STATISTIKA

STATISTICS AND ECONOMY JOURNAL

VOL. **98** (3) 2018

CZECH STATISTICAL OFFICE

CONTENTS

ANALYSES

INFORMATION

About Statistika

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

Publisher

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

Contact us

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

Pension Liabilities in the Czech Republic

Zdeněk Skalák [1] | *Czech Statistical Office, Prague, Czech Republic*
Václav Rybáček [2] | *Jan Evangelista Purkyne University, Usti nad Labem, Czech Republic*

Abstract

The pension liabilities drew a wide attention due to ageing of population, especially the liabilities originating in the operation of the unfunded scheme (pay-as-you-go). Changing demographic situation and the interest of users gave rise to a request on the provision of a new data aiming to provide a basis for analysis and the decision-making of policy-makers. The purpose of this paper is to present the calculations of the total pension liabilities for the Czech Republic which were first released by the Czech Statistical Office in 2018. Besides, the paper presents a sensitive analysis and also compares the final figures internationally.

Keywords

Pension liabilities, pension systems, ageing of population

JEL code

H55, H75, H50

INTRODUCTION

Financial situation of pension systems in developed countries and a need for their reforms has currently become one of the topical issues. Demographic development over the last decades, namely an increasing proportion of retirees in the entire population, has necessitated a number of reform plans. In the field of statistics, this trend manifested itself into a request to compile relevant aggregates chiefly on the operation of the pay-as-you-go system (hereinafter: PAYG) which is commonly referred to as "the first pillar" and which is generally the dominant way how pension protection is institutionalized.

Request to deliver relevant data has been translated into a new table in the Eurostat´s Transmission program (so-called Table 29).[3] At the end of March 2018, the Czech Statistical Office (hereinafter CZSO) firstly released the total pension liabilities for the period ranging from 2011 to 2015. Although the Table 29 covers all pension liabilities including those of private pension schemes, this text is focused on the liabilities arising along with a system of the PAYG only. The reason is that while the liabilities of private pension schemes are explicitly recognized in business and national accounting, PAYG system´s liability must be calculated by statisticians themselves. We will thus present and discuss the way of calculation; then we compare the totals for the Czech Republic internationally as well as with other experimental calculations previously published in the Czech economic journals.

[1] Czech Statistical Office, Na padesátém 81, 100 82 Prague 10, Czech Republic. E-mail: zdenek.skalak@czso.cz.
[2] UJEP, Moskevská 54, Ústí nad Labem, Czech Republic. Also the Czech Statistical Office, Na padesátém 81, 100 82 Prague 10, Czech Republic. E-mail: vaclav.rybacek@czso.cz.
[3] The transmission program whose part is the Table 29 (Accrued-to-date pension entitlements in social insurance) specifies the nature and timeliness of data delivery within the framework of the European System of National and Regional Accounts.

The purpose of the data on pension liabilities is manifold. There is a difference in the treatment of funded and unfunded pension schemes (such as a PAYG system) in national accounts. Liabilities arising in the operation of the PAYG system are recorded off-balance only in the government accounts. This implies that a part of future government liabilities is not explicitly shown let alone to be a part of the general government debt. On the contrary, the funded schemes are recorded under the item AF.6[4] in national accounts adding to the national accounts aggregates.

Different recording of funded and unfunded schemes obviously poses an obstacle for the international comparability of the aggregates on pension liabilities (Bloch-Fall, 2015). Simply, the total amount of pension liabilities, as shown in the sector accounts, depends on the way how pension protection is organized in individual countries. This incomparability is supposedly overcome by introducing the supplementary table covering pension liabilities irrespective of their nature as funded or unfunded.

Furthermore, data on pension liabilities can also serve as an important input into decision-making process of policy-makers. An imbalance in the pension system, which is ideally supposed to be balanced, can create a pressure on the overall government revenues and expenditures. In other words, changes in tax legislation or expenditure cuts can become inevitable in the face of reoccurring imbalances in the pension system, i.e. especially when pension payments exceed pension contributions.

This brings us to an important aspect causing that aggregates discussed in this paper shall be read with great caution. In case of aggregates on the pension liabilities, we pay attention to the liability side only with no reference to the asset-side (Goebel, 2017). Simply, it can be hardly deduced from Table 29 whether the pension system is sustainable or not, because of missing information on revenues from contributions which will accrue to government in the future.[5] Table 29 thus offers only partial information on the financial soundness of the pension system.

1 BACKGROUND

The supplementary table was created in a situation which still poses fiscal challenges to authorities. We can only remind the trends of growing proportion of people at the retirement age creating pressure on the government expenditures, or decreasing share of economically active persons in the population bringing insufficient funds into the pension system. As the following table published in the Ageing report[6] demonstrates, the old-age dependency ratio (i.e. the age group of 65 and older to the age group of 15–64) will nearly double across the EU countries (increasing from the current 29.0% to 50.2%). In other words, during the reference period, the number of persons at the productive age per a person at the retirement age is expected to fall from four to two. In the Czech Republic, the expected changes in the coefficient are alarming, so the increase in the age group of 65+ from the current 18 to 28% in 2060 can create a strong need for interventions into the pension system.

Among the main causes of this trend, both declining birth rates and mortality rates leading to a prolonged life expectancy should be mentioned in particular. Gradually increasing life expectancy has a significant effect on the mortality tables which are an essential input into the calculations of pension liabilities. Technically speaking, the life expectancy represents the average period of time that a person of the age "x" may expect to live under the current immutable conditions. By 2060, this indicator is expected to increase in the Czech Republic to 85 years for men and 88 years for women.

[4] The item with the code AF.6 covers stock of insurance, pension and guarantee schemes.

[5] For interesting discussion on generational accounting calculating "lifetime net tax rates" for a given cohort, see Ruffing, de Water, Kogan (2014).

[6] The Ageing report for the European Population is published by the European Commission every three years. This document serves a platform for debates at the European level. Based on this document, the sustainability of public finance is assessed in the form of expenditures closely linked to demographic developments (retirement pensions, health care, education or unemployment benefits). It is also used to assess the impact of demographic developments on the labour market and potential economic growth.

Table 1 Projection of population ageing expressed by the share of selected population groups, comparison between the EU and the Czech Republic, 2013–2060

Age group	Year / Area	2013	2020	2025	2030	2035	2040	2045	2050	2055	2060
0–14	EU	15.6	15.6	15.2	14.9	14.6	14.6	14.8	15.0	15.0	15.0
	CZ	15.3	16.0	15.3	14.7	14.3	14.7	15.3	15.8	15.7	15.4
15–64	EU	65.4	63.9	62.6	61.1	59.6	58.4	57.5	56.9	56.6	56.6
	CZ	66.6	63.8	63.3	63.0	62.7	60.6	58.0	56.7	56.2	56.4
65+	EU	19.0	20.5	22.2	24.1	25.8	27.0	27.7	28.2	28.4	28.4
	CZ	18.1	20.2	21.4	22.3	23.0	24.7	26.7	27.5	28.1	28.2

Source: Ageing Report 2015

Table 2 suggests that the Czech Republic has been slowly approaching the EU-levels in terms of life expectancy; moreover, this trend is expected to continue in the future. This holds for both life expectancy at birth and life expectancy at age 65. For the latter, the average life expectancy for men is expected to grow from nearly 16 years, as of today, to 21 years in 2060. In case of women, the projection counts with an increase on a similar proportion as for men, i.e. from 19.5 years to 24.5 years. It can be concluded that the Czech Republic, as well as other developed countries, has been going through the demographic changes which pose a risk to the PAYG system´s financial balance.

Table 2 Life expectancy at birth and at age 65, EU and the Czech Republic, 2015–2060

Life expectancy	Area	Year / Sex	2015	2020	2025	2030	2035	2040	2045	2050	2055	2060
E_0	EU	Men	78	78.9	79.7	80.5	81.3	82.0	82.8	83.5	84.1	84.8
		Women	83.4	84.1	84.8	85.5	86.1	86.8	87.4	88.0	88.5	89.1
	CZ	Men	75.5	76.5	77.4	78.3	79.2	80.1	80.9	81.7	82.5	83.3
		Women	81.5	82.3	83.1	83.8	84.5	85.3	85.9	86.6	87.3	87.9
E_{65}	EU	Men	17.9	18.4	18.9	19.5	20.0	20.5	21.0	21.5	22.0	22.4
		Women	21.2	21.8	22.3	22.8	23.3	23.8	24.3	24.7	25.2	25.6
	CZ	Men	15.9	16.6	17.2	17.8	18.4	19.0	19.6	20.1	20.7	21.2
		Women	19.4	20.0	20.6	21.2	21.8	22.4	22.9	23.5	24.0	24.5

Source: Ageing Report 2015

European countries thus clearly face a demographic trend which can bring about a chronicle imbalance to the pension system due to contrasting trends in a number of beneficiaries and contributors. This impacts mainly the un-funded systems where there are no individual pension plans which would make current retirement savings explicit. As a matter of fact, benefits currently payable are financed by social contributions which are paid to the State by currently employed workers and theirs employers. This makes demographic situation very closely interlinked to the financial balance of unfunded pension system.

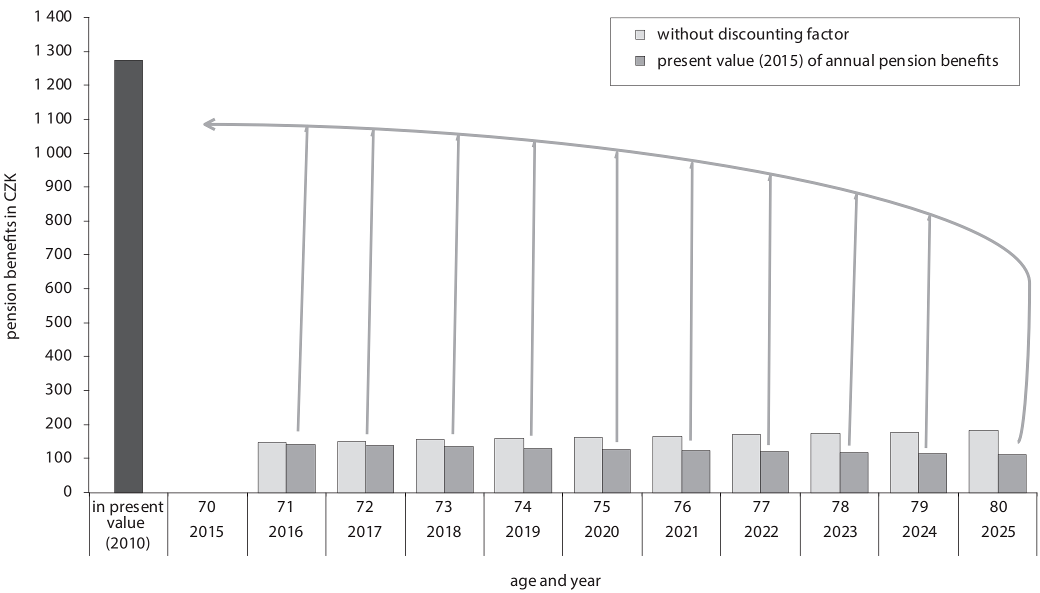
Before going into greater detail, it is worth making few more remarks on the nature of the PAYG system. It can be categorized into defined benefit schemes where the amount of benefits that participants receive at the retirement age is predefined by the law. Besides, the law stipulates the amount of compulsory contributions which applies to all economically active people of the age over 18 years who contributes a certain percentage of their income which is set regardless of level of the income. These funds are then distributed in the form of pensions to current pensioners by main actors.[7]

2 METHODOLOGY

In the methodological part of the paper, we will describe the way of calculation as well as assumptions which is essential to be aware of for an assessment of the explanatory power of aggregates.[8] Accepting some simplistic assumption regarding the future macroeconomic development is inevitable not only because these aggregates fall within the area of macroeconomic figures, but chiefly because we are dealing with projections reaching far into the future. One of the key issues which will be discussed in greater detail is the choice of relevant discount rate. To use a discount rate is an essential part of the calculation as the final aggregates are compared with nominal GDP for given period; future pensions must be thus presented in their current values.

