011, 79(2), pp. 184–201.
TÜMTÜRK, O. Monetary Policy Shocks and Empirical Anomalies in Turkey. *Finans Politik ve Ekonomik Yorumlar Dergisi*, 2020, 57(652), pp. 137–164.
WU, X. AND XU, J. Drivers of food price in China: A heterogeneous panel SVAR approach. *Agricultural Economics*, 2021, 52(1), pp. 67–79.

APPENDIX

Table A1 Optimal lag length selection criteria

Lag	Log likelihood	LR	FPE	AIC	SIC	HQ
0	1338.618	NA	1.68e–16	–16.45488	–16.18693	–16.34608
1	1502.263	308.9951	4.06e–17*	–17.87905*	–16.67328*	–17.38946*
2	1540.330	68.56769*	4.67e–17	–17.74323	–15.59965	–16.87285
3	1572.894	55.82412	5.77e–17	–17.53906	–14.45765	–16.28788
4	1599.099	42.64308	7.80e–17	–17.25588	–13.23666	–15.62391
5	1626.422	42.08805	1.05e–16	–16.98661	–12.02956	–14.97385
6	1664.616	55.51198	1.25e–16	–16.85237	–10.95751	–14.45882
7	1702.684	52.01806	1.52e–16	–16.71657	–9.883889	–13.94222
8	1732.203	37.76998	2.09e–16	–16.47457	–8.704073	–13.31943

Source: Author's compilation

Table A2 SVAR model stability condition check

Root	Modulus
0.345211	0.345211
0.163059 – 0.211560i	0.267106
0.163059 + 0.211560i	0.267106
–0.096922	0.096922
0.038205	0.038205

Source: Author's compilation

Table A3 VAR residual serial correlation LM test

Lags	LM-Stat	Prob
1	66.67531	0.1418
2	61.01115	0.1169
3	49.90338	0.4379
4	50.96735	0.3968
5	44.26877	0.6656

Source: Author's compilation

An Overview of Methodological Issues in Data Envelopment Analysis: a Primer for Applied Researchers

Viera Mendelová | *Matej Bel University, Banská Bystrica, Slovakia*
Pavol Kráľ[1] | *Matej Bel University, Banská Bystrica, Slovakia*

Abstract

Data envelopment analysis (DEA) as a method of measuring the efficiency of decision-making units (DMUs) has become an attractive tool used in managerial decision-making in many real-world applications. However, like other sophisticated methods, the application of DEA to a selected decision problem is not straightforward and requires to be carried out in a sequence of successive phases. Moreover, the practical application of DEA presents a range of procedural issues to be examined and resolved. The paper provides a concise guidance through individual steps of DEA analysis process focusing on applied researchers. It includes comprehensive overview of possible issues together with recommendations and hints of possible solutions supported by references to relevant literature presenting more details and alternative viewpoints.

Keywords	***JEL code***
DEA, methodology	*C60*

INTRODUCTION

Data Envelopment Analysis (DEA), initially presented in (Charnes et al., 1978) and built on the earlier work of Farrell (1957), provides a non-parametric linear programming methodology for assessing the relative efficiency of a set of Decision Making Units (DMUs). The last four decades witness the appearance of a set of articles and books on the theory of DEA and its application to various practical settings, e.g., see (Cook and Seiford, 2009). Since 1999 the number of applied papers exceeds the number of those devoted to theory of the method (see Liu et al., 2013). In other words, DEA became a theoretically well-founded methodology with a proven record of successful applications. All that makes it attractive for decision makers as potential addition in their methodology toolbox for supporting decision making in real world applications. However, for someone not specialized in DEA and thus with limited theoretical

[1] Faculty of Economics, Matej Bel University, Tajovského 10, 975 90 Banská Bystrica, Slovakia. Corresponding author: e-mail: pavol.kral@umb.sk, phone: (+421)484466614.

knowledge of the vast DEA methodology, the application of DEA methods is not as straightforward as it seems to be, based on available software tools and seemingly simple requirements put on data. Everyone who intends to use DEA as a part of his/her decision-making toolbox correctly should have some core knowledge of the underlying theory behind various DEA models and/or be able to consult a DEA expert. It mimics the situation in applications of statistics where availability of easy-to-use software enables powerful statistical methods to be misused due to inadequate statistical knowledge of those applying them. In general, an application of DEA to measure the efficiency of DMUs of interest is carried out in a sequence of successive phases (see Golany and Roll, 1989; Ramanathan, 2003; Ozbek et al., 2009): (1) definition and selection of DMUs, (2) definition, selection, and measurement of input and output variables, (3) selection and formulation of the DEA model, (4) application of the DEA model, (5) post-analysis procedures, and (6) presentation and analysis of results.

In each of these phases, there are some key assumptions and procedures to be followed when applying DEA as certain problems may arise and need to be addressed accordingly in order to achieve credible results and avoid "garbage in-garbage out" issues. It is important to help applied researchers in the process of developing necessary skills to detect possible issues when applying the DEA methods and even guide them directly how to proceed through the above-mentioned individual steps and avoid/resolve problems they encounter during the process. In that regard, methodologically oriented papers summarizing important aspects of the methodology and providing directions how to proceed when applying it can play an important role by speeding up the process and make it leaner. The aim of the paper is to provide a concise and comprehensive guidance on how to tackle individual steps of the DEA analysis process, including minimalist and comprehensive overview of possible issues and hints of solutions with references to relevant literature presenting more details how to address them properly. The paper is structured as follows. In Section 1, we discuss how to identify whether DEA is an appropriate method to tackle a research problem of interest. In Section 2, we give an overview how to define and select DMUs and their inputs and outputs. In Section 3, we focus on selection of an appropriate DEA model. In Section 4, we list some software alternatives which can be used to actually perform DEA analyses. In Section 5, we present a compact overview of sensitivity analysis procedures related to the DEA analyses. Then, in Section 6, we show how results of the DEA analyses should be presented and further analysed. Finally, our last section details our conclusions and we discuss possible future lines of research.

1 IS DEA APPLICABLE TO THE PROBLEM AT HAND?

Having in your toolbox a tool like DEA supports "if I have a hammer everything looks like a nail" approach, so it is important to know how to identify if DEA is the way to go when trying to resolve a problem you are facing. As DEA produces non-parametric estimations of a production function, problem we are dealing with, e.g., identification of most efficient companies/branches/departments, identification of possible ways how to improve a relative efficiency of the company/country, allocation of funds etc., should involve a production process. This process could be either a real one where both inputs and outputs as well as mechanism of transformation of inputs to outputs are well known, or an artificial one, where we have just some insights which variables might play the role of inputs and outputs, but we do not have a clear picture of the production process involved. Although the DEA methods might be used in both cases, we should keep it in mind that the nature of the process sets limits for interpretation or application of the results. Results produced by DEA applied when a real well known production process is in place could have more straightforward and clearer interpretation and could be more easily transformable into insights what key sources of identified inefficiencies of DMUs are and what modifications of the production process are needed, i.e., what actions of DMUs are desirable to improve their relative efficiencies. Once we have established that our problem does indeed involve some kind of production process, we identify, guided by a problem we are trying to resolve, all objects of interest such that insight about their relative efficiencies with respect to an identified production process related to a problem at hand could be essential to resolve the problem, i.e., we identify all potential DMUs. With a set of potential DMUs in place, we identify all relevant variables related to the potential DMUs. From the technical point of view, it is often possible to get results from the DEA specialized or general linear programming software when your data consists of several objects you intend to relatively compare to each other (DMUs) and at least two variables measured for each of the DMUs (possible input and output). To ensure that all relevant pieces of information needed to find a solution to a posed problem are in place, it is essential to check here if our data consists of all relevant DMUs and the initial list of variables includes all variables relevant for efficiency evaluation process, i.e., they cover every dimension, the changes in which may affect the DMUs (see Golany and Roll, 1989; Ozbek et al., 2009), or at least it is as wide as possible. Very often we are not at liberty to freely collect all variables of interest with respect to the production process but rather restricted to selection of appropriate input and output variables from variables available in already collected data which may or may not be an appropriate input-output representation of the production process of interest. If the production process takes place, it usually manifests itself by existence of correlations between some of the variables measured on DMUs. If there is a clear picture of the production process involved, presence of significant correlations between variables identified as inputs and variables identified as outputs is expected. At the same time, weak or non-existent correlations are expected to be identified among input and output variables. On the other hand, if the production process is an artificial one, the presence or lack of correlations can be used to guide identification and selection of appropriate inputs and outputs. However, absence or presence of pairwise correlations in data should not be used as a tool to decide if DEA should be applied at problem at hand or not. The role of correlations is more a supportive and an indicative one. When meaningful correlations coherent with our understanding of the production process are present in data, feasible results of application of the DEA methodology, including meaningful interpretation, are more likely.

2 DEFINITION AND SELECTION OF DMUS AND DEFINITION, SELECTION AND MEASUREMENT OF INPUT AND OUTPUT VARIABLES

In the previous step, we did some preliminary evaluation of feasibility of applying DEA to studied problem, created the first iteration of data containing relevant pieces of information and gathered some initial insight if data seems to be promising with respect to DEA. We proceed with refinement of the data based on assumptions imposed on them by DEA methodology. The next two phases, definition and selection of DMUs, and identification and selection of input and output variables are from the practical point of view so intertwined that we will present them simultaneously. On one hand, the number and definition of DMUs determines the set of potential input and outputs, on the other hand, input and output variables represent properties of DMUs some of them being essential to get meaningful results from application of DEA. Therefore, practically these two phases merge in a multistep nonlinear iterative process which, under ideal circumstances, results in the set of DMUs and set of inputs and outputs satisfying various kinds of theoretical assumptions posed by theory of DEA and ready for analysis. In particular, the following assumptions should be met:

(1) the numbers of DMUs and variables are balanced,
(2) variables are clearly separated into inputs and outputs,
(3) DMUs are mutually independent,
(4) DMUs are homogeneous, i.e., comparable.

2.1 Balance between the number of DMUs and variables, reduction of the number of variables

Balanced numbers of DMUs and variables mean that in order to get meaningful results we should apply DEA on data only if the number of DMUs is sufficient, i.e., not too small in comparison to the number of variables (inputs and outputs). The use of too many variables in DEA tends to shift the compared DMUs towards the efficiency frontier which can result in a relatively large number of DMUs with high efficiency scores, and thus, in practice, to reduce, discriminatory power of the DEA. It has been consistently suggested in the DEA literature that there should be a sufficient number of observations (n) in comparison with the numbers of variables, inputs (m) and outputs (s). We recommend using a rule of thumb presented by Cooper et al. (2011) that n should be greater than $\max(m \times s, 3(m + s))$ to determine if the number of DMUs is sufficient. If that check indicates that we need to think about reduction of the number of variables, we can use one of the following main approaches or combination of them:

(1) selection ex-ante only a subset of the original variables,
(2) the approaches based on using the aggregate measures, and
(3) the most recent approaches based on the inclusion of additional cardinality constraints directly into the DEA program in order to select the relevant inputs and outputs automatically.