As shown below, the final results are highly discount-rate-sensitive. From the households´ perspective, the discount factor reflects the time-value of money. Current payments are considered more valuable

Figure 1 An example of a discount factor application



Source: Technical Compilation Guide for Pension Data in National Accounts

[7] In the Czech Republic, the following institutions are involved: the Ministry of Labour and Social Affairs, the Ministry of Justice, the Ministry of Interiors and the Ministry of Defence.

[8] The calculation of liabilities follows basic principles stipulated in the Technical Compilation Guide released by Eurostat and the European Central Bank (2011).

than future payments because of potential risks associated with future payment flows. This means that the discount factor is generally less than 1, i.e. the present is preferred to the future. In terms of the pension manager (the State), the discount factor is used to calculate the pension reserves currently required to finance future pension liabilities. The following chart shows a model example of pension liability for pensioners in 2015 at the current age of 70 who are expected to live for 80 years. Such a pensioner will be reimbursed a pension of 151 thousand CZK a year to 184 thousand CZK a year at the end of the period. These benefits will be reduced by the discount factor to arrive at their present value.

Discount factor thus significantly impacts the extent of pension liabilities. For the sake of our calculation, nominal discount rate of 5 percent was used in line with the recommendation of the Manual on Pension Liabilities. It is obvious that this assumption is very strong and it calls for the availability of alternative scenarios which are presented below. Besides, the future course of pensions will be, of course, directly affected by the performance of economy (translated into wage growth), amendments in the law (retirement age) or by development in the monetary sphere (rate of inflation). Due to these, it is necessary to incorporate further assumptions concerning the future macroeconomic conditions. These are the followings:

- Rate of inflation – it is defined as an increase in the general price level of goods and services in the economy over a certain period of time. According to the Act on Pension Valuation, they are adjusted annually, among other things, by the rate of inflation; therefore, this indicator is applied in the calculation of the liabilities of current pensioners. In the model, inflation rate of 2% is considered according to the CNB inflation target of 2015.
- Coefficient for adjusting the general assessment base – this indicator determined by the government regulation based on the growth of the average wage. All new types of pension are adjusted accordingly. We apply to the commitments of those generations that are not yet retired. Based on the average of this coefficient from 2008 to 2015.
- Real wage growth – future course of real wages, i.e. real value of monetary rewards of workers participating in the productions process, plays an important role as it serves as basis for the pension calculation. Since 2008, it is part of the annual valorisation of existing pensions to which one-third contributes. Real wage growth is assessed as the average in the period from 2008 to 2016.

Furthermore, as we will see below, assumptions regarding the future demographic situation must be incorporated. Concretely, the final values are directly influenced by the expected probability to be alive at the year x + 1, ongoing migration or the development of fertility rate which all already serve as an input into the compilation of demographic projection. As the demographic projection constitutes one of the key inputs into the calculation of pension liabilities, no further specific calculations are needed in this respect.[9]

Model of pension liabilities under the first pillar is based on the German Freiburg model – it is referred to as "The ADL model" (accrued-to-date liabilities), which represents the present value of the pensions to be paid in the future. ADLs measure the amount of money required by the pension system to meet its commitments in the theoretical case of closing down the system. In case of government- sponsored unfunded pension schemes, such a scenario is practically ruled out. Expressing the ADLs of the unfunded defined benefit schemes of government employees and social security schemes in terms of GDP gives an idea of the number of annual products a country would have to spend to meet its pension commitments.

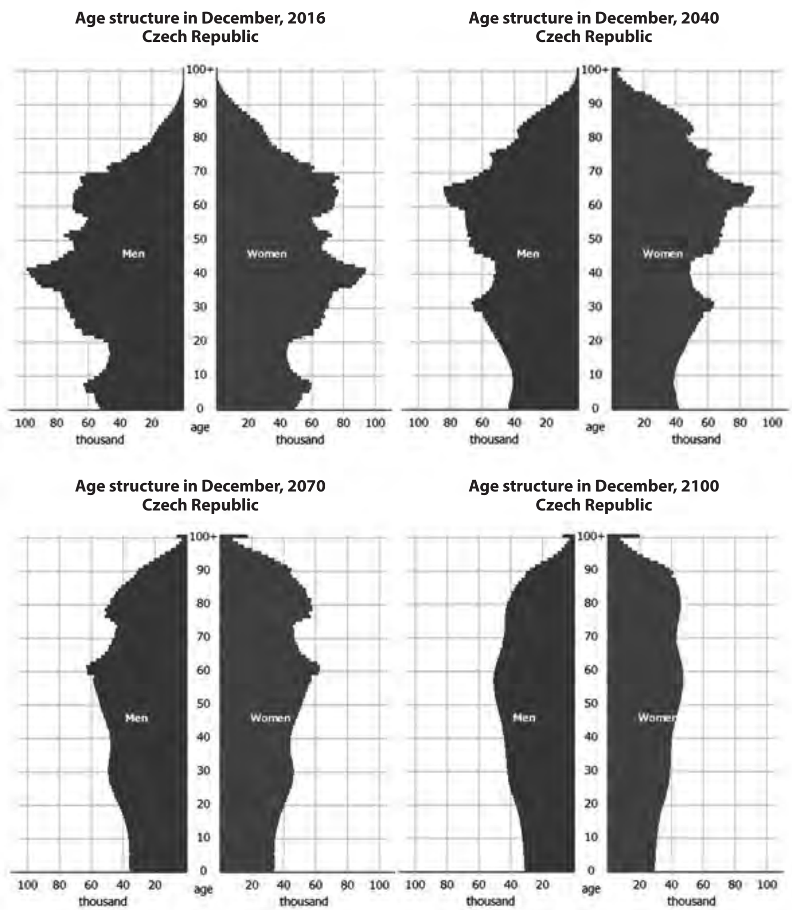
As already mentioned, modelling of commitments is only focused on the first pension pillar where these data are not captured in the national accounts system. The model calculates the liabilities of current and future pensioners, i.e. entitlements accrued to persons who have not yet reached retirement age, are counted in the final total of liabilities.

[9] The demographic projection as published by the CZSO in 2013 was incorporated into the calculation.

The model calculation itself consists of several basic steps. The very first thing is to acquire relevant primary data. To briefly recall, the model incorporates data on pensions which are structured by gender and age of beneficiaries. Relevant information is retrieved from the databases containing several types of pensions paid by the state institutions engaged in the pension protection (see footnote 7). Another key input already mentioned is the demographic projection as published by the CZSO.

The demographic projection is displayed in the following age pyramids population projection by 2100 (CZSO, 2013). There is a visible population ageing process as the age pyramid is increasingly changing into a regressive type of decreasing size of population. The least people are found in the child's component and most in the post-reproduction (or with reproduction at a similar level). The base of the chart is narrow and the sides are convex, which means that the number of newly born children is steadily decreasing and, in the longer term, the total population will decrease. In this case, the population is ageing and the economic burden on the entire population is increasing. Graphically, this state is represented by a significantly wider vertex.

Picture 1 Age structure of the Czech Republic in selected years



Source: CZSO

As a further step, the average pension allowances for a single representative of each generation are determined. The purpose of the calculation is to find out the commitment the system brings to the annual cohort of the population. Once the average amount for each cohort is set, the number of people in each cohort (according to the projection) in the following years is multiplied by particular estimated average income (up to the age of 100 in old-age pensions - according to the population projection). After doing so, we will arrive at the amount of pension liabilities by individual generations. Formally, the estimation of the existing retirees' benefits is based on the following identity:

$$p_{b,k}=\frac{B_{b,k}\cdot M_{b,k}}{C_{b,k}} \quad \wedge \quad P_b=\sum_{k=b-D}^{b} p_{b,k}\cdot C_{b,k}\ , \tag{1}$$

where:

$p_{b,k}$ = the pension benefits in the base year b of the cohort born in k,
$B_{b,k}$ = average pension benefit of a certain age x,
$M_{b,k}$ = number of scheme retirees of a certain age x,
$C_{b,k}$ = cohort size of the overall population.

This identity suggests that the sum of age-specific individual pension benefits $p_{b,k}$ (in the base year b of the cohort born in k) weighted by the cohort size $C_{b,k}$ should equal the corresponding overall aggregate pension expenditure, denoted by P_b.

It is necessary to consider the annual valorisation of pensions, i.e. adjustment of pensions according to the law amendments. In 2008–2017, this indicator was set as a sum of the rate of inflation in given year and one third of the real wage growth. Economic crisis in the year 2013 and 2014 was an exception from this rule as only one-third of the inflation rate was taken into account. Since 2018, half of real wage growth has been counted in as part of the valorisation. Based on these, we adjust the inflation rate and one third of wage growth which are discussed below. Formula (2) states that an individual already retired in base year b receives the same pension in a specific year t as in the base year b, only corrected by the indexation g of a pension in payment.

$$p_{t,k}^{exis} = p_{b,k}^{exis}(1+g)^{t-b}, \tag{2}$$

where:

g = indexation, according to established rules.

The age-sex-specific pension profile for future new retirees, which is the basis for the estimation of accrued-to-date entitlements, is calculated by adjustments of existing retirees` benefits in the base year. A new retirees` benefit $p_{t,k}^{new}$ in a specific year t of a cohort k is developed by calculating the absolute change in the benefit of the existing retirees of the cohort $b-(t-k)$ (the cohort with the same age $(t-k)$ in the base year b) to the cohort one year younger in the base year, namely $b-(t-1-k)$.

The following equation sums up the calculations of pension benefits for newly incoming beneficent in given future year t. The amount of future pensioners' allowances is then estimated from the current retirement benefits. The value of such benefit must be then adjusted by the coefficient of the general assessment base, according to which the newly awarded pensions are increased each year. The average of the last 8 years represents a value of 0.026 to 2015. In other words, the current claim shall be multiplied by the coefficient for each year remaining year before a beneficent reaches retirement age as shown in Formula (3).

$$p_{t,k}^{new} = \left[p_{b,b-(t-k)}^{exis} - p_{b,b-(t-1-k)}^{exis}\right]\cdot(1+v)^{t-b}, \tag{3}$$

where:

$p_{t,k}^{new}$ = new retiree`s benefit in a specific year t of a cohort k,
$p_{b,b-(t-k)}^{exis}$ = benefit of the existing retirees of the cohort $b - (t - k)$,
$p_{b,b-(t-1-k)}^{exis}$ = benefit of the existing retirees of the cohort one year younger,
v = valorisation rate.