Approaches from the first group select ex-ante a subset of the original variables either based on heuristic decision or value judgement about which variables are the most relevant for the given problem (see Allen et al., 1997; Golany and Roll, 1989); or on statistical techniques as proposed by Banker (1996), Simar and Wilson (2001), Pastor et al. (2002), Fanchon (2003), Jenkins and Anderson (2003), Ruggiero (2005), Wagner and Shimshak (2007). We prefer here a balanced approach where an expert judgment coming from knowledge about the production process is emphasized and results of statistical analyses play the secondary role. Our recommendation is supported by Harrell (2015) claiming that statistical criteria should not be the only driver determining if variables are or are not included in the models. Application of hypothesis testing could be especially dangerous here as our potential DMUs are usually not selected utilizing statistical sampling techniques and, therefore, applicability of statistical tests in this context is debatable. Moreover, as Peyrache et al. (2020) state, the use of statistical testing brings many additional problems:

(1) the result about which variables are relevant strongly depends on the order of the testing,
(2) we can encounter a severe problem of multitesting when testing for every possible combination between inputs and outputs and thus severely underestimating probability of error in the whole process,
(3) if the number of potential inputs and outputs is very large, the whole procedure of variable selection can become also very tedious and time consuming.

If we identify multiple variables as candidates to be dropped from our data, we can use their measurement character as an additional criterion for final decision. The conventional DEA models assumes that all the data are quantitative variables taking positive or semipositive (i.e., each DMU has at least one positive input and one positive output) real values (see Charnes et al., 1978, 1991). Consequently, if some of the preselected variables are irregular from this point of view, i.e., are qualitative, ordinal or negative, we should prefer their elimination to that of standard quantitative data. The ratio data can too be seen as good candidates for dropping because they violate requirement of convexity (Emrouznejad and Amin, 2009). Approaches from the second group propose to reduce the dimensionality of the problem by substituting some of the original variables by an aggregate measure. This is usually a two-step procedure where in the first step weights of individual variables are determined and in the second step an aggregation technique is selected, and, finally, new aggregated variables created. We face two main problems here. First, there exists a vast number of techniques to determine weights of variables ranging from completely subjective expert selection of ad-hoc weights, through preference-based techniques to some objective techniques utilizing statistical properties of variables to determine the weights, e.g., correlations, variability, information entropy etc. (e.g., Diakoulaki et al., 1995; Wang and Lee, 2009). Regardless the process we used to determine the weights, for the aggregation itself we can use various aggregation functions. We simple have too many options to be able effectively find an optimal combination of weights and aggregation function and to establish arguments defending our selection against the alternatives. Moreover, we do not have a clear picture of what effect various techniques would have on properties of DEA models. A simple workaround for this decision problem is to restrict ourselves to use some well established statistical technique like Principal Component Analysis (PCA) (see Jolliffe and Cadima, 2016) where selection of the variables in based on some statistical criteria, in the case of PCA on preserved variability, and which was extensively studied in the context of its use in DEA modelling. Based on the above, if we see the use of aggregations techniques as necessity, we recommend the most popular approach, DEA-PCA, proposed by Ueda and Hoshiai (1997), and Adler and Golany (2001) which substitutes inputs and outputs with principal components. This technique was studied extensively in comparison to other variable reduction techniques (see Adler and Golany, 2001; Nataraja and Johnson, 2011; Wilson, 2018) and the published results indicate that PCA-DEA consistently provides more accurate results and performs better especially with highly correlated inputs, even for small data sets. More detailed guidelines for the application of PCA-DEA based on the concept of a rule-of-thumb which considers the trade-off between the two types of erroneous classification, namely efficient decision-making units defined as inefficient (under-estimation) and inefficient DMUs defined as efficient (over-estimation) can be found in Adler and Yazhemsky (2010). The second problem related to the use of aggregation techniques to reduce the number of variables is that of interpretability. The new aggregated variable is not generally easily interpretable which could significantly limit our ability to transform results of the DEA model to managerial decisions regarding modifications of inputs and outputs. The third, latest group of approaches is based on the inclusion of additional conditions directly in the DEA model program in order to select the most appropriate inputs and outputs. Majority of these approaches (see Limleamthong and Guillén-Gosálbez, 2018; Benítez-Peña et al., 2020) is based on the introduction of binary (zero-one) variables into the program, so the model is transformed into a Mixed Integer Linear Programming (MILP) formulation. The relevant inputs and outputs are here selected without any previous statistical analysis, heuristic decision making or expert judgement, but the resulting models differ from the most popular models and, consequently, are more demanding on researcher with respect to theoretical knowledge. As these techniques are quite recent, the corresponding models might not be yet implemented in DEA software tools preferable by the researcher.

2.2 Classification of variables into inputs and outputs

Selection of the appropriate number of variables relevant for efficiency evaluation process of DMUs is closely related to classification of these variables into inputs and outputs. If a well-known real production process is in place, classification should be more or less straightforward. Resources utilized by the DMUs or conditions affecting their operation constitute typical inputs, while measurable benefits generated by the DMUs represent the outputs. However, in some cases, especially if the underlying production process is not fully understood or it is an artificial one, some variables may be interpreted in both ways and play either input or output roles. These variables are usually referred to as flexible measures (Cook and Zhu, 2007) or dual-role factors (Cook and Zhu, 2006). Suppose, for example, that one wishes to evaluate the efficiency of bank branch operations. In this case, a factor such as deposits, could serve as either an input or an output. In some of the previous studies devoted to efficiency of banks, such a factor is regarded as an output, as it is a source of revenue for the branch. At the same time, however, it can be argued that the deposits, in particular, the time spent by employees in processing of customers who are making deposits or opening deposit accounts, is an input since it could be used to better advantage to sell more profitable products to the customer. For more examples of flexible measures see e.g. (Cook and Zhu, 2007). The DEA literature offers several methodological suggestions for deciding the status of such flexible measures. Golany and Roll (1989) propose to carry out a series of regression analyses, of such factors, one at a time, on the factors known to be inputs and outputs. A weak relation to inputs and strong relation to outputs indicates a preference towards classifying the factor as an input, and vice versa a weak relation to all the factors may indicate a need to re-examine the factor and possibly remove it from the analysis. Alternatively, strong relationships may indicate that the information contained in that factor is already represented by other factors and, again, its removal should be considered. Golany and Roll (1989) also pointed out one of the basic assumptions for the DEA models, the condition of "isotonicity", i.e., that an increase in any input should not result in a decrease in any output, may also be helpful. Besides statistical properties of flexible measures, additional information might be used to determine their appropriate status. Bala and Cook (2003) investigated the situation where bank branch consultants provided additional "classification data" specifying quality of each branch (good versus poor branch). Their intention was to assign a status to each flexible measure such as to provide efficiency scores that are in the best agreement with expert opinion. Many other approaches how to deal with the flexible measures were published (see e.g. Cook and Zhu, 2007; Amirteimoori and Emrouznejad, 2011; Amirteimoori et al., 2013; Toloo et al., 2018; Kordrostami et al., 2019; Sedighi Hassan Kiyadeh et al., 2019; Boďa, 2020). Detailed description of these approaches is beyond the scope of the paper. Our recommendation is to start with the classical approach where the status of a flexible measure is determined by its statistical properties reinforced by available additional information and move towards more elaborate techniques only if the results of the modelling are not satisfactory to solve the problem.

2.3 Independence of DMUs

It is commonly assumed that the DMUs are independent of one another, i.e., each DMU has its own values of inputs and outputs not connected in any way to values of the inputs and outputs of other DMUs. In other words, if the assumption of independence of DMUs holds, decreasing the inputs or increasing the outputs of one DMU will not affect the inputs or outputs of other DMUs. However, this natural assumption is not always fulfilled in real conditions. The condition of mutual independence of DMUs is not fully met in the following two cases. First, if the DMUs have some input in common. The efficiency measurement difficulty created by this "shared" resource phenomenon is that in attempting to move an inefficient DMU towards the frontier by reducing that shared resource, the other DMUs in the same set of DMUs will be equally penalised. Second, if some internal or linking activities between DMUs are present. For example, many companies are comprised of several divisions that are linked by some intermediate products or activities. Some DMUs in this multi-stage structure consume resources produced by other DMUs and some of which produce resources consumed by other DMUs. Another example is the supply chain system where supply chain members (suppliers, manufacturers, distributors, and retailers) are interconnected through intermediates, i.e., some measures linked to related supply chain members that cannot be simply classified as outputs or inputs of the supply chain. For example, the supplier's revenue is an output for the supplier, and it is in the supplier's interest to maximize it, and, at the same time, it is an input to the manufacturer who wishes to minimize it. These possible conflicts between supply chain members may result in that one member's inefficiency may be caused by another's efficient operations. Measuring efficiency of the interconnected DMUs becomes a difficult and challenging task because of the need to deal with the multiple intermediate products, and to integrate and coordinate the efficiency of those DMUs. If, based on the presented examples, we identify interdependencies among DMUs, we must address it with selection of an appropriate DEA model as the conventional DEA models will fail to address this important issue. Within the context of DEA, several methods have the potential to be used in these situations. To capture interdependence caused by shared inputs, Avilés-Sacoto et al. (2019) developed a new DEA-like methodology and illustrated it using the problem of evaluating a set of departments in a university setting, where the departments are grouped under various faculties. To deal with interdependence caused by linking of DMUs, Zhu (1996, 2009) presented a DEA-based supply chain model to both define and measure the efficiency of a supply chain and its members and yield a set of optimal values of the intermediate performance measures that establish an efficient supply chain. Liang et al. (2006) developed two types of DEA-based models for supply chain efficiency evaluation, using a seller-buyer supply chain as an example.

2.4 Homogeneity of DMUs

A homogenous set of DMUs is a key assumption within DEA and can be obtained by considering the following criteria (see Golany and Roll, 1989; Dyson et al., 2001; Ozbek et al., 2009):

(1) DMUs should perform the same activities and undertake the same processes with similar objectives,

(2) the input-output variables characterizing the activity of DMUs in the data set should be identical except for the differences in their intensity or magnitude, and

(3) DMUs should operate under the same market and environmental conditions.

While the first two criteria represent the criteria of homogeneity of the DMUs themselves, the last criterion is the criterion of the homogeneous environment of DMUs. However, in practice, homogeneity of DMUs or environment is seldom present. In the cases when the condition of homogeneity across all of the DMUs is not met, it is necessary to use one of the special techniques proposed for the given cases as application of the conventional DEA models may result in efficiency scores reflecting the underlying differences in the conditions of DMUs rather than any inefficiencies. Examples of applied problems where some of the above-mentioned conditions are violated are the following:

- When efficiency of bank branches is examined, differences in activities may occur as large branches usually carry out most banking activities, whilst smaller branches may only engage in some of them.
- Differences in input-output variables may be encountered when assessing the efficiency of faculties at one university. Usually, not all faculties shared the same inputs, i.e., the science faculties required laboratories and equipment, while the humanities faculty did not.
- Non-homogeneous environment can be illustrated on efficiency comparison of individual urban public transport (UPT) lines in a city. Efficiency of UPT lines is highly dependent on the conditions in which transport takes place, i.e., on environmental factors. For example, lines running in the city centre are exposed to a high level of competition because many lines connecting the city centre share the same routes and it is not important for passengers which particular line they choose, in effect of which the numbers of passengers are highly influenced by the possibility of substitution lines.

In general, environmental factors express the influence of the environment on a DMU and characterize the environment and its characteristics in which the DMU operates. These factors are not traditional inputs and are assumed not being under the control of a manager. Environmental factors include (i) external factors such as the current political and macroeconomic situation, the purchasing power of the population, the demographic structure of the population or other location characteristics, or (ii) internal factors such as the ownership differences, e.g., public/private or corporate/non-corporate, and organizational structure differences of DMUs.

2.4.1 Non-homogeneity induced by differences in activities, inputs and outputs

When dealing with non-homogeneity coming from (1) and (2), there are two basic options on how to proceed, either (i) restrict the analysis to a limited number of activities / inputs and outputs shared by all DMUs; or (ii) to cluster DMUs into homogeneous groups and thus study only a limited number of DMUs engaged in identical activities/having the same inputs and outputs (see Athanassopoulos and Thanassoulis, 1995; Molinero, 2008). Unfortunately, these two basic options are in many cases neither possible (e.g., due to unavailable data or due to the low number of DMUs in relation to the number of inputs and outputs) nor desirable (if we wish to evaluate the overall relative efficiency capturing all activities of DMUs in the whole group of evaluated DMUs). In cases when these basic options are not applicable, one of the more advanced approaches must be applied. If non-homogeneity of DMUs is caused by non-homogeneous output sets and we can assume that the input set was common across all DMUs, the approach of Cook et al. (2012, 2013) can be followed. Cook et al. (2013), by extending the earlier research of Cook et al. (2012), developed a general setting encompassing the lack of homogeneity on the output side where some DMUs may produce a certain set of products, but not all products are produced in all DMUs. If non-homogeneity of DMUs is caused by non-homogeneous input sets, the approach of Li et al. (2016) extending earlier research of Cook et al. (2012, 2013) can be employed. Li et al. (2016) examined a three-step procedure to evaluate efficiencies of DMUs in the presence of non-homogeneous input sets. In each DMU, the process of inputs generating outputs is divided into the separate processes. In the first step, for each DMU in each process (in which that DMU is involved), an appropriate split of the inputs and outputs is determined. In the second step, the process scores for each DMU are computed, and in the third step, the aggregate scores computed as a weighted average of the process scores for any given DMU are calculated. In the most general case, two types of non-homogeneity mentioned above are present simultaneously, i.e. some DMUs do not carry out the same activities as others, and they do not share the same set of inputs/outputs with the remaining DMUs, and the model proposed by Molinero (2008) could be applied. The proposed model is presented on the example of three types of university institutions: those, such as standard universities, that engage in both teaching (T) and research (R) activities; those that engage in the T activity but not in the R activity; and those, such as research institutes, that engage in the R activity but not in the T activity. Under the proposed joint efficiency model, it is assumed that some inputs/outputs are related only with the T activity, some inputs/outputs are shared between/reflect the effort devoted to the T and R activities, and some inputs are allocated/depend only to the R activities. The DMU under observation must decide how to allocate shared inputs to the R or to the T activities, and how much effort should be devoted to produce outputs from the T or the R activities. It does that by taking into account the importance attached to the T and R activities (captured by the weights determined outside the model and reflecting the priorities of the decision maker), and the desire to be operating as efficiently as possible under both activities when compared with other DMUs. The rationale of the joint efficiency DEA algorithm is based on the same philosophy as the standard DEA model in the ratio formulation: once the DMU under observation has decided how to allocate shared inputs, and how to attribute shared outputs, this split is applied to all other DMUs and efficiency calculations take place as usual.

2.4.2 Non-homogeneity induced by non-homogeneous environment

There are numerous ways in which non-homogeneous environment can be accommodated in a DEA analysis. They differ mainly based on permissible character of environmental factors, i.e., ordinal, categorical or continuous. There are four traditional approaches (Coelli et al., 2005) to tackle this issue. If the values of the environmental factor can be ordered from the least to the most detrimental effect upon efficiency, then the first approach of Banker and Morey (1986a) can be followed. In this approach, the DMU is compared with those DMUs in the sample that have a value of the environmental factor which is less than or equal to that of the given DMU. This would ensure that no DMU is compared with another DMU that has a more favourable environment. Example of application of this approach can be found in (Roháčová, 2015). The second approach proposed by Charnes et al. (1981) can be employed in the cases when the environmental factor is of a categorical character (e.g., public versus private ownership). This approach involves three stages: (i) divide the DMUs into relatively homogeneous sub-groups and solve a DEA model for each sub-group, (ii) project all observed data points onto their respective frontiers, and (iii) solve a single DEA model using the projected points and assess any difference in the mean efficiency of the two sub-groups. Example of application of this approach can be found in (Soteriou and Zenios, 1999; Fandel et al., 2019). Note that there are two main limitations of the previous two approaches: (i) a high number of DMUs is necessary (subdivision of DMUs into sub-groups may significantly reduce the number of DMUs compared, resulting in many DMUs being found to be efficient and thus reducing the discriminatory power of the analysis), and (ii) the possibility of considering only one environmental factor. The second approach has the additional limitation that it requires that the environmental variable be a categorical variable, while the first approach suffers from the problem that it requires that the direction of the influence of the environmental variable (upon efficiency) be known a priori. The third possible approach is to include the environmental factor(s) directly into the optimization task of a DEA model, either in the form of a non-controllable input variable, a non-controllable output variable, or a non-controllable neutral variable (a non-discretionary variable), (see e.g., Coelli et al., 2005; Cooper et al., 2007). In this approach, it is necessary to determine the direction of influence of the environmental factor first, i.e., whether higher values of the environmental factor will contribute to improving or deteriorating efficiency. If the environmental factor is likely to have a favourable (detrimental) effect upon efficiency, then the environmental factor can be included in the optimization task of a DEA model in the form of non-controllable input (output) variable. In this way, the DMU is compared with a theoretical DMU that has an environment that is no better than that of the given DMU. If the direction of influence of the environmental factor is uncertain, then it can be included in the DEA model as a non-controllable neutral variable. This will ensure that the assessed DMU is only compared with a (theoretical) frontier DMU that has the same environment (neither better nor worse). Although this approach does not require a predetermined direction of the environmental variable influence, it can significantly reduce the reference set for each DMU and hence inflate the obtained efficiency scores. Finally, it should be noted that this third approach also has the disadvantage that the environmental factors have to be continuous variables (i.e., they cannot be categorical or ordinal variables). If there are categorical variables considered as environmental factors, then the more complicated mixed-integer linear programming models, suggested by Banker and Morey (1986b), can be used. For some extensions of this model, the reader is referred to Kamakura (1988), and Rousseau and Semple (1993). The last, and relatively most widely used, approach to include environmental factors consists of a two-stage procedure. In the first stage, the efficiency of the DMUs is measured without including environmental factors, e.g., using traditional inputs and outputs, and then in the second stage, the efficiency scores from the first stage are regressed upon the environmental factors. The environmental factors can be of both continuous and categorical nature, i.e., the effects upon efficiency of the age, experience, education and training of the manager(s) can be estimated. Using this approach, it is possible to examine the direction and intensity of the influence, as well as to correct the efficiency scores for environmental factors by using the estimated regression coefficients to adjust all efficiency scores to correspond to a common level of environment (e.g., the sample means). The two-stage procedure approach has several significant advantages over previous approaches, e.g., it can accommodate more than one variable, it works for both continuous and categorical variables, it does not make prior assumptions regarding the direction of the influence of the environmental factor and one can determine significance and effect size of factor influence upon efficiencies. Moreover, it is easy to calculate, and the approach is simple and therefore transparent. See e.g. (Ray, 1991; Fizel and Nunnikhoven, 1992, 1993; Oum and Yu, 1994; Sexton et al., 1994; Nolan, 1996; Mancebon and Molinero, 2000; Mendelová and Kanderová, 2016) for application of this approach and (Haas and Murphy, 2003; Banker and Natarajan, 2008; Simar and Wilson, 2007, 2011 and Banker et al., 2019) for comparisons and some extensions of this approach.

3 SELECTION AND FORMULATION OF THE DEA MODEL

The right choice of the DEA model used for tackling our problem is a very important and definitely non-trivial issue. To date, a huge variety of the DEA models have been developed so it would be unattainable to carry out systemisation of all of them including some guidelines when they should be selected in a single paper. For this reason, in this section the main attention will be paid to the basic considerations that could guide the analyst in the right direction in choosing an appropriate DEA model in relation to the problem being solved and various selected DEA model alternatives connected to these directions will be presented. In the process of selecting a suitable DEA model, the critical areas are in particular:

(1) the choice of the DEA model orientation,
(2) the choice of a radial or a non-radial approach,
(3) the choice of a returns-to-scale assumption, and
(4) violation of basic assumptions, i.e., non-homogeneity and/or interdependence of DMUs, presence of data irregularities.

3.1 DEA model orientation

The following considerations (see Golany and Roll, 1989; Ramanathan, 2003) may be helpful in selecting the DEA model orientation. In applications where inputs are rather inflexible (e.g., determined to a certain extend by higher managerial levels and thus not fully under control of DMUs), output-oriented DEA model would be more appropriate. On the other hand, in applications that involve inflexible outputs (e.g., matched closely with goals set by management or restricted by environmental conditions), input-oriented DEA model may be more appropriate.

3.2 Radial vs non-radial models

There are two types of approaches in DEA: radial and non-radial. The radial approach, represented by the CCR (Charnes et al., 1978) and BCC models (Banker et al., 1984), neglects the non-radial input/output slacks and provides a radial measure of efficiency. This measure is referred as Farrell efficiency (for the CCR model) or pure technical efficiency (for the BCC model) and, in the case of input-oriented models, represents the maximum equiproportionate, i.e., radial reduction in all inputs that is feasible with given technology and outputs. On the other hand, the non-radial approach, represented by the additive or the SBM models (Tone, 2001), deals with slacks directly and measures a non-radial efficiency that is referred to as Pareto-Koopmans efficiency or strong efficiency. Both approaches have some benefits but also drawbacks. While, the radial approaches neglect the non-radial input/output slacks, the non-radial approaches neglect the radial characteristics of inputs and/or outputs. It is therefore necessary to consider in detail which of these approaches to use in relation to the specifics of the problem addressed. The final selection should be based primarily on the consideration of potential differences in the characterization of the inputs and outputs. The general rule can be formulated as follows. If the non-radial slacks have an important role in evaluating managerial efficiency, the non-radial approaches should be preferred. And, on the other hand, if the loss of the original proportionality of inputs and/or outputs is inappropriate for the analysis, the radial approaches are more suitable. In other words, if all the inputs and/or outputs (depending on the model orientation) are non-radial (substitutional), i.e., they do not have to change proportionally, the non-radial approach should be selected. And, if, all the inputs and/or outputs are radial, i.e., they have to change proportionally, the radial approach should be preferred.