For generations not yet retired which do not meet the criteria of the minimum time of participation in the social insurance program, pension benefits is not considered in its full amount. In this case, the "lambda" factor is introduced which represents a reduced entitlement to future pension claims. To put an example of participants borne in the year 1980, we consider only 15 years of social insurance out of a total of 40 years. Then, the primary income of this generation is reduced by the share 15/40. The following years include the same pension adjusted for the revaluation of pensions. We will make all these adjustments for all generations over the age of 20 (with the exception of the orphan's pension we are considering from 0 to 26 years of age, who do not yet have a retired pension (up to the age of 67 when we no longer consider retirement).

Lastly, the accumulated future benefits of new retirees need to be calculated. Thus, for example, a 55 year old representative (in the base year) will have a certain probability of retiring at the age of 56, 57 and so on. Formally, this is done by cumulating year-by year $p_{t,k}^{new}$ according to the following equation. The accumulated age-sex-specific future pension benefits $p_{t,k}^{fut}$ of a retiree for a specific year t of the cohort k are defined as follows:

$$p_{t,k}^{fut} = p_{t-1,k}^{fut}(1+g) + \lambda_{t,k} \cdot p_{t,k}^{new} \ , \tag{4}$$

where:

$p_{t,k}^{fut}$ = future pension benefits of a retiree for a specific year t of the cohort k,
λ = lambda factor, which represents a reduced entitlement to future pension claims,
$p_{t,k}^{new}$ = the pension benefits for new retirees in the base year.

From this equation it follows that an average individual born in year k receives a future benefit in year t (t>b) which consists of the accumulated pension payment one period earlier (t – 1) corrected by the pension indexation g plus the pension paid to new retirees in that year. Finally, the ADLs of the pension scheme are calculated by discounting and adding up the above projected pension benefits over the cohorts living in the base year as expressed by Formula (5).

$$ADL_b = \sum_{t=b}^{b+D} \sum_{k=b-D}^{b} \frac{\left(p_{t,k}^{exis} + p_{t,k}^{fut}\right)}{(1+r)^{t-b}} C_{t,k} \ , \tag{5}$$

where:

$ADLb$ = accrued-to-date liabilities of the base year b,
r = discount factor.

This means that, in every period t, the pension benefits of the existing retirees ($p_{t,k}^{exis}$) and the pension rights accrued until the base year ($p_{t,k}^{fut}$) – which are both discounted by the factor $(1 + r)$ for every future year $(t – b)$ – are multiplied by the number of members of this age cohort Ct, k. This is done for each age group, beginning with those born in $k = b – D$, which goes back 100 years prior to the base year.

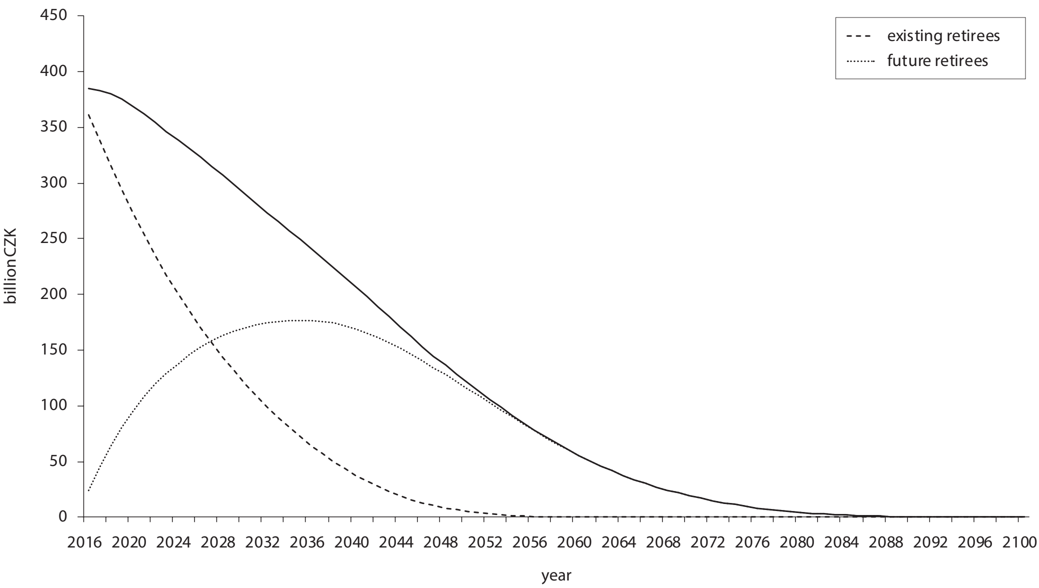
3 ANALYSIS

In this section, we will analyse the results obtained from the model just described. Figure 3 shows the comparison of pension liabilities arisen in the first pillar between the current and future pensioners as it was in 2015. The value of current pensioners' claims is approximately CZK 4.2 trillion amounting to 40 percent of all the pension liabilities. Most of the liabilities take the form of old-age pension claims

(84 percent). The number of expenditures falls over time due to a gradual decrease in the status of these cohorts and reaches a zero value by paying out the commitments to the last participant in the model (up to 100 years).

The claims of future pensioners who have not yet reached the age to benefit from the system offer a completely different picture. Due to a low lambda factor which determines the entitlements in the future, the liabilities are lower in the first years for the future pensioners as illustrated by the red line in Figure 2. As our analysis shows, the share of future retirees will peak around 2035 representing the balance between the most numerous generations retiring around this year and the high lambda coefficient. Similarly to the case of current pensioners, pension entitlements amounting to 82 percent takes the highest share in the total pension entitlements.

Figure 2 Share of pension liabilities for current and future pensioners, the Czech Republic, 2015



Source: CZSO

Let´s take a look at the development of the total pension liabilities in terms of GDP which is shown in Table 3. Over the analysed period, pension liabilities grew by more than CZK 700 billion. Major part of the total amount takes the form of the liabilities of old-age pensions (more than 80 percent in all years), while the smallest share belongs to the orphan's pensions (8 percent). From the gender point of view, we can draw a fall in the value of men's disability pensions (by CZK 50 billion), while in the case of women this decrease reached only CZK 10 billion. In total, the share of pension liabilities from the first pillar stood at 230 percent of GDP at the end of 2015.

Although there was an increase in the value of pension liabilities over this period, it had been continuously decreasing if expressed in terms of GDP. This is chiefly due to a strong growth in nominal GDP, especially in 2014–2015. This trend may be expected to continue also in next years as the Czech economy was strongly growing in both years 2016 and 2017. To sum up, while the total value is nominally growing, its share on GDP went down due to high nominal growth of the Czech economy.

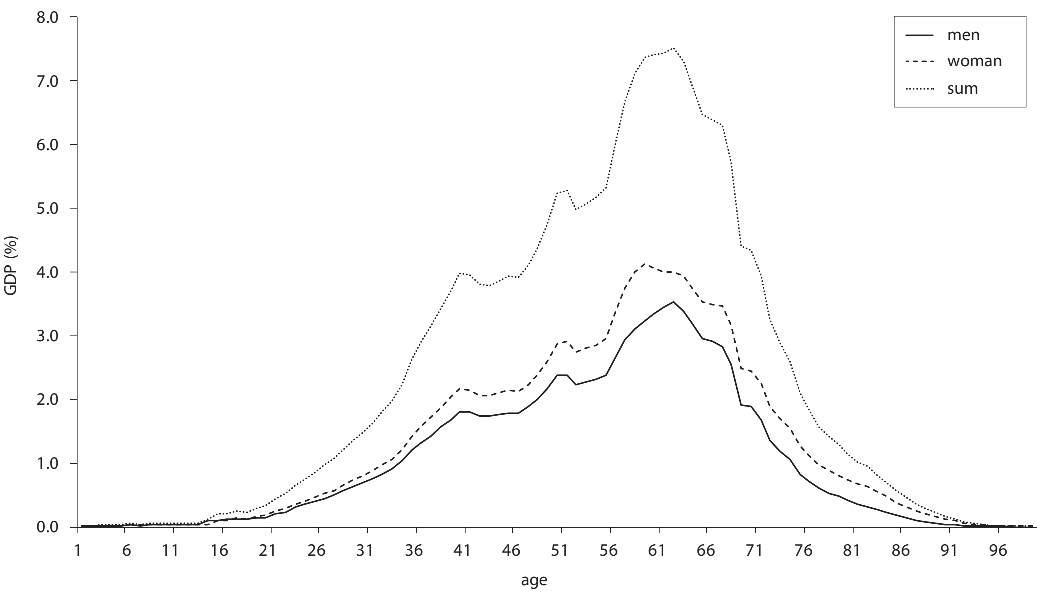
Table 3 Pension liabilities of the first pillar by sex and type, the Czech Republic, 2011–2015, CZK mil, percentage of GDP

Type of pension	Gender / Year	2011	2012	2013	2014	2015
Survivor	Women	699	744	745	735	704
	Men	110	122	125	126	125
Disability	Women	460	462	454	451	445
	Men	545	550	537	519	502
Old age	Women	4 368	4 492	4 548	4 627	4 706
	Men	3 626	3 843	3972	3 962	4 067
Total		9 808	10 213	10 381	10 420	10 549
Percentage of GDP		243	251	253	242	230

Source: CZSO

Pension liabilities can be, of course, analysed from many other perspectives. Let´s move on to the structure of liabilities by age and gender. As illustrated in Figure 3, the largest proportion of both male and female commitments falls into the age group of 60–64 years (approximately 20% of all commitments). Evidently, the amount of liability is growing over the working age for both men and women. Observable exceptions are the age groups from 41 to 45 which may be attributed to a lower number of people in these cohorts. From the age of 65, the amount is on declining path as the number of people in these

Figure 3 Pension liabilities by age and sex, percentage of GDP, the Czech Republic, 2015



Source: CZSO

cohorts is decreasing. From the gender point of view, similar trends are observable. However, higher values are identifiable in all age groups compared to men, mainly due to larger sizes of generations in terms of number of people.

Let´s make a step beyond the headline figures. The values presented so far were based on the simplistic assumption on the 5-percent discount rate as requested by the Manual (Eurostat-ECB, 2011). Because the future conditions can change or a particular situation in individual countries can differ, statisticians are expected to provide the sensitivity analysis showing the alternative results at different level of discount rate. By doing so, we can better assess the extent of sensitivity, not least the importance of discount factor in the compilation process.

Table 4 summarizes the results of the sensitivity analysis for the Czech Republic.

Table 4 Sensitive analysis, the Czech Republic, 2015

Liabilities of the pay-as-you-go system	Discount rate 4% (nominal)	Discount rate 5% (nominal)	Discount rate 6% (nominal)
Closing stock of pension liabilities (CZK bn)	12 693	10 549	8 928
Level of pension entitlement as a share of GDP (%)	276.2	229.5	194.3

Source: CZSO, own calculation

Not surprisingly, the lower discount rate used the higher share of pension entitlements on GDP is, as future payments take higher present values. Furthermore, we can see that the sensitivity on the discount rate is really significant. The total value for the lowest discount rate (4 percent) is by 82 percentage points higher than that for the 5 percent discount rate.