3.3 Returns-to-scale assumption

One of the most critical problems for setting up a DEA model is the identification of suitable Returns To Scale (RTS) for the data. In principle, either the Constant Returns to Scale (CRS), Increasing (non-decreasing) Returns to Scale (IRS), Decreasing (non-increasing) Returns to Scale (DRS) or Variable Returns to Scale (VRS) can be assumed. The identification of the RTS is one of the most discussed areas in DEA. At the outset, it should be emphasized that it is done at two levels: at the technology that is referred as the Technological Returns to Scale (TRTS), and at the DMU level. For identifying the RTS at the DMU level the basic approaches were developed by Färe et al. (1983), Banker et al. (1984), and Banker and Thrall (1992). In the process of selecting the appropriate DEA model for analysis, the proper identification of the TRTS is crucial. Unlike the RTS at the DMU level, the RTS at the technology level has been investigated by only a few researchers. In the DEA literature, there are two main approaches for the determination of the TRTS, subjective, where the TRTS is determined by expert opinions and, objective, where the TRTS is identified through a mathematical model. In the subjective approach, an analyst has to take into consideration that the DEA models under CRS assumption do not allow for the potential existence of economies or diseconomies of scale. Thus, when the performances of DMUs are not normally expected to depend on the scale of operation (e.g., comparisons several large monopolies), the assumption of CRS seem appropriate. However, a unit may be too small to operate with optimal efficiency or so large that it becomes difficult to manage. The DEA models under VRS assumption have been developed specifically to accommodate scale effects in analysis. It is well known, that the VRS model will always envelop the data more closely than the CRS model, irrespective of whether VRS exist. If the VRS model is used, where there are no inherent scale effects, the efficiency of small and large units will tend to be over-rated. Alternatively, if the CRS model is used, where there are evident inherent scale effects, the small and large units can be greatly discriminated. It is therefore advisable, when it is not known a priori if the production technology exhibits CRS or VRS, to use some of the objective approaches that test the data separately for the scale effects. The TRTS investigation is still an active area of scientific interest. The objective approaches for the identification of the TRTS are based on either the statistical methods (see Banker, 1993, 1996; Read and Thanassoulis, 2000; Simar and Wilson, 2002; Banker and Natarajan, 2011), or the non-statistical methods (Alirezaee et al., 2018). Banker (1993, 1996), and Banker and Natarajan (2011) explored the statistical properties of the production frontiers generated by DEA models and developed DEA-based hypothesis tests for addressing a wide range of issues including the TRTS. Read and Thanassoulis (2000), inspired by the approach of Banker (1996), suggested a measure of cross-mix scale size for single-output, multiple-input cases where the production function can be approximated by a homothetic function. Their method can be used to help to decide which of the DEA models to use even in small data sets. Simar and Wilson (2002) used non-parametric statistical tests for recognizing the TRTS. However, since applying a statistical method usually requires some background assumptions, Alirezaee et al. (2018), to overcome this deficiency of the statistical methods, focused on identifying the TRTS using data mining. They designed a novel method, the Angles method, for mining the dataset and discovering its TRTS characteristics.

3.4 Model selection if basic assumptions are violated

In this section, we assume that we determined the final set of DMUs and their corresponding inputs and outputs. Further, we assume that one of the following violation of basic assumptions occurs:

(1) non-homogeneity and/or interdependence of DMUs;

(2) presence of data irregularities: ordinal variables, negative data, ratio data.

The issues of non-homogeneity and interdependence of DMUs were already discussed in Sections 2.3 and 2.4 and some recommendations regarding suitable models were presented. Therefore, we restrict ourselves here to the case when non-homogeneity and interdependence occurs simultaneously. Castelli et al. (2001) dealt with nonhomogeneous subunits by identifying sensibly compared groups of them evaluated through three new-introduced efficiency concepts. Imanirad et al. (2013) extended the conventional DEA models to situations where only partial input-to-output impacts exist. They considered the DMU as a business unit consisting of a set of mutually exclusive subunits, each of which can be evaluated in the conventional DEA sense. Du et al. (2015) pointed out two main deficiencies of the Imanirad et al. (2013) approach: splitting inputs may not be applicable when non-separable inputs exist, and the intermediate measures that link subunits are not considered. Du et al. (2015) expanded the presence of non-homogeneity to a network setting and proposed DEA models that tackle the problem of the parallel network in non-homogeneous situations, where subunits operate with intermediate products but are not restricted to identical input/output variables. Barat et al. (2019) followed the approach of Du et al. (2015) and generalised the idea of nonhomogeneity in parallel network structures to the case where DMUs have various subunits. In the presence of ordinal data, we need to resolve the issue that direct incorporation of ordinal data into the standard linear DEA model leads to DEA model which is a non-linear and non-convex program. Such a DEA model is called Imprecise DEA (IDEA) in the literature (Cooper et al., 1999) as in DEA terminology ordinal data, together with interval or bounded data, and ratio bounded data, is referred to as imprecise data. The modified DEA structure incorporating ordinal data was first presented in Cook et al. (1993, 1996). In general, there are two approaches how to deal with ordinal variables in DEA evaluation. One uses scale transformations and variable alternations to convert the non-linear IDEA model into a linear program (Cooper et al., 1999; Park, 2007). The other identifies a set of exact data from the ordinal (imprecise) inputs and outputs and then uses the standard linear DEA model (Zhu, 2003a, 2004; Chen, 2007). We recommend using the second approach. A comprehensive survey of the IDEA approach that uses the standard DEA model is presented in Chen (2007). In the presence of negative data, there are three ways how to proceed:

(1) to interchange roles between inputs and outputs,

(2) to transform the negative data to be positive and then use a translation invariant DEA model,

(3) to use the modified DEA models capable of handling negative data without data translation.

In the input-output exchange approach suggested by Scheel (2001), the absolute values of negative outputs are treated as inputs and the absolute values of negative inputs are treated as outputs. The rationale is that more negative values, as an output, indicate worse performance, while larger positive values, as an input, also indicate worse performance. It is a very simplistic approach not having influence on selection of DEA model, but as pointed out by Kao (2017), this approach has also its drawbacks, namely:

(1) restriction of this approach to cases when the values of all DMUs are negative and

(2) the fact that it does not reflect the true production process.

We do not recommend the second approach, based on transformation of positive values into negative, as it poses a serious restriction on models which can be used. Essentially, we are restricted to the use of the additive model or the BCC model. However, both models have their drawbacks. The additive model provides the "furthest" target on the production frontier for inefficient DMUs (Cheng et al., 2013) and cannot provide any measure of efficiency (Portela et al., 2004); the BCC model is just restricted translation invariant. In the third approach, the use of modified DEA models, we recommend one of the following models. Cooper et al. (1999), based on the additive model, developed the Range Adjusted Measure (RAM). Portela et al. (2004), to deal with negative data, utilized the directional distance function and develop two variants of a Range Directional Measure (RDM), labelled RDM+ and RDM−, respectively. The proposed RDM model provides an efficiency score that results from the comparison of the unit under assessment with the so-called ideal point by using the corresponding directional distance function. Through combining the RDM+ model with the SBM model, Sharp et al. (2007) proposed a Modified SBM (MSBM) model where both negative outputs and negative inputs could be handled. Tone and Tsutsui (2009) propose the approach how to deal with non-positive data in the SBM and Network DEA model. Emrouznejad et al. (2010) introduced a Semi-Oriented Radial Measure (SORM) model by breaking down each variable into two non-negative variables. Kazemi Matin and Azizi (2011) proposed a new additive based approach to provide a target with non-negative value associated with negative components for each observed unit. Kerstens and Van de Woestyne (2011) recommended a generalized Farrell proportional distance function that handles negative data and maintains a proportional interpretation under mild conditions. Cheng et al. (2013) developed a variant of the traditional radial model, where original values of inputs or outputs are replaced with their absolute values as the basis to quantify the proportion of improvements to reach the efficiency frontier. Some imprecisions of the Cheng et al. (2013) model were corrected by Kerstens and Van de Woestyne (2014). Finally, it should be added that if the researcher is interested in finding out the full ranking of the evaluated units, i.e., to further discriminate efficient units, it is possible to apply one of the most recent approaches that address the issue of negative data for super efficiency evaluation. Hadi-Vencheh and Esmaeilzadeh (2013), Lin and Chen (2017), Wei et al. (2019), and Lin et al. (2019) proposed alternative super-efficiency DEA models, which can fully rank all DMUs, including the efficient DMUs, and can deal with negative data. In the presence of ratios in our variables, standard DEA models methods cannot be directly used as it may lead to incorrect results caused by violation of convexity (see Bogetoft, 1996; Bogetoft et al., 2000; Cooper et al., 2007; Emrouznejad and Amin, 2009; Olesen and Petersen, 2009; Cook and Zhu, 2014; Olesen et al., 2015). This problem can be resolved if we know the original numerator and denominator by placing them accordingly as additional inputs and outputs in the model (Emrouznejad and Amin, 2009). However, if there are too many ratio variables present in the data this might not be feasible as it can lead to the significantly increased number of variables and thus destroy the balance established between the number of DMUs and the number of variables. Alternatively, the ratio variable could be replaced by volume variable (Thanassoulis et al., 1995, 2008; Dyson et al., 2001). An alternative approach can be found in (Hatami-Marbini and Toloo, 2019).

4 APPLICATION OF DEA MODELS

There are relatively many software packages on the market that allow straightforward calculations of DEA models without any particular knowledge of DEA methodologies. Some commercial specialized DEA programs are presented in Table 1 and Table 2 provides a list of some non-commercial alternatives.

Since DEA is a method based on solving linear and non-linear optimization tasks, besides specialized DEA software or macros, we can use general programming languages widely adopted for scientific computing and providing means for solving both linear and non-linear optimization problems.

Table 1 Commercial DEA software packages

Software	Authors/developers	Available from
DEA-Solver-Pro™	K. Tone (Saitech, Inc.)	http://www.saitech-inc.com/Products/Prod-DSP.asp
Frontier Analyst®	Banxia Software Ltd.	https://banxia.com/frontier
PIM-DEAsoft	A. Emrouznejad, E. Thanassoulis	http://www.deasoftware.co.uk
xlDEA	Productivity Tools	https://xldea.windows10compatible.com
LIMDEP	Econometric Software, Inc.	http://www.limdep.com
DEAFrontierTM	J. Zhu	http://www.deafrontier.net
DEA Online Software (DEAOS)	DssBridge Decision Group Inc.	https://www.deaos.com
MaxDEA Ultra	Beijing Realworld Software Company Ltd.	http://maxdea.com/MaxDEA.htm

Source: Own construction

Table 2 Non-commercial DEA software packages

Software	Authors/developers	Available from
DEAP	T. Coelli	https://economics.uq.edu.au/cepa/software
EMS: Efficiency Measurement System	H. Scheel	http://www.holger-scheel.de/ems
OSDEA	H. Virtos	https://opensourcedea.org
MaxDEA Basic	Beijing Realworld Software Company Ltd.	http://maxdea.com/MaxDEA.htm
DEAOS Lite	DssBridge Decision Group Inc.	https://www.deaos.com
DEAFrontierTM Free Version	J. Zhu	http://www.deafrontier.net/deafree.html

Table 2 (continuation)

Software	Authors/developers	Available from
R packages		
Benchmarking	P. Bogetoft, L. Otto	https://CRAN.R-project.org/package=Benchmarking
nonparaeff	D. Oh, D. Suh	https://CRAN.R-project.org/package=nonparaeff
FEAR	P. W. Wilson	http://media.clemson.edu/economics/faculty/wilson/Software/FEAR/FEAR-1.15/FEAR-manual.pdf
npsf	O. Badunenko et al.	https://CRAN.R-project.org/package=npsf
MultiplierDEA	A. K. Puthanpura	https://CRAN.R-project.org/package=MultiplierDEA
Python Library pyDEA	A. Raith et al.	https://pypi.org/project/pyDEA

Source: Own construction

This option is of particular importance if the DEA model contains several special settings (e.g., extending the model to include uncontrollable variables, non-discretionary variables, bounded variables, undesirable variables, etc.) and if we plan to adopt more advanced sensitivity analyses, post-hoc statistical analyses and visualisations. The drawback is a steeper learning curve. Alternatively, the versatile Microsoft Excel spreadsheet editor can also be used using the Solver Add-in. For instructions on how the Excel Solver can be used to solve various DEA models see e.g. (Zhu, 2003b).