High volatility was found out also in other empirical studies dealing with this issue. For example, in the study made by Langhamrova, Sixta and Simkova (2016), the relative amount of pension liabilities for the year 2014 ranges from 679 percent of GDP (if 1-percent discount rate is used) to 230 percent of GDP for 5-percent discount rate)[10] depending on a particular discount rate. To compare the results of this study with the data published by the CZSO, only small differences can be found. Using identical assumption, i.e. 5 percent discount rate and 2 percent growth in wages, the officially published ratio is by 12 percentage points higher than that calculated in the study of Langhamrova, Sixta and Simkova.

4 INTERNATIONAL COMPARISON

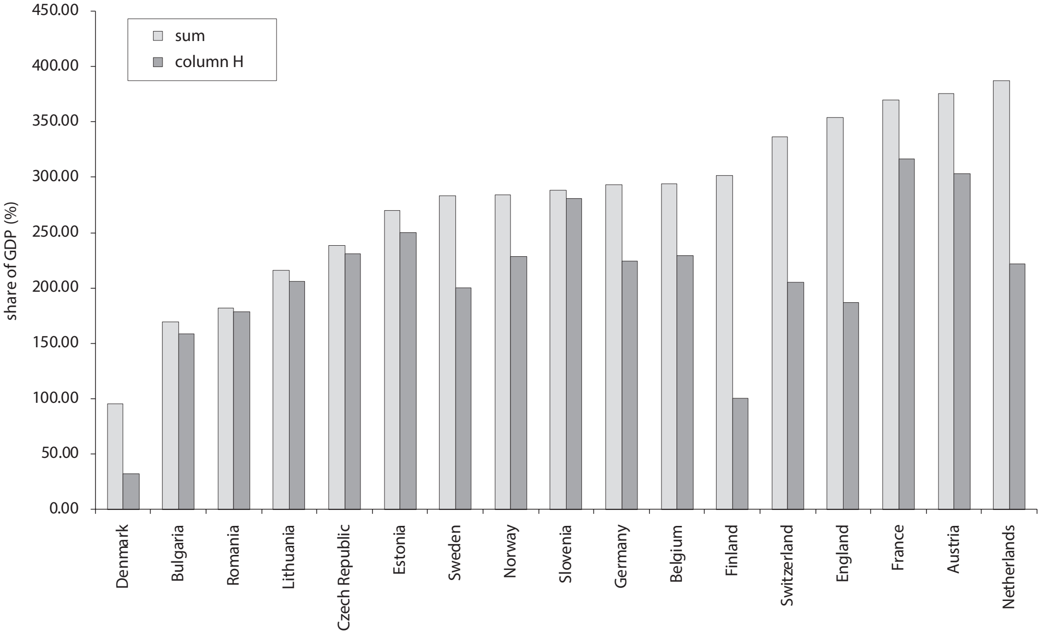
Finally, we will take a look at the pension liabilities in the international context. Figure 4 illustrates the role which the PAYG system (the column H in the Table 29)[11] plays in individual countries. The lowest shares were reported by Finland, Netherland or Denmark. On the other hand, the unfunded system dominates in Eastern Europe, i.e. in countries like Romania, Bulgaria or the Czech Republic where the share of the PAYG system reached almost 97 percent of all pension liabilities.

Let´s take a look at the PAYG system and its annual changes in Figure 5. In most cases, there are no large differences between year t and $t + 1$. One of the reasons is that it is a long-term projection of an accumulated value of pensions in the base year that will be paid over coming decades. More substantial annual changes are thus mainly driven by economic cycle and the development in nominal GDP. This is exactly the case of the Czech Republic where the share decreased by 6 percentage points between the years 2014 and 2015.

[10] These values correspond to the assumption of 2-percent growth in wages as is the case of the calculation made by the CZSO

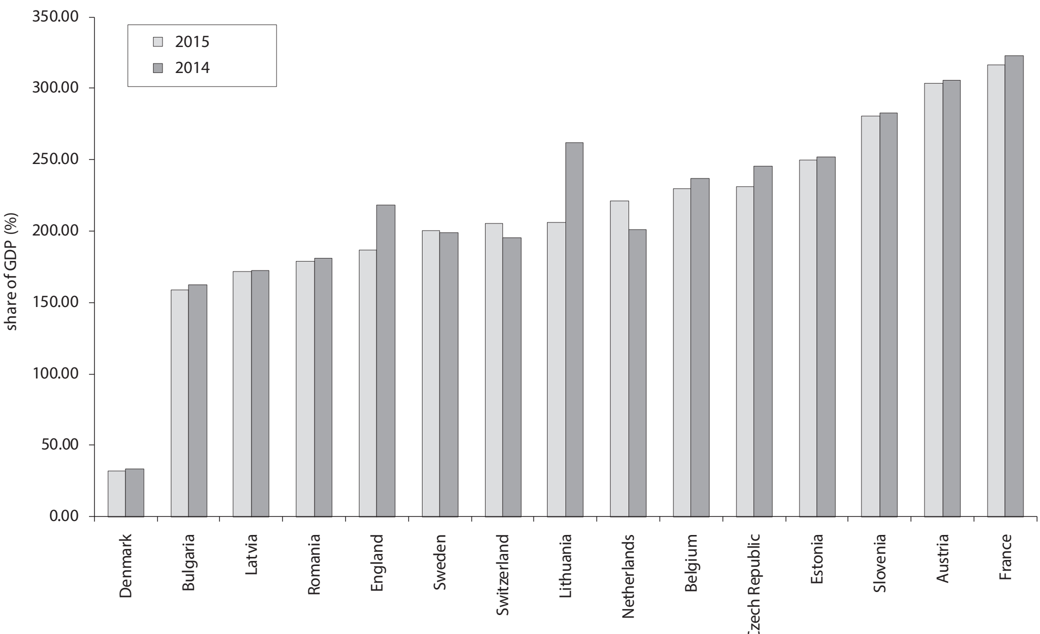
[11] Column H in the respective table concerns the social security pension schemes which are not an integral part of the core national accounts.

Figure 4 The total amount of pension liabilities compared to Column H of Table 29 (2015)



Source: CZSO, Eurostat, own calculation

Figure 5 Table 29 – Pension Entitlements in the column H, % of GDP (2014–2015)



Source: CZSO, Eurostat, own processing

In the first place, Figure 5 clearly reveals the extent of pension liabilities from the operation of the PAYG system (column H in Table 29) in individual countries. The 300-percent threshold in terms of GDP was exceeded in France and Austria. On the other hand, the lower value was reported by Denmark where, however, the funded scheme takes a larger share in pension protection as shown in Figure 5. For most of the countries, the share ranges between 150 percent and 250 percent of GDP. As we can draw on the table, the Czech Republic ranks among the countries with rather higher share of the PAYG system´s liability on GDP.

CONCLUSION

As we have shown, the pension liabilities of the Czech Republic are primarily associated to the operation of the unfunded pension scheme (PAYG system, the first pillar). To quantify the corresponding liability, the Freiburg model ADL was used where the future households' claims are determined for the current base year. Using particular assumption on the macroeconomic and demographic development in the future, the pension liabilities from the operation of the first pillar in the Czech Republic reached 230 percent of GDP at the end of 2015.

As we have tried to show, the finals are highly sensitive to the choice of relevant discount rate. Discount rate of 4 percent, i.e. by 1 percentage point lower than that used for the headline figure calculation, would bring the pension liabilities up to 276 percent of GDP. The international comparison further revealed that the unfunded scheme played quite exceptionally major role in pension protection in the Czech Republic. However, the total shares in terms of GDP, capturing both funded and unfunded schemes, are not significantly different from other countries analysed in this paper.

Very finally, we can only reiterate that values on the pension liabilities do not themselves indicate on the sustainability of the pension system as the asset-side of the system is entirely left out. On the other hand, they can point to a potential risk for the government finance if future liabilities are not to be matched by corresponding payments of contributors. Thus, although pension liabilities are not a debt in a strict sense, because it does not originate in money borrowing, they clearly represent an obligation to future generation.

References

BLOCH, D. AND FALL, F. *Government Debt Indicators: Understanding the Data.* Economic Department Working Papers No. 1228, OECD, 2015.

CZECH STATISTICAL OFFICE. *Main Economic Indicators* [online]. Prague: CZSO, 2017. <*https://www.czso.cz/csu/czso/hmu_cr*>.

CZECH STATISTICAL OFFICE. *Population projection of the Czech Republic until 2100* [online]. Prague: CZSO, 2013. <*https://www.czso.cz/csu/czso/projekce-obyvatelstva-ceske-republiky-do-roku-2100-n-fu4s64b8h4*>.

EUROSTAT, EUROPEAN CENTRAL BANK. *Technical compilation Guide for Pension Data in National accounts* [online]. Eurostat Methodologies and Working papers, 2011. <*https://www.ecb.europa.eu/pub/pdf/other/techn_comp_gd_pens_dt_nat_accts_201201en.pdf*>.

EUROSTAT. *European system of accounts* [online]. Publications Office of the European Union, 2013. <*http://ec.europa.eu/eurostat/documents/3859598/5925693/KS-02-13-269-EN.PDF/44cd9d01-bc64-40e5-bd40-d17df0c69334*>.

EUROSTAT. *The 2015 Ageing report: Economic and budgetary projections for the 28 EU Member States (2013–2060)* [online]. Publications Office of the European Union, 2015. <*http://ec.europa.eu/economy_finance/publications/european_economy/2015/pdf/ee3_en.pdf*>.

EUROSTAT. *Statistical Database* [online]. Eurostat, 2018. <*http://ec.europa.eu/eurostat/web/pensions/other-information*>.

GOEBEL, P. *A Primer on Government-Sponsored Pension Schemes in the National Accounts and Their Impact on the Interpretations of Government Debt Statistics.* Working Paper No. 81, OECD, 2017.

LANGHAMROVÁ, J., SIXTA, J., ŠIMKOVÁ, M. Penzijní závazky stárnoucí populace České republiky. *Politická ekonomie*, 2016, Vol. 64, No. 5, pp. 591–607. DOI: https://doi.org/10.18267/j.polek.1090.

MINISTRY OF SOCIAL AND LABOUR AFFAIRS. *Český důchodový systém* [online]. Prague, MoLSA, 2011. *<http://www.mpsv.cz/files/clanky/11969/Analyza.pdf >*.

RUFFING, K., VAN DE VATER, P., KOGAN, R. *"Generational Accounting" is Complex, Confusing, and Uninformative.* Washington, DC: Center on Budget and Policy Priorities, 2014.

Compilation of Physical Supply and Use Tables as a Tool for Increasing Analytical Potential of Economy-Wide Material Flow Analysis and Indicators

Jan Kovanda [1] | *Charles University Environment Centre, Prague, Czech Republic*

Abstract

The study described in this article presents the first physical supply and use tables (PSUT) ever based on the recently published methodological standard for System of Environmental-Economic Accounting (SEEA). The tables were compiled for the Czech Republic for 2014 and can be used for increasing analytical potential of the economy wide material flow analysis and indicators. The subsequent compilation procedure was described in detail so that it can serve as a source of inspiration and a benchmark for other researchers and/or statisticians. The major shortcoming of the PSUT is that not all needed data was readily available in physical units and required estimations based on proxies. Some parts of the tables are therefore burdened with a degree of uncertainty.

Keywords

System of environmental-economic accounting (SEEA), economy-wide material flow analysis (EW-MFA), material flow indicators, physical supply and use tables (PSUT), Czech Republic

JEL code

Q56

INTRODUCTION

The overall environmental pressure and impact caused by human societies is to a large extent induced by the consumption of energy and materials (e.g. Ayres and Simonis, 1994; Weizsäcker et al., 2009; Fischer-Kowalski et al., 2011). The relevant pressures and impacts include structural landscape changes, loss of biodiversity, acidification, eutrophication, global climate change and others (Giljum et al., 2005). In order to measure material and energy flows and to mitigate the related problems, material flow analysis has been developed. The aim of this accounting and analytical approach is to monitor material and energy flows at various levels of detail,

[1] Charles University Environment Centre, J. Martiho 2/407, 162 00 Prague 6, Czech Republic. E-mail: jan.kovanda@czp.cuni.cz, phone: (+420)220199476.

and to provide indicators which contribute to the management of resource use and output emission flows from both economic and environmental points of view (OECD, 2008). As convenient measures of sustainability, material flow indicators focusing mostly on an economy-wide level have been compiled for a range of both developed and developing countries (e.g. Adriaanse et al., 1997; Matthews, 2000; Giljum, 2004; Eurostat, 2015; Schandl and West, 2012; Gonzalez-Martinez and Schandl, 2008).

The economy-wide material flow analysis (EW-MFA) treats the economy as a black-box devoted exclusively to monitoring overall input and output flows while inter-industry physical flows are neglected. In order to increase the analytical potential of this tool, it is advisable to construct a physical input-output table (PIOT) which shows the input of raw materials and products by industries, inter-industry deliveries of products and a breakdown of output products and waste residues by industries. Data from PIOT can be used to analyse physical flows, considering the economic activities and structural changes that lie behind these flows, and to construct industry-specific waste or material accounts based on the material balance principle. They can further be used to generate information on the raw material equivalents of final demand; to analyse technological change, material substitution, shifts in consumption and to assess the effectiveness of policies targeting sustainable consumption and production (OECD, 2008). The implementation of PIOT is a labour-intensive task involving many data entries. This is the reason why it has only been compiled for a few countries so far including Denmark (Mulalic, 2007), Finland (Mäenpää, 2004), Germany (Stahmer et al., 1997), Italy (Nebbia, 2000), New Zealand (McDonald and Patterson, 2006), Spain (Gasco et al., 2005), the Netherlands (Hoekstra and Van den Bergh, 2006) and the EU (Giljum and Hubacek, 2004). Moreover, as no standardized approach for PIOT compilation was available until 2014, the above studies use different approaches and the resulting PIOTs thus have different formats and are not fully comparable.

The aim of this article is to compile physical supply and use tables (PSUT) for the Czech Republic for 2014 with the use of recently published methodological standard for System of Environmental-Economic Accounting (SEEA) (United Nations et al., 2014). These PSUT can be transformed to any type of PIOT by standard procedures prescribed by Eurostat (2008). As far as we know no PSUT based on SEEA have been published yet. The compilation of PSUT can be understood as an extension of and addition to the compilation of supply and use tables and input-output tables in monetary units which has had a long tradition at the Czech Statistical Office. The tables in monetary units are regularly published at: <*http://apl.czso.cz/pll/rocenka/rocenka.indexnu*> (Czech Statistical Office, various years-a).

The rest of the article is structured as follows: section 1 describes in detail the procedure of compilation of PSUT which might serve as an inspiration and methodological benchmark for other countries. This section is the core part of the article. Section 2 presents results for PSUT while section 3 discuses the benefits and shortcomings of their compilation. Last section concludes with lessons learnt.

1 METHODS AND DATA COMPILATION

1.1 General description of physical supply and use tables

SEEA (United Nations et al., 2014) focuses on the compilation of physical supply and use tables (PSUT) which are the building stones of PIOT rather than on the compilation of PIOT itself. The reason for this is that there is a standard procedure for transforming PSUT into any type of PIOT (product-by-product, sector-by-sector) (Eurostat, 2008) while the procedure for PSUT compilation has not been standardized yet.

PSUT are composed of a sequence of detailed tables. Their structure is shown in Figure 1.

Figure 1 Structure of physical supply and use tables (PSUT)

Supply table

	Production; Generation of residuals		**Accumulation**	**Flows from the rest of the world**	**Flows from the environment**	**Total**
	Production; Generation of residuals by industries	**Generation of residuals by households**	**Industries**			
Natural inputs					A. Flows from the environment	Total supply of natural inputs
Products	C. Output			D. Imports of products		Total supply of products
Residuals	I1. Residuals generated by industry I2. Residuals generated following treatment	J. Residuals generated by households	K1. Residuals from scrapping and demolition of produced assets K2. Emissions from controlled landfill sites	L. Residuals received from the rest of the world	M. Residuals recovered from the environment	Total supply of residuals
Total supply						

Supply table

	Intermediate consumption of products; Use of natural inputs; Collection of residuals	**Final consumption**	**Accumulation**	**Flows to the rest of the world**	**Flows to the environment**	**Total**
	Industries	**Households**	**Industries**			
Natural inputs	B1. Extraction of natural inputs used in production B2. Extraction of natural resource residuals					Total use of natural inputs
Products	E. Intermediate consumption	F. Household final consumption	G. Gross capital formation (including fixed assets and inventories)	H. Exports of products		Total use of products
Residuals	N. Collection and treatment of residuals		O. Accumulation of waste in controlled landfill sites	P. Residuals sent to the rest of the world	Q1. Direct residual flows from industry and households	0.166
Total use						

Source: United Nations et al., 2014

PSUT contains several categories of flows in the rows: Flows from the environment into the economy are natural inputs, flows within the economy consist of either products or residuals, and flows from the economy to the environment are residuals. Natural inputs which are not used in production such as mining overburden are called natural resource residuals. They correspond to unused extraction in EW-MFA (Eurostat, 2001).

The columns of the PSUT are structured to reflect both the activity underlying the flows and the economic units involved. The second column covers the use of natural inputs, the production and intermediate consumption of products, and the generation and receipt of residuals by all enterprises in the economy. The third column covers the consumption of products by households and the generation of residuals from this consumption. The activity of households in extracting and collecting natural inputs from the environment for their own consumption is a productive activity and is recorded in the second column under a relevant industry class. Unlike the monetary supply and use table, no entries in physical terms are made in relation to government final consumption expenditure and final consumption expenditure of non-profit organizations. All physical flows related to the intermediate consumption and generation of residuals by governments and non-profit organizations are recorded in the first column under the relevant industry class (commonly, public administration). The fourth column covers changes in the stock of materials in the economy. From a supply perspective, this column records reductions in the physical stock of produced assets through demolition or scrapping. From a use perspective, the accumulation column records additions to the physical stock of produced assets (gross capital formation). The fifth column recognizes the exchanges between national economies in the form of imports and exports of products and flows of residuals. Residuals received from the rest of the world and residuals sent to the rest of the world primarily relate to the movement of solid waste between different economies. The sixth column is a significant addition to the monetary supply and use table structure and records flows to and from the environment. The incorporation of the environmental column allows a full accounting for flows of natural inputs and residuals which would otherwise not be possible (United Nations et al., 2014).

The PSUT contains a range of accounting and balancing identities. The two most important identities include supply and use identity which indicates that the total supply of products (i.e. the sum of Tables C and D) is equal to the total use of products (sum of Tables E, F, G and H). The input-output identity implies that flows of materials into an economy (sum of Tables A, D, L and M) are equal to the flows of materials out of an economy (sum of tables Q, H, P) plus any net additions to stock in the economy (sum of Tables G and O minus Table K) (United Nations et al., 2014). The above identities are very useful when compiling PSUT. When some underlying data are missing, these identities can be used for their estimation.

1.2 Procedure of compilation of PSUT for the Czech Republic

The Czech PSUT were compiled for 2014 and on the level of NACErev2/CPA[2] for 88 industries and product groups (see Supplementary Information 1 for the full list of the industries and product groups – Annex 1). The resulting tables are shown in section 2 while this section describes in detail the procedure of PSUT compilation. The selection of year 2014 was driven by the fact that data needed for the compilation is available with 2–3 year delay.

The aim of the project was to base PSUT on the economy-wide material flow accounts currently available in the Czech Republic (Kovanda et al., 2010; Czech Statistical Office, 2017; Kovanda, 2018). This required some adjustments of the SEEA framework. For instance, landfilled wastes were considered flows out of the economy, which led to inclusion of landfilled wastes in Table Q and skipping Tables K2 and O. The Czech EW-MFA system is fully balanced. It means that it includes so called balancing items such as the oxygen needed for the combustion of fossil fuels by industries and households or water vapor from combustion (Eurostat, 2001). In order to balance PSUT these balancing items had to be included as well.

[2] NACE: statistical classification of economic activities in the European Community; CPA: classification of products by activity.

They were ranked among natural inputs and residuals as other natural inputs and other residuals and incorporated into Tables A, I, J, B and Q. The Tables further included natural resource residuals (unused extraction) which were incorporated into Tables A, I, B and Q.

The procedure of compilation of PSUT was as follows:

1) Compilation of data readily available from statistics

Data from statistics was readily available for Tables A. Flows from the environment, D. Imports of products, J. Residuals generated by final household final consumption, L. Residuals received from the rest of the world, M. Residuals recovered from the environment, H. Export of products, N. Collection and treatment of residuals, P. Residuals sent to the rest of the world and Q. Residual flows to the environment. Major sources of data were the database on full economy-wide material flow accounts for the Czech Republic (Kovanda, 2018), Czech Statistical Office and its datasets on extraction of raw materials (Czech Statistical Office, 2017), foreign trade (Czech Statistical Office, various years-b), waste management (Czech Statistical Office, 2015a) and emissions to water (Czech Statistical Office, 2015c). Other sources included the Czech Hydrometeorological Institute and its data on emissions to air (Czech Hydrometeorological Institute, various years) and the Czech Geological Survey with data on natural resource residuals (Czech Geological Survey, 2015). Table M. was set equal to zero, as no residuals are recovered from the environment in the Czech Republic (Czech Statistical Office, 2015a).

2) Table C. Output of products

The estimation of Table C. Output of products comprised the following steps:

a) Data on production of industrial products

The Czech Statistical Office maintains a dataset on the production of industrial products by industries (Czech Statistical Office, 2015b). Various entries, however, are not expressed in physical units, but in monetary units. These entries were transformed into physical units using unit prices in CZK per tonne calculated with the help of export statistics which are available both in physical and monetary units.

b) Data on production of agriculture product

This data was taken from the Czech Statistical Office datasets on extraction of raw materials (Czech Statistical Office, 2017) and attributed to their native industries, i.e. NACE 01-03.

c) Data on the production of buildings and transport infrastructures

No data was found which would give any leads to the production of buildings and transport infrastructures in physical units. We calculated this item later on in step 6 under the assumption that all these infrastructures go to accumulation (Table G).