5 POST-ANALYSIS PROCEDURES (SENSITIVITY ANALYSIS OF DEA RESULTS)

The results of DEA as a non-parametric and non-statistical method of measuring efficiency are generally relatively sensitive to both the selection of the compared DMUs and the selection of inputs and outputs considered. There is no guarantee that the initial selection of such DMUs and inputs and outputs are correct and that this serves the best purpose of the analysis. DEA is an extreme point technique where the efficiency frontier is formed by the actual performance of the best-performing DMUs. A direct consequence of this aspect is that errors in measurement can affect DEA results significantly. Hence, it is expected that the robustness of the DEA results be verified using some form of sensitivity analysis. This problem can be mitigated also by using a specially developed robust DEA models, e.g., a model presented in (Hladík, 2019).

5.1 Basic sensitivity analysis approaches

The sensitivity (stability or robustness) analysis has been one of the widely studied topics in the DEA literature. Some basic sensitivity analysis procedures (Smith and Mayston, 1987; Ahn and Seiford, 1993; Ramanathan, 2003) are:

(1) peer analysis,
(2) analysis based on diminution and augmentation in the number of inputs and outputs,
(3) analysis based on diminution and augmentation in the number of DMUs, and
(4) analysis based on the use of different DEA models.

Smith and Mayston (1987) pointed out that according to the DEA technique, it is possible for a DMU to become efficient if it achieves remarkably better results in terms of one output but performs below average in terms of other outputs. These kinds of efficient DMUs can be tested by identifying the peers for inefficient DMUs. Thus, if a DMU is initially identified as efficient by DEA, a supplementary sensitivity analysis should be conducted by checking the number of inefficient DMUs for which it is a peer. If the number is high, then the DMU is genuinely efficient. However, if the number is low, then its best performance is questionable and other evidence for establishing the superiority of its performance is necessary. An alternative way of testing the robustness of the DEA results is to verify whether the efficiency score of a DMU is affected appreciably if either only one input or output is omitted from the DEA analysis (Smith and Mayston, 1987), or different input/output combinations are considered in the analysis (Banker, 1996). An efficient DMU that is ranked inefficient due to these input/output changes should be viewed with caution. Another part of the literature studies responses with given data when DMUs are deleted or added to the set being considered. Smith and Mayston (1987) and Ramanathan (2003) proposed a sensitivity analysis based on omitting efficient DMUs from the analysis. The purpose of this analysis is to assess the sensitivity of the DEA results for this change. Another part of the literature extends sensitivity analyses to eliminating some DMUs entirely and/or introducing additional DMUs. Wilson (1995) studied the effects of removing DMUs in order to determine "influential observations" (i.e., DMUs). A method based on a similar basis is "window analysis", originally proposed by Klopp (1985) for studying trends. Here, however, it is also treated as a method for studying the stability of DEA results because such window analyses involve the removal of entire sets of observations and their replacement by other previously not considered observations. Another topic of interest revolves around effects that might be associated with choices of different DEA models. Ahn and Seiford (1993) focused on whether different DEA models would result in different conclusions in a hypothesis test setting. Nevertheless, there are systematic differences between the different DEA models because of the differences in efficiency characterizations, the models should generate the same hypothesis test results. If DEA results are proved to be sensitive to the model selected for an analysis, the credibility of the results obtained would be seriously weakened.

5.2 Advanced sensitivity analysis approaches

Building on previous basic approaches to assessing the credibility of DEA results, analytically formulated (mathematical) methods for examining stability and sensitivity of results to data variations with given variables and given models have been developed in recent years. These methodological approaches to sensitivity analysis are associated with the development of tools and concepts that can be used to determine the degree of sensitivity to data variations in any particular application of DEA. There has been a great progress in the methodological approaches. The early work in this area, which focused on a single input or output analyses for a single DMU, has now moved to sensitivity analyses aimed at the evaluation of the stability of DEA results when all inputs and outputs are varied simultaneously in all DMUs. According Cooper et al. (2001, 2011), we can distinguish:

(1) algorithmic approach,
(2) metric approach,
(3) multiplier model approach,
(4) a two-stage alternative,
(5) envelopment approach.

Algorithmic approaches are extensions of algorithmic approaches used in linear programming and they are discussed by Charnes and Neralic (1992), and Neralic (1997). The metric approaches, see Charned et al. (1992), use the concept of distance to determine for a DMU its radius of stability, i.e., the range of data variations not causing alteration in DMU being efficient or vice versa. Both algorithmic and metric approaches, work with one DMU at a time. This drawback can be overcome by using a multiplier model approach of Thompson et al. (1994, 1996) where a special multiplier model is present solution of which provides as with information about stability of all DMUs simultaneously. Unfortunately, it involves the use of special algorithms like the interior point methods developed in Thompson et al. (1996). To avoid such difficulties Cooper et al. (2011) proposed the use of a two-stage alternative, where special algorithms as the one mentioned above are used only if standard methods fail. As an alternative to multiplier model approach, Seiford and Zhu (1998), and Zhu (2009) have developed the envelopment approach, where envelopment boundaries are determined for situations where (unequal) changes in all DMUs (both efficient and inefficient DMUs) are considered. Alternatively, we can focus on analysis of inefficient DMUs and define, for each inefficient DMU, so called "NecessaryChangeRegion" (Jahanshahloo et al., 2011) which encompasses ways how this DMU can reach some predetermined level $\alpha < 1$ of efficiency. Besides techniques developed specifically for DEA, we can use also bootstrap to estimate variance of efficiency and determine the related confidence intervals (see Bogetoft and Otto, 2011).

6 PRESENTATION AND ANALYSIS OF RESULTS

When discussing presentation and analysis of the results coming from applications of DEA, two main questions need to be answered. First, what should be presented and/or analysed and second, how it should be presented and/or analysed. We are well aware that detailed answers to these questions depend on the intended audience and character of the problem we are dealing with. Therefore, we restrict ourselves here to some general recommendations.

The direct results of DEA models always need to be presented in such a way that they can be used as a basis for decision making in an economic environment. In the presentation, all previous phases of DEA modelling should be sufficiently covered. The key pieces of information which should be shared with our audience are the following:

(1) goal of the analyses, i.e., the problem we are trying to solve or get insight into,
(2) description of production process, definition and properties of DMUs, selected inputs and output, their properties and their connection to the problem at hand,
(3) what model were used, assumptions of these models and why we opted for the selected models,
(4) software tools used for analysis,
(5) identification of efficient and inefficient DMUs, and pears for inefficient DMUs,
(6) efficiency scores and the impact of efficient DMUs on those identified as inefficient,
(7) results of sensitivity analyses and post-hoc statistical analyses,
(8) recommendation for DMUs regarding ways how to improve their performance.

Points (1)–(4) mean presentation of crucial parts of phases discussed in Sections 1–4 of the paper. From the point of view of presentation these are not particularly demanding and could be satisfactory covered by standard tabulations and series of simple graphs like e.g., boxplots, scatterplots and web charts. If dimension reduction technique, like PCA, was used when identifying appropriate inputs and outputs of analyses, barplots and biplots related to this analysis could be beneficial. The most appropriate approach to points (5)–(6) depends on the number of DMUs. When the analysed dataset comprises of small amount of DMUs and, consequently, of quite few inputs and outputs, it is again satisfactory to present all elements of the analysis via standard tabulation techniques and simple graphing techniques. However, with an increase in the number of DMUs and related inputs and outputs, more advance techniques are required to be implemented in order to quickly identify key pieces of information and avoid information overload. In (Porembski et al., 2005), Sammon's Mapping (Sammon, 1969) is applied to DEA results of efficiency of 140 bank branches in Germany and the paper illustrates using this specific example that this approach allows us to identify simultaneously efficient DMUs, severity of inefficiency of inefficient DMUs, pears for inefficient DMUs, outliers and influential DMUs. An alternative approach called Co-plots can be found in (Adler et al., 2007) applied to 19 Finnish Forestry Boards. Co-plots are a visualization technique based on multidimensional scaling which simplifies multivariate results of DEA modelling into two dimensions where DMUs are represented as points and inputs and outputs are represented as arrows. One of the key advantages of Co-plots is that they preserve the relative distance between DMUs in multidimensional space while transforming them to plain, i.e., proximity of DMUs in Coplot indicates how similar those DMUs are. Similar presentation techniques could be used also for analyses connected to point (7). Finally, point (8) is closely related to the previous points and it is context dependent, so we just point out here that we should use insight from presentation of the previous parts to guide us what actions/recommendations are feasible. Besides the aforementioned, we also fully support a reproducible approach to research, so we recommend making the data used for the modelling available, if not restricted by confidentiality agreements and licensing policies, so the presented results could be independently verified by their recipient.

Before concluding our paper, let us briefly discuss, for completeness, also two problems that may be of interest when applying DEA, ranking of DMUs and analysis of effectiveness over time. When we, in points (5) and (6), identify efficient and inefficient DMUs and compute their efficiency scores, we may not be interested in identifying how their performance could be improved, but our goal could be a ranking of the DMUs with respect to their efficiency. The problem is that more often than not multiple DMUs are seen as efficient, i.e., they attain the best possible efficiency score and thus are incomparable. To resolve this problem, we can use one of so-called super-efficiency methods which allows us to achieve a complete ranking. An example of such method can be found in (Andersen and Petersen, 1993). Regardless if we are interested in ranking of DMUs or in some recommendations how to improve their efficiency, we may be interested how technical efficiency of the DMUs changes and evolves over time with changing economic or technological conditions. In such cases dynamic DEA can be applied. The traditional tools that allow evaluation of the development of technical efficiency over time are window analysis (Klopp, 1985) and the Malmquist Index (MI). While the window analysis is based on the assumption of a constant level of technology and applies so called moving averages, MI includes in the concept of measuring changes in efficiency also a technological progress. The concept of MI was first introduced by Malmquist (1953) and has been further developed in the non-parametric framework by Caves et al. (1982), Färe et al. (1992), Thrall (2000), and Zofio (2007). MI is an index representing Total Factor Productivity (TFP) change of a DMU and in that it reflects both a change in technical efficiency as well as a change of the frontier technology between two periods of time.

CONCLUSION

We present here a guideline researcher can use to navigate through individual steps of applications of the DEA models. For each modelling phase, some general principles and issues as well as main approaches are outlined. We also stressed some of the problems which can encountered during the analysis alongside with some recommendations how to proceed when resolving them. Although we aimed for being as concise as possible, all argumentation is supported with quite exhaustive list of references which can be used as a starting point if further details regarding a particular modelling phase, model or issue need to be retrieved. Nevertheless, we plan to provide a more detailed treatise of some important issues connected to applications of DEA in the future and also provide some software solution which can help researcher in selection of the appropriate models in the context of principles and issues outlined in the paper.

ACKNOWLEDGMENT

The paper was supported by the grant scheme VEGA 1/0843/18 "Methodological aspects of DEA application on efficiency assessment of production units" of the Ministry of Education, Science, Research and Sport of the Slovak Republic.

References

ADLER, N. AND GOLANY, B. Evaluation of Deregulated Airline Networks Using Data Envelopment Analysis Combined with Principal Component Analysis with an Application to Western Europe. *European Journal of Operational Research*, 2001, 132, 2, pp. 260–273.

ADLER, N., RAVEH, A., YAZHEMSKY, E. DEA Presented Graphically Using Multi-Dimensional Scaling. In: *Modeling Data Irregularities and Structural Complexities in Data Envelopment Analysis*, Boston, MA: Springer US, 2007, pp. 171–187.

ADLER, N. AND YAZHEMSKY, E. Improving Discrimination in Data Envelopment Analysis: PCA–DEA or Variable Reduction. *European Journal of Operational Research,* 2010, 202, 1, pp. 273–284.