3) Table B. Extraction of natural inputs

Extraction of natural inputs used in production was set equal to the production of natural inputs in Table C (NACE 01-08) assuming no losses between extraction and the production of natural inputs for further use. The accompanying natural resource residuals such as overburden from mining were disaggregated by sectors proportionally to the extraction of corresponding natural inputs. Disaggregation of other natural inputs was straightforward for items such as consumption of oxygen and water by livestock which was attributed to NACE 01 Agriculture. Consumption of oxygen and nitrogen for combustion was disaggregated according to the combustion of fuels by particular industries provided by Ministry of Industry and Trade (Ministry of Industry and Trade, 2015).

4) Tables I. Residuals generated by industries and following treatment

Waste and emissions generated by industries and following treatment were available from waste, water emission and air emission statistics, but on a lower level of sectoral disaggregation than needed. The detailed disaggregation of data into 88 industries was achieved with the use of relationships from

Table C for industrial products and detailed monetary supply tables for services (Musil, 2017). Natural resource residuals by industries were set equal to consumption of natural resource residuals in Table B, as they are only extracted by particular sectors and then returned back to the environment without entering the production and consumption processes of other sectors. Other residuals generated by sectors were disaggregated using similar approaches to the disaggregation of other natural inputs in Table B.

5) Tables E. Intermediate consumption, F. Household final consumption and G. Gross capital formation
Data on intermediate consumption, household final consumption and gross capital formation was not available in physical units at all. The attribution of produced and imported commodities (Tables C and D) to industries, households and accumulation was therefore based on relationships in the monetary use tables. Detailed use tables compiled separately for domestic production and imports and disaggregated by 204 industries and product groups (Musil, 2017) were used for this attribution. The detailed results were then aggregated back into a single table broken down by 88 industries and product groups. The use of detailed tables addressed the inhomogeneity issue mentioned in the Introduction at least in part. It was assumed that this approach would produce quite severe inhomogeneity on the level of 88 industries and product groups, as e.g. crude oil and natural gas or industrial non-metallic minerals and construction non-metallic minerals composed one product group. On the other hand, they were treated separately in the detailed tables.

After this calculation step, Table G did not contain accumulation of buildings and transport infrastructures, as their production was not shown in Table C (see step 2c).

6) Accumulation of buildings and transport infrastructures in Table G and production of buildings and transport infrastructures in Table C
An estimation of these entries was based on the overall input-output balance. Flows of materials into and out of the economy (Tables A, D, L, M, Q, H and P) were compiled in step 1 while total K1 was available from the Czech and Eurostat trade statistics (Czech Statistical Office, 2015a; Eurostat, various years). Table G was compiled in step 5, but it did not contain accumulation of building and transport infrastructures: the missing part of Table G was thus calculated from the overall input-output balance. Under the assumption that total production of building and transport infrastructures is assigned to accumulation, the calculated figure had to be equal to the total production of building and transport infrastructures in Table C as well. This total production was further split by CPA 41 and 42 and particular industries in Table C with the use of relationships in detailed monetary supply tables (Musil, 2017).

7) Table K1. Residuals from scrapping and demolition of produced assets
This data was available on a lower level of disaggregation than that necessary in Czech and Eurostat waste statistics (Czech Statistical Office, 2015a; Eurostat, various years). We based the further disaggregation on accumulation figures (Table G. Gross capital formation) assuming that the higher accumulation is, the higher the volume of discarded infrastructures also is, because accumulation replaces discarded infrastructures and equipment.

After this last step, the full sequence of physical supply and use tables as prescribed by SEEA framework (United Nations et al., 2014) was available for the Czech Republic and the tables contained both supply and use and input-output identities. The presence of these identities was caused by the calculation methods of Tables E, F and G and buildings and transport infrastructures in Table C.

2 RESULTS

Tables 1 and 2 show aggregated physical supply and use tables for the Czech Republic for 2014. The full PSUT in tonnes are shown in Supplementary Information 2 (Annex 2 is available

Table 1 Physical supply table (million tonnes), Czech Republic, 2014

Particular tables		Production; Generation of residuals					Accumulation	Flows from the rest of the world	Flows from the environment	Total
		NACE 01-36	NACE 37-39	NACE 41-43	NACE 45-99	Households				
1	Natural inputs								559.8	559.8
A	Flows from the environment								559.8	559.8
A1	Natural resource inputs								157.7	157.7
A2	Natural resource residuals								257.5	257.5
A3	Other natural inputs								144.7	144.7
2	Products	323.6	5.7	60.7	15.9			71.9		477.8
C	Output of products	323.6	5.7	60.7	15.9					405.8
D	Imports of products							71.9		71.9
3	Residuals	392.4	10.2	19.2	23.5	48.0	4.6	1.4	0	499.3
I	Residuals generated by industries and following treatment	392.4	10.2	19.2	23.5					445.3
I1	Residuals generated by industries	106.0	3.2	1.7	17.7					128.6
I2	Residuals generated following treatment		7							6.5
I3	Natural resource residuals generated by industries	240.0	0.4	16.9	0.2					257.5
I4	Other residuals generated by industries	46.4	0.2	0.6	5.5					52.7
J	Residuals generated by household final consumption					48.0				48.0
J1	Wastes generated by household final consumption					19.3				19.3
J2	Other residuals generated by household final consumption					28.7				28.7
K	Residuals from scrapping and demolition of produced assets						4.6			4.6
L	Residuals received from rest of the world							1.4		1.4
M	Residuals recovered from the environment								0	0
4	Total supply	716.0	15.9	79.9	39.3	48.0	4.6	73.4	559.8	

Source: Own calculation

at the webpage of *Statistika: Statistics and Economy Journal*, see the online version of No. 3/2018 (Excel file) at: <*http://www.czso.cz/statistika_journal*>).

Table 2 Physical use table (million tonnes), Czech Republic, 2014

Particular tables		Intermediate consumption of products; Use of natural inputs; Collection of residuals				Final consumption	Accumulation	Flows from the rest of the world	Flows from the environment	Total
		NACE 01-36	NACE 37-39	NACE 41-43	NACE 45-99	Households				
1	Natural inputs	524.2	1.1	18.5	16.0	31.0				590.8
B	Extraction of natural inputs	524.2	1.1	18.5	16.0					559.8
B1	Extraction of natural inputs used in production	157.6	0.1	0.0	0.0					157.7
B2	Extraction of natural resource residuals	240.0	0.4	16.9	0.2					257.5
B3	Extraction of other natural inputs	126.6	0.7	1.6	15.8					144.7
2	Products	210.8	7.4	49.9	26.2	31.0	85.7	66.7		477.8
E	Intermediate consumption	210.8	7.4	49.9	26.2					294.4
F	Household final consumption					31.0				31.0
G	Gross capital formation						85.7			85.7
H	Exports of products							66.7		66.7
3	Residuals		20.4					4.0	481.4	505.7
N	Collection and treatment of residuals		20.4							20.4
P	Residuals sent to the rest of the word							4.0		4.0
Q	Residual flows to the environment								481.4	481.4
Q1	Direct flows of residuals								136.0	136.0
Q2	Flows of residuals following treatment								6.5	6.5
Q3	Flows of natural resource residuals								257.5	257.5
Q4	Flows of other residuals								81.4	81.4
4	Total use	735.1	28.9	68.4	42.2	61.9	85.7	70.7	481.4	

Source: Own calculation

Total supply of products to the Czech economy was 477.8 million tonnes in 2014 (Table C + Table D) which was equal to total use of products (sum of Tables E, F, G and H). The total input of materials into the economy was 633.2 million tonnes (sum of Tables A, D, L and M) which was equal to the total output of materials (sum of Tables Q, H, P and G minus table K1).

3 DISCUSSION

The compilation of physical supply and use tables increased the analytical potential of the Czech EW-MFA accounts by breaking down the input and output material flows by industries. Other benefits which were not available from the original EW-MFA include the data in physical units on output of products, on intermediate consumption and distinguishing between residuals generated by industries and after treatment and between residuals from production and from scrapping and demolition. A significant advantage of the physical supply and use tables is their compatibility with accounts in monetary units. This allows for calculation of indicators which combine physical and monetary values such as material intensities and productivities of particular sectors. Last but not least the PSUT can be used for calculation of materials embodied in foreign trade and RMI and RMC indicators.

The major shortcoming of the PSUT compiled for the Czech Republic is that by far not all the necessary data was readily available in physical units and required estimations based on proxies. The most prominent examples are Tables E, F and G on intermediate consumption, final consumption of households and gross capital formation. These Tables were estimated with the use of detailed monetary use tables and it can be expected that the results will be burdened by some uncertainties. This is because even on the detailed level of 204 sectors and product groups the particular product groups were not fully homogenous and thus did not fully fulfil the requirement for homogenous sectoral prices of commodity outputs, imports and exports. Other data which was not available at all were production of buildings and transport infrastructures and their accumulation. These data were estimated with the use of overall input-output balance which has both advantages and shortcomings. The major advantage is that the whole system is balanced thanks to the calculation procedure. The shortcoming is that all inaccuracies present in the tables which compose the balance are propagated in the Tables used for the final balancing.

The full sectoral breakdowns were not available for some other Tables, namely Tables I and K1. Here at least the values for total economy and partial sectoral disaggregation were available before applying proxies for full disaggregation. It can therefore be expected that these Tables are affected by smaller uncertainties compared to Tables E, F and G and data on the production of buildings and transport infrastructures.

CONCLUSIONS

In the study described in this article we compiled the first ever physical supply and use tables (PSUT) based on the recently published methodological standard for System of Environmental-Economic Accounting (SEEA) (United Nations et al., 2014). The tables were compiled for the Czech Republic for 2014 and can be used for increasing analytical potential of the economy wide material flow analysis and indicators such as allowing for depicting productivities of particular sectors and calculating materials embodied in foreign trade. We thoroughly described the procedure we followed so that it can serve as a source of inspiration and a benchmark for researchers and/or statisticians who would wish to compile PSUT for other countries. The major shortcoming of the PSUT is that by far not all needed data was readily available in physical units and thus required estimations based on proxies. For this reason some parts of the tables can be burdened by quite large uncertainties. The most significant examples include the Tables E, F and G on intermediate consumption, final demand of households and gross capital formation and production of buildings and transport infrastructures in Table C. The author of this study is in regular contact with statisticians at the Czech Statistical Office (CZSO) responsible for material flow accounting. We plan to hand over our findings from PSUT compilation to them so that they can be reflected in regular statistical surveys arranged by CZSO. This should make the compilation of PSUT much easier in the future when new surveys on missing data are introduced.

ACKOWLEDGMENT

This work was supported by the Czech Science Foundation under grant No. 16-00262S "Anthropogenic material flows in the Czech Republic: Analysis of structure and trends and opening a black-box of the physical economy". This support is gratefully appreciated.