AHN, T. AND SEIFORD, L. M. Sensitivity of DEA to Models and Variable Sets in a Hypothesis Test Setting: the Efficiency of University Operations. In: IJIRI AND YUJI. (eds.) *Creative and Innovative Approaches to the Science of Management*, New York: Quorum Books, 1993.

ALIREZAEE, M., HAJINEZHAD, E., PARADI, J. C. Objective Identification of Technological Returns to Scale for Data Envelopment Analysis Models. *European Journal of Operational Research*, 2018, 266, 2, pp. 678–688.

ALLEN, R., ATHANASSOPOULOS, A., DYSON, R., THANASSOULIS, E. Weights Restrictions and Value Judgements in Data Envelopment Analysis: Evolution, Development and Future Directions. *Annals of Operations Research*, 1997, 73, pp. 13–34.

AMIRTEIMOORI, A. AND EMROUZNEJAD, A. Flexible Measures in Production Process: A DEA-based Approach. *RAIRO – Operations Research*, 2011, 45, 1, pp. 63–74.

AMIRTEIMOORI, A., EMROUZNEJAD, A., KHOSHANDAM, L. Classifying Flexible Measures in Data Envelopment Analysis: A Slackbased Measure. *Measurement*, 2013, 46, 10, pp. 4100–4107.

ANDERSEN, P. AND PETERSEN, N. C. A Procedure for Ranking Efficient Units in Data Envelopment Analysis. *Management Science*, 1993, 39, pp. 1261–1264.

ATHANASSOPOULOS, A. D. AND THANASSOULIS, E. Separating Market Efficiency from Profitability and its Implications for Planning. *The Journal of the Operational Research Society*, 1995, 46, 1, pp. 20–34.

AVILÉS-SACOTO, S. V., COOK, W. D., GÜEMES-CASTORENA, D., ZHU, J. Measuring Efficiency in DEA in the Presence of Common Inputs. *Journal of the Operational Research Society*, 2019, 71, 11, pp. 1-13.

BALA, K. AND COOK, W. D. Performance Measurement with Classification Information: An Enhanced Additive DEA Model. *Omega*, 2003, 31, 6, pp. 439–450.

BANKER, R. D. Maximum Likelihood, Consistency and Data Envelopment Analysis: a Statistical Foundation. *Management Science*, 1993, 39, 10, pp. 1265–1273.

BANKER, R. D. Hypothesis Tests Using Data Envelopment Analysis. *Journal of Productivity Analysis*, 1996, 7, 2–3, pp. 139–159.

BANKER, R. D., CHARNES, A., COOPER, W. W. Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis. *Management Science*, 1984, 30, 9, pp. 1078–1092.

BANKER, R. D. AND MOREY, R. C. Efficiency Analysis for Exogenously Fixed Inputs and Outputs. *Operations Research*, 1986, 34, 4, pp. 513–521.

BANKER, R. D. AND MOREY, R. C. The Use of Categorical Variables in Data Envelopment Analysis. *Management Science*, 1986, 32, 12, pp. 1613–1627.

BANKER, R. D. AND NATARAJAN, R. Evaluating Contextual Variables Affecting Productivity Using Data Envelopment Analysis. *Operations Research*, 2008, 56, 1, pp. 48–58.

BANKER, R. D. AND NATARAJAN, R. Statistical Tests Based on DEA Efficiency Scores. In: *Handbook on Data Envelopment Analysis*, Springer, 2011, pp. 273–295.

BANKER, R. D., NATARAJAN, R., ZHANG, D. Two-Stage Estimation of the Impact of Contextual Variables in Stochastic Frontier Production Function Models Using Data Envelopment Analysis: Second Stage OLS Versus Bootstrap Approaches. *European Journal of Operational Research*, 2019, 278, 2, pp. 368–384.

BANKER, R. D. AND THRALL, R. Estimation of Returns to Scale Using Data Envelopment Analysis. *European Journal of Operational Research*, 1992, 62, 1, pp. 74–84.

BARAT, M., TOHIDI, G., SANEI, M., RAZAVYAN, S. Data Envelopment Analysis for Decision Making Unit with Nonhomogeneous Internal Structures: An Application to the Banking Industry. *Journal of the Operational Research Society*, 2019, 70, 5, pp. 760–769.

BENÍTEZ-PEÑA, S., BOGETOFT, P., ROMERO MORALES, D. Feature Selection in Data Envelopment Analysis: A Mathematical Optimization approach. *Omega*, 2020, 96. <https://doi.org/10.1016/j.omega.2019.05.004>.

BOGETOFT, P. DEA on Relaxed Convexity Assumptions. *Management Science*, 1996, 42, 3, pp. 457–465.

BOGETOFT, P. AND OTTO, L. Benchmarking with DEA, SFA, and R. *International Series in Operations Research & Management Science*, New York, NY: Springer New York, 2011, Vol. 157.

BOGETOFT, P., TAMA, J. M., TIND, J. Convex Input and Output Projections of Nonconvex Production Possibility Sets. *Management Science*, 2000, 46, 6, pp. 858–869.

BOĎA, M. Classifying Flexible Measures in Data Envelopment Analysis: A Slacks-Based Measure – A Comment. *Measurement*, 2020, 150, Article 107045.

CASTELLI, L., PESENTI, R., UKOVICH, W. DEA-like Models for Efficiency Evaluations of Specialized and Interdependent Units. *European Journal of Operational Research*, 2001, 132, 2, pp. 274–286.

CAVES, D. W., CHRISTENSEN L. R., DIEWERT, W. E. The Economic Theory of Index Numbers and the Measurement of Input, Output and Productivity. *Econometrica*, 1982, 50, 6, pp. 1393–1414.

CHARNED, A., HAAG, S., JASKA, P., SEMPLE, J. Sensitivity of efficiency classifications in the additive model of data envelopment analysis. *International Journal of Systems Science*, 1992, 23, 5, pp. 789–798.
CHARNES, A., COOPER, W., RHODES, E. Measuring the efficiency of decision making units. *European Journal of Operational Research*, 1978, 2, 6, pp. 429–444.
CHARNES, A., COOPER, W. W., RHODES, E. Evaluating Program and Managerial Efficiency: an Application of Data Envelopment Analysis to Program Follow Through. *Management Science*, 1981, 27, 6, pp. 668–697.
CHARNES, A., COOPER, W. W., THRALL, R. M. A structure for classifying and characterizing efficiency and inefficiency in Data Envelopment Analysis. *Journal of Productivity Analysis*, 1991, 2, 3, pp. 197–237.
CHARNES, A. AND NERALIC, L. Sensitivity analysis of the proportionate change of inputs (or outputs) in data envelopment analysis. *Glasnik Matematički*, 1992, 27, 47, pp. 393–405.
CHEN, Y. Imprecise DEA–Envelopment and Multiplier Model Approaches. *Asia-Pacific Journal of Operational Research*, 2007, 24, 2, pp. 279–291.
CHENG, G., ZERVOPOULOS, P., QIAN, Z. A variant of radial measure capable of dealing with negative inputs and outputs in data envelopment analysis. *European Journal of Operational Research*, 2013, 225, 1, pp. 100–105.
COELLI, T. J., RAO, D. S. P., O'DONNELL, C., BATTESE, G. *An Introduction to Efficiency and Productivity Analysis.* New York: Springer-Verlag, 2005.
COOK, W. D., HARRISON, J., IMANIRAD, R., ROUSE, P., ZHU, J. Data envelopment analysis with nonhomogeneous DMUS. *Operations Research*, New York, 2013, 61, 3, pp. 666–676.
COOK, W. D., HARRISON, J., ROUSE, P., ZHU, J. Relative efficiency measurement: the problem of a missing output in a subset of decision making units. *European Journal of Operational Research*, 2012, 220, 1, pp. 79–84.
COOK, W. D., KRESS, M., SEIFORD, L. M. On the Use of Ordinal Data in Data Envelopment Analysis. *Journal of the Operational Research Society*, 1993, 44, 2, pp. 133–140.
COOK, W. D., KRESS, M., SEIFORD, L. M. Data Envelopment Analysis in the Presence of Both Quantitative and Qualitative Factors. *The Journal of the Operational Research Society*, 1996, 47, 7, pp. 945–953.
COOK, W. D. AND SEIFORD, L. M. Data envelopment analysis (DEA) – thirty years on. *European Journal of Operational Research*, 2009, 192, 1, pp. 1–17.
COOK, W. D. AND ZHU, J. Rank order data in DEA: a general framework. *European Journal of Operational Research*, 2006, 174, 2, pp. 1021–1038.
COOK, W. D. AND ZHU, J. Classifying inputs and outputs in data envelopment analysis. *European Journal of Operational Research*, 2007, 180, 2, pp. 692–699.
COOK, W. D., ZHU, J. (eds.) Data Envelopment Analysis. *International Series in Operations Research & Management Science.* Boston, MA: Springer US, 2014, Vol. 208.
COOPER, W. W., LI, S., SEIFORD, L. M., TONE, K., THRALL, R. M., ZHU, J. Sensitivity and stability analysis in dea: Some recent developments. *Journal of Productivity Analysis*, 2001, 15, 3, pp. 217–246.
COOPER, W. W., LI, S., SEIFORD, L. M., ZHU, J. Sensitivity Analysis in DEA. In: COOPER, W. ET AL. (eds.) *Handbook on Data Envelopment Analysis, International Series in Operations Research & Management Science*, Springer Science + Business Media, 2011, pp. 71–91.
COOPER, W. W., PARK, K. S., PASTOR, J. T. RAM: A Range Adjusted Measure of Inefficiency for Use with Additive Models, and Relations to Other Models and Measures in DEA. *Journal of Productivity Analysis*, 1999, 11, 1, pp. 5–42.
COOPER, W. W., SEIFORD, L. M., TONE, K. *Data Envelopment Analysis: a comprehensive text with models, applications, references and DEA-solver software.* 2nd Ed. Boston, MA: Springer US, 2007.
DIAKOULAKI, D., MAVROTAS, G., PAPAYANNAKIS, L. Determining objective weights in multiple criteria problems: the critic method. *Computers & Operations Research*, 1995, 22, 7, pp. 763–770.
DU, J., CHEN, Y., HUO, J. DEA for non-homogenous parallel networks. *Omega*, 2015, 56, pp. 122–132.
DYSON, R., ALLEN, R., CAMANHO, A., PODINOVSKI, V., SARRICO, C., SHALE, E. Pitfalls and protocols in DEA. *European Journal of Operational Research*, 2001, 132, 2, pp. 245–259.
EBRAHIMI, B. Efficiency measurement to identify the best efficient unit in the presence of dual-role factors. *International Journal of Applied Decision Sciences*, 2020, 13, 1, pp. 93–108.
EMROUZNEJAD, A., AMIN, G. R. DEA models for ratio data: Convexity consideration. *Applied Mathematical Modelling*, 2009, 33, 1, pp. 486–498.
EMROUZNEJAD, A., ANOUZE, A. L., THANASSOULIS, E. A semi-oriented radial measure for measuring the efficiency of decision making units with negative data, using DEA. *European Journal of Operational Researc*h, 2010, 200, 1, pp. 297–304.
FANCHON, P. Variable selection for dynamic measures of efficiency in the computer industry. *International Advances in Economic Research*, 2003, 9, 3, pp. 175–188.
FANDEL, P., MARIŠOVÁ, E., MALATINEC, T., LICHNEROVÁ, I. Decentralization Policies in Public Administration in Slovakia and the Czech Republic, and their Impact on Building Offices' Scale Efficiency. *Administrative Sciences*, 2019, 9, 4, pp. 1–15.

FARRELL, M. J. The Measurement of Productive Efficiency. *Journal of the Royal Statistical Society. Series A (General)*, 1957, 120, 3, pp. 253–290.
FIZEL, J. L. AND NUNNIKHOVEN, T. S. Technical Efficiency of For-profit and Non-profit Nursing Homes. *Managerial and Decision Economics*, 1992, 13, 5, pp. 429–439.
FIZEL, J. L. AND NUNNIKHOVEN, T. S. The Efficiency of Nursing Home Chains. *Applied Economics*, 1993, 25, 1, pp. 49–55.
FÄRE, R. AND GROSSKOPF, S. Malmquist Indexes and Fisher Ideal Indexes. *The Economic Journal*, 1992, 102, 410, pp. 158–160.
FÄRE, R., GROSSKOPF, S., LOVELL, C. A. K. The Structure of Technical Efficiency. *The Scandinavian Journal of Economics*, 1983, 85, 2, pp. 181–190.
GOLANY, B. AND ROLL, Y. An Application Procedure for DEA. *Omega*, 1989, 17, 3, pp. 237–250.
HAAS, D. A. AND MURPHY, F. H. Compensating for Non-homogeneity in Decision-Making Units in Data Envelopment Analysis. *European Journal of Operational Research*, 2003, 144, 3, pp. 530–544.
HADI-VENCHEH, A. AND ESMAEILZADEH, A. A New Super-efficiency Model in the Presence of Negative Data. *Journal of the Operational Research Society*, 2013, 64, 3, pp. 396–401.
HARRELL, F. E. *Regression Modeling Strategies.* Springer Series in Statistics, Cham: Springer International Publishing, 2015.
HATAMI-MARBINI, A. AND TOLOO, M. Data Envelopment Analysis Models with Ratio Data: A Revisit. *Computers & Industrial Engineering*, 2019, 133, pp. 331–338.
HLADÍK, M. Universal Efficiency Scores in Data Envelopment Analysis Based on a Robust Approach. *Expert Syst. Appl.*, 2019, 122, pp. 242–252. <https://doi.org/10.1016/j.eswa.2019.01.019>.
IMANIRAD, R., COOK, W. D., ZHU, J. Partial Input to Output Impacts in DEA: Production Considerations and Resource Sharing among Business Subunits. *Naval Research Logistics (NRL)*, 2013, 60, 3, pp. 190–207.
JAHANSHAHLOO, G., HOSSEINZADEH LOTFI, F., SHOJA, N., GHOLAM ABRI, A., FALLAH JELODAR, M., JAMALI FIROUZABADI, K. Sensitivity Analysis of Inefficient Units in Data Envelopment Analysis. *Mathematical and Computer Modelling*, Mar. 2011, 53, 5–6, pp. 587–596.
JENKINS, L. AND ANDERSON, M. A Multivariate Statistical Approach to Reducing the Number of Variables in Data Envelopment Analysis. *European Journal of Operational Research*, 2003, 147, 1, pp. 51–61.
JOLLIFFE, I. T. AND CADIMA, J. Principal Component Analysis: A Review and Recent Developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 2016, Vol. 374, Iss. 2065, Art. No. 20150202. <https://doi.org/10.1098/rsta.2015.0202>.
KAMAKURA, W. A Note on the Use of Categorical Variables in Data Envelopment Analysis. *Management Science*, 1988, 34, 10, pp. 1273–1276.
KAO, C. Network Data Envelopment Analysis. *International Series in Operations Research & Management Science*, Cham: Springer International Publishing, 2017, Vol. 240.
KAZEMI MATIN, R. AND AZIZI, R. A Two-phase Approach for Setting Targets in DEA with Negative Data. *Applied Mathematical Modelling*, 2011, 35, 12, pp. 5794–5803.
KERSTENS, K. AND VAN DE WOESTYNE, I. Negative Data in DEA: A Simple Proportional Distance Function Approach. *Journal of the Operational Research Society*, 2011, 62, 7, pp. 1413–1419.
KERSTENS, K. AND VAN DE WOESTYNE, I. A Note on a Variant of Radial Measure Capable of Dealing with Negative Inputs and Outputs in DEA. *European Journal of Operational Research*, 2014, 234, 1, pp. 341–342.
KLOPP, G. *The Analysis of the Efficiency of Production System with Multiple Inputs and Outputs.* PhD Thesis, University of Illinois at Chicago, 1985.
KORDROSTAMI, S., AMIRTEIMOORI, A., JAHANI SAYYAD NOVEIRI, M. Inputs and Outputs Classification in Integer-valued Data Envelopment Analysis. *Measurement*, 2019, 139, pp. 317–325.
LI, W., LIANG, L., COOK, W. D., ZHU, J. DEA Models for Non-homogeneous DMUs with Different Input Configurations. *European Journal of Operational Research*, 2016, 254, 3, pp. 946–956.
LIANG, L., YANG, F., COOK, W. D., ZHU, J. DEA Models for Supply Chain Efficiency Evaluation. *Annals of Operations Research*, 2006, 145, 1, pp. 35–49.
LIMLEAMTHONG, P. AND GUILLÉN-GOSÁLBEZ, G. Mixed-Integer Programming Approach for Dimensionality Reduction in Data Envelopment Analysis: Application to the Sustainability Assessment of Technologies and Solvents. *Industrial & Engineering Chemistry Research*, 2018, 57, 30, pp. 9866–9878.
LIN, R. AND CHEN, Z. A Directional Distance Based Super-efficiency DEA Model Handling Negative Data. *Journal of the Operational Research Society*, 2017, 68, 11, pp. 1312–1322.
LIN, R., YANG, W., HUANG, H. A Modified Slacks-based Super-efficiency Measure in the Presence of Negative Data. *Computers & Industrial Engineering*, 2019, 135, pp. 39–52.
LIU, J. S., LU, L. Y., LU, W.-M., LIN, B. J. A Survey of DEA Applications. *Omega*, 2013, 41, 5, pp. 893–902.
MALMQUIST, S. Index Numbers and Indifference Surfaces. *Trabajos de Estadistica*, 1953, 4, pp. 209–242.
MANCEBON, M. J. AND MOLINERO, C. M. Performance in Primary Schools. *Journal of the Operational Research Society*, 2000, 51, 7, pp. 843–854.

MENDELOVÁ, V. AND KANDEROVÁ, M. Measuring the Efficiency of Bus Transport Undertaking in the Slovak Republic: An Application of DEA and Tobit. In: *FERNSTAT 2016: conference proceedings*, Šachtička, 22–23 September 2016, Banská Bystrica, Slovenská štatistická a demografická spoločnosť, pp. 84–94.
MOLINERO, C. M. *Non-homogeneity in Data Envelopment Analysis.* University of Kent Canterbury, Canterbury, 2008, Working Paper Number 163.
NATARAJA, N. R. AND JOHNSON, A. L. Guidelines for Using Variable Selection Techniques in Data Envelopment Analysis. *European Journal of Operational Research*, 2011, 215, 3, pp. 662–669.
NERALIC, L. Sensitivity in Data Envelopment Analysis for Arbitrary Perturbations of Data. *Glasnik Matematicki*, 1997, 32, pp. 315–335.
NOLAN, J. F. Determinants of Productive Efficiency in Urban Transit. *Logistics and Transportation Review*, 1996, 32, 3, pp. 319–342.
OLESEN, O. B. AND PETERSEN, N. C. Target and Technical Efficiency in DEA: Controlling for Environmental Characteristics. *Journal of Productivity Analysis*, 2009, 32, 1, pp. 27–40.
OLESEN, O. B., PETERSEN, N. C., PODINOVSKI, V. V. Efficiency Analysis with Ratio Measures. *European Journal of Operational Research*, 2015, 245, 2, pp. 446–462.
OUM, T. H. AND YU, C. Economic Efficiency of Railways and Implications for Public Policy: A Comparative Study of the OECD Countries' Railways. *Journal of Transport Economics and Policy*, 1994, 28, 2, pp. 121–138.
OZBEK, M. E., DE LA GARZA, J. M., TRIANTIS, K. Data Envelopment Analysis as a Decision-Making Tool for Transportation Professionals. *Journal of Transportation Engineering*, 2009, 135, 11, pp. 822–831.
PARK, K. S. Efficiency Bounds and Efficiency Classifications in DEA with Imprecise Data. *Journal of the Operational Research Society*, 2007, 58, 4, pp. 533–540.
PASTOR, J. T., RUIZ, J. L., SIRVENT, I. A Statistical Test for Nested Radial DEA Models. *Operations Research*, 2002, 50, 4, pp. 728–735.
POREMBSKI, M., BREITENSTEIN, K., ALPAR, P. Visualizing Efficiency and Reference Relations in Data Envelopment Analysis with an Application to the Branches of a German Bank. *Journal of Productivity Analysis*, 2005, 23, 2, pp. 203–221.
PORTELA, M. C. A. S., THANASSOULIS, E., SIMPSON, G. Negative Data in DEA: A Directional Distance Approach Applied to Bank Branches. *Journal of the Operational Research Society*, 2004, 55, 10, pp. 1111–1121.
RAMANATHAN, R. *An Introduction to Data Envelopment Analysis.* New Delhi: Sage Publications, 2003.
RAY, S. C. Resource-use Efficiency in Public Schools: A Study of Connecticut Data. *Management Science*, 1991, 37, 12, pp. 1620–1628.
READ, L. E. AND THANASSOULIS, E. Improving the Identification of Returns to Scale in Data Envelopment Analysis. *Journal of the Operational Research Society*, 2000, 51, 1, pp. 102–110.
ROHÁČOVÁ, V. A DEA Based Approach for Optimization of Urban Public Transport System. *Central European Journal of Operations Research*, 2015, 23, 1, pp. 215–233.
ROUSSEAU, J. J. AND SEMPLE, J. H. Categorical Outputs in Data Envelopment Analysis. *Management Science*, 1993, 39, 3, pp. 384–386.
RUGGIERO, J. Impact Assessment of Input Omission on DEA. *International Journal of Information Technology & Decision Making*, 2005, 4, 3, pp. 359–368.
SAMMON, J. W. A Nonlinear Mapping for Data Structure Analysis. *IEEE Transactions on Computers C-18*, 1969, 5, pp. 401–409.
SCHEEL, H. Undesirable Outputs in Efficiency Valuations. *European Journal of Operational Research*, 2001, 132, 2, pp. 400–410.
SEDIGHI HASSAN KIYADEH, M., SAATI, S., KORDROSTAMI, S. Improvement of Models for Determination of Flexible Factor Type in Data Envelopment Analysis. *Measurement*, 2019, 137, pp. 49–57.
SEIFORD, L. M. AND ZHU, J. Sensitivity Analysis of DEA Models for Simultaneous Changes in All the Data. *Journal of the Operational Research Society*, 1998, 49, 10, pp. 1060–1071.
SEXTON, T. R., SLEEPER, S., TAGGART, R. E. Improving Pupil Transportation in North Carolina. *Interfaces*, 1994, 24, 1, pp. 87–103.
SHARP, J. A., MENG, W., LIU, W. A Modified Slacks-based Measure Model for Data Envelopment Analysis with Natural Negative Outputs and Inputs. *The Journal of the Operational Research Society*, 2007, 58, 12, pp. 1672–1677.
SIMAR, L. AND WILSON, P. W. Testing Restrictions in Nonparametric Efficiency Models. *Communications in Statistics – Simulation and Computation*, 2001, 30, 1, pp. 159–184.
SIMAR, L. AND WILSON, P. W. Non-parametric Tests of Returns to Scale. *European Journal of Operational Research*, 2002, 139, 1, pp. 115–132.
SIMAR, L. AND WILSON, P. W. Estimation and Inference in Two-stage, Semi-parametric Models of Production Processes. *Journal of Econometrics*, 2007, 136, 1, pp. 31–64.
SIMAR, L. AND WILSON, P. W. Two-stage DEA: Caveat Emptor. *Journal of Productivity Analysis*, 2011, 36, 2, pp. 205–218.
SMITH, P. AND MAYSTON, D. Measuring Efficiency in the Public Sector. *Omega*, 1987, 15, 3, pp. 181–189.
SOTERIOU, A. AND ZENIOS, S. A. Operations, Quality, and Profitability in the Provision of Banking Services. *Management Science*, 1999, 45, 9, pp. 1221–1238.