References

ADRIAANSE, A. et al. *Resource flows: The material basis of industrial economies.* Washington, D.C.: World Resources Institute, 1997.
AYRES, R. U. AND SIMONIS, U. E. *Industrial metabolism: Restructuring for sustainable development.* Tokyo, New York: United Nations University Press, 1994.
CZECH GEOLOGICAL SURVEY. *Balances and inventory of deposits of raw materials* Prague: Czech Geological Survey, 2015.
CZECH HYDROMETEOROLOGICAL INSTITUTE. *Emission balances o the Czech Republic* [online]. Various years. [cit. May 2017] <*http://portal.chmi.cz/files/portal/docs/uoco/oez/emisnibilance_CZ.html*>.
CZECH STATISTICAL OFFICE. *Generation, recovery and disposal of waste 2014.* Prague: Czech Statistical Office, 2015a.
CZECH STATISTICAL OFFICE. *Production of Selected Industrial Products – 2014.* Prague: Czech Statistical Office, 2015b.
CZECH STATISTICAL OFFICE. *Water supply systems, sewerage and watercourses – 2014.* Prague: Czech Statistical Office, 2015c.
CZECH STATISTICAL OFFICE. *Material flows accounts, (selected indicators) – 2016.* Prague: Czech Statistical Office, 2017.
CZECH STATISTICAL OFFICE. *Annual national accounts database* [online]. Various years-a. [cit. May 2017] <*http://apl.czso.cz/pll/rocenka/rocenka.indexnu*>.
CZECH STATISTICAL OFFICE. *Database on foreign trade* [online]. Various years-b. [cit. May 2017] <*http://apl.czso.cz/pll/stazo/STAZO.STAZO*>.
EUROSTAT. *Economy-wide material flow accounts and derived indicators: A methodological guide.* Luxembourg: Eurostat, 2001.
EUROSTAT. *Eurostat manual of supply, use and input-output tables.* Luxembourg: Eurostat, 2008.
EUROSTAT. *Sustainable development in the European Union: 2015 monitoring report of the EU sustainable development strategy.* Luxembourg: Eurostat, 2015.
EUROSTAT. *Eurostat statistics database* [online]. Various years. [cit. May 2017] <*http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/search_database*>.
FISCHER-KOWALSKI, M. et al. *Decoupling natural resource use and environmental impacts from economic growth. A Report of the working group on decoupling to the International Resource Panel.* Geneva: United Nations Environment Programme, 2011.
GASCO, G., HERMESILLA, D., GASCO, A., NAREDO, J. M. Application of a Physical Input-Output Table to evaluate the development and sustainability of continental water resources in Spain. *Environmental Management*, 2005, 36, pp. 59–72.
GILJUM, S. Trade, material flows and economic development in the South: The example of Chile. *Journal of Industrial Ecology*, 2004, 8(1–2), pp. 241–261.
GILJUM, S., HAK, T., HINTERBERGER, F., KOVANDA, J. Environmental governance in the European Union: strategies and instruments for absolute decoupling. *International Journal of Sustainable Development*, 2005, 8(1–2), pp. 31–46.
GILJUM, S. AND HUBACEK, K. Approaches of physical input–output analysis to estimate primary material inputs of production and consumption. *Economic Systems Research*, 2004, 16(3), pp. 301–310.
GONZALEZ-MARTINEZ, A. C. AND SCHANDL, H. The biophysical perspective of a middle income economy: Material flows in Mexico. *Ecological Economics*, 2008, 68(1–2), pp. 317–327.
HOEKSTRA, R. AND VAN DEN BERGH, J. C. J. M. Constructing physical input-output tables for environmental modeling and accounting: Framework and illustrations. *Ecological Economics*, 2006, 59(3), pp. 375–393.
KOVANDA, J., WEINZETTEL, J., HAK, T. Material flow indicators in the Czech Republic in light of the accession to the European Union. *Journal of Industrial Ecology*, 2010, 14(4), pp. 650–665.
MÄENPÄÄ, I. *Physical flow accounts, Finland 1999.* Oulu: Thule Institute, University of Oulu, 2004.
MATTHEWS, E. *The weight of nations: Material outflows from industrial economies.* Washington, DC: World Resources Institute, 2000.
MCDONALD, G. W. AND PATTERSON, M. G. *Development of a New Zealand Physical Input-Output Table.* Palmerston North: New Zealand Centre for Ecological Economics, 2006.
MINISTRY OF INDUSTRY AND TRADE. *Fuel and Energy Consumption – 2014.* Prague: Ministry of Industry and Trade, 2015.
MULALIC, I. *Material flows and physical Input-output tables – PIOT for Denmark 2002 based on MFA.* Copenhagen: Statistics Denmark, 2007.

NEBBIA, G. Contabilià monetaria e contabilità ambientale. *Economia Pubblica*, 2000, 30(6), pp. 5–33.
OECD. *Measuring material flows and resource productivity.* Paris: OECD, 2008.
SCHANDL, H. AND WEST, J. Material Flows and Material Productivity in China, Australia, and Japan. *Journal of Industrial Ecology*, 2012, 16(3), pp. 352–364.
STAHMER, C., KUHN, N., BRAUN, M. *Physical input–output tables for Germany, 1990.* Wiesbaden: DESTATIS, 1997.
UNITED NATIONS et al. *System of Environmental-Economic Accounting 2012: Central Framework.* New York: United Nations, 2014.
WEIZSÄCKER, E. U., HARGROVES, K., SMITH, M. H., DESHA, C., STASINOPOULOS, P. *Factor five. Transforming the global economy through 80% improvements in resource productivity: A report to the Club of Rome.* London; Sterling, VA: Earthscan/The Natural Edge Project, 2009.

SUPPLEMENTARY INFORMATION 1 – ANNEX 1

Table A1 Industry and product groups used for PSUT compilation

NACE/CPA	Name
01	Products of agriculture, hunting and related services
02	Products of forestry, logging and related services
03	Fish and other fishing products; aquaculture products; support services to fishing
05	Coal and lignite
06	Crude petroleum and natural gas
07	Metal ores
08	Other mining and quarrying products
09	Mining support services
10	Food products
11	Beverages
12	Tobacco products
13	Textiles
14	Wearing apparel
15	Leather and related products
16	Wood and of products of wood and cork, except furniture; articles of straw and plaiting materials
17	Paper and paper products
18	Printing and recording services
19	Coke and refined petroleum products

Table A1 (continuation)

NACE/CPA	Name
20	Chemicals and chemical products
21	Basic pharmaceutical products and pharmaceutical preparations
22	Rubber and plastics products
23	Other non-metallic mineral products
24	Basic metals
25	Fabricated metal products, except machinery and equipment
26	Computer, electronic and optical products
27	Electrical equipment
28	Machinery and equipment n.e.c.
29	Motor vehicles, trailers and semi-trailers
30	Other transport equipment
31	Furniture
32	Other manufactured goods
33	Repair and installation services of machinery and equipment
35	Electricity, gas, steam and air conditioning
36	Natural water; water treatment and supply services
37, 38, 39	Sewerage services; sewage sludge; waste collection, treatment and disposal services; materials recovery services; remediation services and other waste management services
41	Buildings and building construction works
42	Constructions and construction works for civil engineering
43	Specialised construction works
45	Wholesale and retail trade and repair services of motor vehicles and motorcycles
46, 47	Wholesale trade services, except of motor vehicles and motorcycles; retail trade services, except of motor vehicles and motorcycles
49	Land transport services and transport services via pipelines
50	Water transport services
51	Air transport services
52	Warehousing and support services for transportation
53	Postal and courier services

Table A1 (continuation)

NACE/CPA	Name
55	Accommodation services
56	Food and beverage serving services
58	Publishing services
59	Motion picture, video and television programme production services, sound recording and music publishing
60	Programming and broadcasting services
61	Telecommunications services
62	Computer programming, consultancy and related services
63	Information services
64	Financial services, except insurance and pension funding
65	Insurance, reinsurance and pension funding services, except compulsory social security
66	Services auxiliary to financial services and insurance services
68	Real estate services
69	Legal and accounting services
70	Services of head offices; management consulting services
71	Architectural and engineering services; technical testing and analysis services
72	Scientific research and development services
73	Advertising and market research services
74	Other professional, scientific and technical services
75	Veterinary services
77	Rental and leasing services
78	Employment services
79	Travel agency, tour operator and other reservation services and related services
80	Security and investigation services
81	Services to buildings and landscape
82	Office administrative, office support and other business support services
84	Public administration and defence services; compulsory social security services
85	Education services

Table A1 (continuation)

NACE/CPA	Name
86	Human health services
87	Residential care services
88	Social work services without accommodation
90	Creative, arts and entertainment services
91	Library, archive, museum and other cultural services
92	Gambling and betting services
93	Sporting services and amusement and recreation services
94	Services furnished by membership organisations
95	Repair services of computers and personal and household goods
96,97,98, 99	Other personal services; undifferentiated goods and services produced by private households for own use; services provided by extraterritorial organisations and bodies

Source: CZSO <*https://www.czso.cz/csu/czso/classifications*>

SUPPLEMENTARY INFORMATION 2 – ANNEX 2

Annex 2 is available at the webpage of *Statistika: Statistics and Economy Journal*, see the online version of No. 3/2018 (Excel file) at: <*http://www.czso.cz/statistika_journal*>.

Model of Socio-Ecological and Economic System: the Central Federal District Regions of the Russian Federation

Roman Zhukov [1] | *Financial University under the Government of the Russian Federation, Tula, Russian Federation*

Abstract

This paper aims to present the model of socio-ecological and economic system for the regions of the Central Federal district and to make calculations related to the assessment of their state, functioning, management efficiency and harmony. We apply the author's methodological toolkit that includes the formation of individual and integrated indicators of the functioning and the management efficiency of complex systems, considered as socio-ecological and economic systems. The coefficient of harmony is a measure of the equilibrium of the region's functioning, which is constructed using the author's methodology. The paper results are as follows: The model is presented with 9 generalized performance indicators, 26 individual performance indicators and 49 factor indicators (state and impact factors) using open data from the Federal State Statistics Service of the Russian Federation for 2004–2015. Also the assessment of state, functioning, management efficiency and harmony of the Central Federal district are described. Included is also the analysis of the results.

Keywords

Regional economy, socio-ecological and economic systems, econometric modeling, integral indicators balance, efficiency

JEL code

C10, P25, R11, R15

INTRODUCTION

At the moment Russian economy is based on resource-exploration and resource-intensive sectors. It leads to a deterioration of the environmental health and the depletion of natural resources. At the same time, shifting the focus towards the economy without the social component support decreases the living standards, which should not be present in developing and developed countries, including the Russian Federation. Regions represent their territorial administrative units and they should be considered as complex socio-ecological and economic systems (SEES). Their management determines the economic growth and well-being of the entire state's population. However, the managerial efficiency is determined

[1] Department of Mathematics and Informatics, Financial University under the Government of the Russian Federation (Tula Branch), Oruzheynaya 1-a, Tula, Russian Federation. E-mail: pluszh@mail.ru, phone: (+7)4872223-483.

first of all by the validated assessment of the SEES's state and functioning as well as by studying the nature of the impact on the part of the management entity. It is the complexity of systems (socio-ecological-economic systems) that defines the diversity of approaches to their study. At the same time, most authors emphasize the social, environmental or economic aspects, applying various evaluation methods and using various models of the state and functioning of complex systems. The purpose of the article is to build a model for the functioning of the socio-ecological and economic system for the regions of the Central Federal District, and offer an assessment of the efficiency of their management. We are going to use the results of author's previous studies related to the research of individual components (social, environmental and economic).