THANASSOULIS, E. *Introduction to the Theory and Application of Data Envelopment Analysis*. Boston, MA: Springer US, 2001.
THANASSOULIS, E., BOUSSOFIANE, A., DYSON, R. Exploring Output Quality Targets in the Provision of Perinatal Care in England Using Data Envelopment Analysis. *European Journal of Operational Researc*h, 1995, 80, 3, pp. 588–607.
THANASSOULIS, E., PORTELA, M. C. S., DESPIĆ, O. Data Envelopment Analysis: The Mathematical Programming Approach to Efficiency Analysis. In: *The Measurement of Productive Efficiency and Productivity Change*, Oxford University Press, 2008, pp. 251–420.
THOMPSON, R., DHARMAPALA, P. S., THRALL, R. M. Sensitivity Analysis of Efficiency Measures with Applications to Kansas Farming and Illinois Coal Mining. In: *Data Envelopment Analysis: Theory, Methodology, and Applications*, Dordrecht: Springer Netherlands, 1994, pp. 393–422.
THOMPSON, R. G., DHARMAPALA, P. S., DIAZ, J., GONZÁLEZ-LIMA, M. D., THRALL, R. M. DEA Multiplier Analytic Center Sensitivity with an Illustrative Application to Independent Oil Companies. *Annals of Operations Research*, 1996, 66, 2, pp. 163–177.
THRALL, R. M. Measures in DEA with an Application to the Malmquist Index. *Journal of Productivity Analysis*, 2000, 13, 2, pp. 125–137.
TOLOO, M., KESHAVARZ, E., HATAMI-MARBINI, A. Dual-role Factors for Imprecise Data Envelopment Analysis. *Omega*, 2018, 77, pp. 15–31.
TONE, K. A Slacks-based Measure of Efficiency in Data Envelopment Analysis. *European Journal of Operational Research*, 2001, 130, 3, pp. 498–509.
TONE, K. AND TSUTSUI, M. Network DEA: A Slacks-based Measure Approach. *European Journal of Operational Research*, 2009, 197, 1, pp. 243–252.
UEDA, T. AND HOSHIAI, Y. Application of Principal Component Analysis for Parsimonious Summarization of DEA Inputs and/or Outputs. *Journal of the Operations Research Society of Japan*, 1997, 40, 4, pp. 466–478.
WAGNER, J. M. AND SHIMSHAK, D. G. Stepwise Selection of Variables in Data Envelopment Analysis: Procedures and Managerial Perspectives. *European Journal of Operational Research*, 2007, 180, 1, pp. 57–67.
WANG, T.-C. AND LEE, H.-D. Developing a Fuzzy TOPSIS Approach Based on Subjective Weights and Objective Weights. *Expert Systems with Applications*, 2009, 36, 5, pp. 8980–8985.
WEI, F., SONG, J., JIAO, C., YANG, F. A Modified Slacks-based Ranking Method Handling Negative Data in Data Envelopment Analysis. *Expert Systems*, 2019, 36, 1, article ID e12329.
WILSON, P. W. Detecting Influential Observations in Data Envelopment Analysis. *Journal of Productivity Analysis*, 1995, 6, 1, pp. 27–45.
WILSON, P. W. Dimension Reduction in Nonparametric Models of Production. *European Journal of Operational Research*, 2018, 267, 1, pp. 349–367.
ZHU, J. Robustness of the Efficient DMUs in Data Envelopment Analysis. *European Journal of Operational Research*, 1996, 90, 3, pp. 451–460.
ZHU, J. Imprecise Data Envelopment Analysis (IDEA): a Review and Improvement with an Application. *European Journal of Operational Research*, 2003, 144, 3, pp. 513–529.
ZHU, J. Solving DEA via Excel. In: *Multi-Objective Programming and Goal Programming*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2003, pp. 301–306.
ZHU, J. Imprecise DEA via Standard Linear DEA Models with a Revisit to a Korean Mobile Telecommunication Company. *Operations Research*, 2004, 52, 2, pp. 323–329.
ZHU, J. Quantitative Models for Performance Evaluation and Benchmarking: Data Envelopment Analysis with Spreadsheets. *International Series in Operations Research & Management Science*, Boston, MA: Springer US, 2009, Vol. 126.
ZOFIO J. L. Malmquist Productivity Index Decompositions: A Unifying Framework. *Applied Economics*, 2007, 39, 18, pp. 2371–2387.

Papers
We publish articles focused at theoretical and applied statistics, mathematical and statistical methods, conception of official (state) statistics, statistical education, applied economics and econometrics, economic, social and environmental analyses, economic indicators, social and environmental issues in terms of statistics or economics, and regional development issues.

The journal of *Statistika* has the following sections:

The *Analyses* section publishes high quality, complex, and advanced analyses based on the official statistics data focused on economic, environmental, and social spheres. Papers shall have up to 12 000 words or up to twenty (20) 1.5-spaced pages.

Discussion brings the opportunity to openly discuss the current or more general statistical or economic issues, in short what the authors would like to contribute to the scientific debate. Contribution shall have up to 6 000 words or up to 10 1.5-spaced pages.

The *Methodology* section gives space for the discussion on potential approaches to the statistical description of social, economic, and environmental phenomena, development of indicators, estimation issues, etc. Papers shall have up to 12 000 words or up to twenty (20) 1.5-spaced pages.

Consultation contains papers focused primarily on new perspectives or innovative approaches in statistics or economics about which the authors would like to inform the professional public. Consultation shall have up to 6 000 words or up to 10 1.5-spaced pages.

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

The *Information* section includes informative (descriptive) texts, information on latest publications (issued not only by the CZSO), recent and upcoming scientific conferences. Recommended range of information is 6 000 words or up to 10 1.5-spaced pages.

Language
The submission language is English only. Authors are expected to refer to a native language speaker in case they are not sure of language quality of their papers.

Recommended Paper Structure
Title (e.g. On Laconic and Informative Titles) — Authors and Contacts — Abstract (max. 160 words) — Keywords (max. 6 words / phrases) — JEL classification code — Introduction — (chapters: 1, 2, …) — Conclusion — Acknowledgments — References — Annex — Tables and Figures (for print, for review process in the text)

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 underline in the text. Paper parts numbering: 1, 1.1, 1.2, etc.

Headings
1 FIRST-LEVEL HEADING (Times New Roman 12, bold)
1.1 Second-level heading (Times New Roman 12, bold)
1.1.1 Third-level heading (Times New Roman 12, bold italic)

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

References in the Text
Place reference in the text enclosing authors' names and the year of the reference, e.g. "White (2009) points out that…", "… recent literature (Atkinson and Black, 2010a, 2010b, 2011; Chase et al., 2011, pp. 12–14) conclude…". Note the use of alphabetical order. Include page numbers if appropriate.

List of References
Arrange list of references alphabetically. Use the following reference styles: [for a book] HICKS, J. *Value and Capital: An inquiry into some fundamental principles of economic theory*. 1st Ed. Oxford: Clarendon Press, 1939. [for chapter in an edited book] DASGUPTA, P. et al. Intergenerational Equity, Social Discount Rates and Global Warming. In: PORTNEY, P. AND WEYANT, J., eds. *Discounting and Intergenerational Equity*. Washington, D.C.: Resources for the Future, 1999. [for a journal] HRONOVÁ, S., HINDLS, R., ČABLA, A. Conjunctural Evolution of the Czech Economy. *Statistika: Statistics and Economy Journal*, 2011, 3 (September), pp. 4–17. [for an online source] CZECH COAL. *Annual Report and Financial Statement 2007* [online]. Prague: Czech Coal, 2008. [cit. 20.9.2008]. <http://www.czechcoal.cz/cs/ur/zprava/ur2007cz.pdf>.

Tables
Provide each table on a separate page. Indicate position of the table by placing in the text "insert Table 1 about here". Number tables in the order of appearance Table 1, Table 2, etc. Each table should be titled (e.g. Table 1 Self-explanatory title). Refer to tables using their numbers (e.g. see Table 1, Table A1 in the Annex). Try to break one large table into several smaller tables, whenever possible. Separate thousands with a space (e.g. 1 528 000) and decimal points with a dot (e.g. 1.0). Specify the data source below the tables.

Figures
Figure is any graphical object other than table. Attach each figure as a separate file. Indicate position of the figure by placing in the text "insert Figure 1 about here". Number figures in the order of appearance Figure 1, Figure 2, etc. Each figure should be titled (e.g. Figure 1 Self-explanatory title). Refer to figures using their numbers (e.g. see Figure 1, Figure A1 in the Annex).

Figures should be accompanied by the *.xls, *.xlsx table with the source data. Please provide cartograms in the vector format. Other graphic objects should be provided in *.tif, *.jpg, *.eps formats. Do not supply low-resolution files optimized for the screen use. Specify the source below the figures.

Formulas
Formulas should be prepared in formula editor in the same text format (Times 12) as the main text and numbered.

Paper Submission
Please email your papers in *.doc, *.docx or *.pdf formats to statistika.journal@czso.cz. All papers are subject to double-blind peer review procedure. Articles for the review process are accepted continuously and may contain tables and figures in the text (for final graphical typesetting must be supplied separately as specified in the instructions above). Please be informed about our Publication Ethics rules (i.e. authors responsibilities) published at: http://www.czso.cz/statistika_journal.

Managing Editor: Jiří Novotný
Phone: (+420) 274 054 299 | **fax:** (+420) 274 052 133
E-mail: statistika.journal@czso.cz | **web:** www.czso.cz/statistika_journal
Address: Czech Statistical Office | Na padesátém 81 | 100 82 Prague 10 | Czech Republic

Subscription price (4 issues yearly)
CZK 66 per copy + postage.
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
Czech Statistical Office | Information Services
Na padesátém 81 | 100 82 Prague 10 | Czech Republic
Phone: (+420) 274 052 733, (+420) 274 052 783
E-mail: objednavky@czso.cz

Design: Toman Design
Layout: Ondřej Pazdera
Typesetting: Družstvo TISKOGRAF, David Hošek
Print: Czech Statistical Office

101st year of the series of professional statistics and economy journals of the State Statistical Service in the Czech Republic: ***Statistika*** (since 1964), ***Statistika a kontrola*** (1962–1963), ***Statistický obzor*** (1931–1961) and ***Československý statistický věstník*** (1920–1930).

Published by the Czech Statistical Office
ISSN 1804-8765 (Online)
ISSN 0322-788X (Print)
Reg. MK CR E 4684