The main method is supposed to be the author's approach presented in the part Data and Methods. It is based on the methods of system analysis, econometric modeling, correlation and regression analysis and the convolution algorithm of data.

1 SURVEY AND LITERATURE REVIEW

The socio-ecological and economic system is the unity of the social, ecological and economic subsystems (Tatarkin et al., 2016). This division is rather conventional as the terminological analysis revealed such systems in definitions that also correspond to other categories. For example, the social system contains economic factors, environmental - social elements, etc. Therefore, the term SEES allows to avoid these inaccuracies and form a general concept of a complex system comprising economic, social, environmental objects, processes, environments and projects.

Anopchenko and Murzin (2012) describe the socio-ecological and economy system as a set of structural components of nature, society and economy with different interconnection types. Meanwhile, they are the interconnections but not the components that are more significant. So it is partly true.

Davankov and Ezhova define the territorial socio-ecological and economy system (TSEES) as "an interconnected combination of natural, industrial, demographic, social and institutional components that function intentionally at the certain territory" (Tatarkin et al., 2016).

Rozanova (2001) determines TSEES as "part of the territory where the intensity of the links between the elements of nature and the economy greatly exceeds the intensity of the connections directed from outside and outwards the system, or assembly of elements itself".

Therefore, we conclude that socio-ecological and economy system is the integral set of interrelated objects, processes, systems and environments having social, ecological and economic relations, as well as their combinations. The system's functioning is aimed at ensuring its survival in the space-time continuum through production, distribution, exchange and consumption of material and non-material resources, substance, energy and information.

The different approaches to the study of socio-ecological and economic systems and the formation of SEES' conception started to develop in the 70's of the last century, when the "nature-population-economy" scheme was used as the basis for the study (Tatarkin et al., 2016). At that date, a number of studies describing the methodology of EES and SEES's research and modeling were published, among them you can find works by Jacobson and Jacobson (1987), Gurman (1981) and also co-authored by Ryumina (2001). Muhina et al. (1978) offered an analytical scheme of "impact – changes – consequences" for the study of complex systems. It correlated with socio-economic geography and suggested the use of the process approach. Bashirova (2010) comes up with the idea of targeted approach and studies SEES from the aspect of meeting the needs of system's elements within the framework of three components: environmental protection, protection and improvement of the human environment, and economic development. Achieving the goals ensures the balance of the state and functioning of SEES. Herewith, the object scheme "nature–man–economy" is used, similar to the Muhina and Preobrazhenskiy's triad presented in the late 1970's.

The study of the sustainable development of complex economic systems is based on the works of the founders of economic thought such as Leontiev (1997), Rothschild et Stiglitz (1976), Akerlof et al. (2001), Solow (1956), Schumpeter (1935), who determined the possibility of equilibrium growth. The issues of environmental safety are highlighted in the works by Zerkalov (2012) and other authors. At the same time, the studies concern only the social, ecological or economic component, although the presence of other different factors is supposed a priori. Summarizing the approaches presented above, we come to the conclusion that all of them use the basic methodology of system analysis in full or in part. The system analysis is based on 3 fundamental methodologies, including analysis, synthesis and behaviourism (Gharajedaghi, 2011).

The authors rely on a variety of models, applying them in compliance with the goals and objectives of the study.

The economic aspect of research at the regional level includes the following most known models of the regional economy.

Individual regional model. The authors use the economic growth models that exploit external demand (export base model) and also Keynes multiplier (multiplier multiplied by the initial change in investment gives the increment of GNP) as the basic factor. Nonlinear approaches (Zhulanov, 2016) are introduced within the framework of well-known Leontiev's input output model (Leontiev, 1997). A detailed analysis of models used by modern scientists, in particular input-output tables, is presented in (Baranov et al., 2016).

The Neumann model of economic growth (linear model of production) by J. von Neumann is also often applied. It describes the possibility of an object's outgate to a time-independent trajectory or a trajectory of equilibrium growth, and, unlike Leontiev's model, that used industries as production units, he relies on technological processes (Neumann, 1945–1946). The application of optimized interindustry models of the region offered by Kantorovich (1939) within the framework of his theory of optimum allocation of resources is widespread. They also take the advantage of Cobb-Douglas model to describe the results of economic system's production depending on labour and capital (Cobb and Douglas, 1928). A number of authors use regional econometric models, including for assessing the sustainability of socio-economic systems (Latypova and Chertykovtsev, 2008).

There are works that use an aggregated model of the regional economy functioning (6 blocks), with separate allocation models (cargo transportation, population migration, production location), as well as interregional models of the national economy (Larionov et al., 2017). Kondratiev's cyclic model (Kondratiev, 1993) is also known. The model presented in the report to the Club of Rome in 1972 by Meadows et al. (1972) is widely known too. It describes the limits of economic and demographic growth under the depletion of natural resources. It is a system of 16 nonlinear differential equations with more than 30 variables. It can be referred to the class of socio-ecological-economic models (Meadows et al., 2012).

Borodin (2006) took as a basis the notorious D. W. Pearce and R. K. Turner model describing the well-being of the population according to natural resources and services, which can be called the model of interaction between economic and ecological systems (Pearce and Turner, 1990). He related pollution (production waste) with the production and economic activities of the SEES, which results in a well-being of the population. These models are used to assess the equilibrium growth of complex systems as a factor in their steady state and functioning (Tatarkin et al., 2016). Partial or general indicators are used as efficiency indicators for evaluating complex systems, the latter of which are defined in the framework of a component analysis or an expert-statistical approach.

As for the first method the first principal component of private unified indicators is the integral indicator (Aivazian, 2003), with its informative value exceeding 55%. If it is impossible to figure out the first principal component, the weighted methodology is used. The squares of the lengths of the eigenvectors of the correlation matrix of the partial indices act as weight coefficients.

One of the options for dimension reduction, which is similar to Principal Components, is the Multidimensional Scaling (an alternative to factor analysis), taking into account the distances (proximity) between objects, but the latest data representations are difficult to interpret (Tolstova, 2006).

The expert-statistical method is supposed to execute an examination of weight coefficients value formation, which ultimately leads to compiling a rating of each of the individual indices significance.

An analysis of recent works on integral estimates has shown that most of them contain a calculation of averages of different types (arithmetic, geometric simple or weighted) and a weighted assessment that defines the importance of this or that partial indicator (subindex) is carried out mostly by expert methods.

Krivonozhko and Lychev (2010) present the Functioning Environment Analysis (FEA) method that develops Data Envelopment Analysis (DEA) by the research group of Charnes et al. (1978). The weighted coefficients of partial indicators that make part of summarized index are defined by solving the tasks of non-linear programming.

From the latter it can be concluded that a number of methods for constructing integral indicators can be used not only to assess the results rating of complex system's functioning, but also its efficiency, although the boundaries between these concepts, when using appropriate methods, are not generally defined.

The variety of existing approaches to the modeling and assessment of the complex systems' functioning led the authors to the idea of developing a methodology that would include the advantages of the presented studies and highlight the main characteristics of socio-ecological and economic systems.

2 DATA AND METHODS

As the model for the functioning of complex systems, a system of linear equations is chosen which are the standardized models of the additive form as the models for forming the norms $\hat{y}_i^*$:

$$\hat{y}_i^* = \sum_{j=1}^{n} C_{i,j} \cdot x_j^* + \sum_{s=1}^{s} D_{i,s} \cdot z_s^* , \tag{1}$$

where n is the number of state factors, s is the number of impact factors, $C_{i,j}$, $D_{i,s}$, are corresponding weight coefficients between i productive (result of functioning of system) and j and s standardized factors of x_j^* state and z_s^* impact. State factors present a set of essential properties that the system possesses at a given moment in time. Impact factors are a set of controlled properties that lead to changes in results of functioning of the system. Subjects of management can change the impact factors. When substituting actual values x_j^* and z_s^* in (1) for k region you can get a individual norm. Herewith:

$$x_j^* = \frac{x_j - M(x_j)}{\sigma(x_j)} , \tag{2}$$

$$z_s^* = \frac{z_s - M(z_s)}{\sigma(z_s)} , \tag{3}$$

where x_j , z_s are the actual values of factors of state and impact in absolute units of measurement.

At the same time, the assessment of the state and functioning of the SEES is determined by means of individual and generalized performance indicators, which describe the results of functioning of the system (Zhuravlev et al., 2013):

$$\xi_{k,i}(t) = \frac{y_{k,i}^0(t)}{\hat{y}_{k,i}^0(t)} , \tag{4}$$

where $y_{k,i}^0(t)$, $\hat{y}_{k,i}^0(t)$ are actual and normative values of standardized individual performance indicators which are specific for explored region, k is the region number, t is the time parameter ($t = 1..T$), $i = 1..m$,

m is the amount of individual performance indicators, the index "0" indicates that the normalization procedure has been carried out (bringing to the scale from 0 to 1), and after standardization:

$$y^0_{k,i}(t) = \frac{y^*_{k,i} - \min\{y^*_{k,i}, \hat{y}^*_{k,i}\}}{\max\{y^*_{k,i}, \hat{y}^*_{k,i}\} - \min\{y^*_{k,i}, \hat{y}^*_{k,i}\}}, \tag{5}$$

$$\hat{y}^0_{k,i}(t) = \frac{\hat{y}^*_{k,i} - \min\{y^*_{k,i}, \hat{y}^*_{k,i}\}}{\max\{y^*_{k,i}, \hat{y}^*_{k,i}\} - \min\{y^*_{k,i}, \hat{y}^*_{k,i}\}}. \tag{6}$$

Here $y^*_{k,i}$, $\hat{y}^*_{k,i}$ are standardized individual performance indicators, defined by the formulas:

$$y^*_{k,i}(t) = \frac{y_{k,i} - M(y_i(t))}{\sigma(y_i(t))}, \tag{7}$$

$$\hat{y}^*_{k,i}(t) = \frac{\hat{y}_{k,i} - M(\hat{y}_i(t))}{\sigma(\hat{y}_i(t))}, \tag{8}$$

where $M(y_i(t))$, $M(\hat{y}_i(t))$, $\sigma(y_i(t))$, $\sigma(\hat{y}_i(t))$ are expected value and standard deviation, respectively.

A generalized performance indicator is calculated as the ratio of individual performance indicators (actual and normative) (Zhukov, 2014):

$$\xi_k(t) = \frac{\sqrt{\sum_{p=1}^{m}\sum_{q=1}^{m} r_{pq} \cdot y^0_{k,p}(t) \cdot y^0_{k,q}(t)}}{\sqrt{\sum_{p=1}^{m}\sum_{q=1}^{m} \hat{r}_{pq} \cdot \hat{y}^0_{k,p}(t) \cdot \hat{y}^0_{k,q}(t)}}, \tag{9}$$

where r_{pq} and $\hat{r}_{pq}$ are the corresponding paired correlation coefficients.

If $\xi_k(t) \geq 1$, then we can assume them satisfactory otherwise we are to take measures aimed at the achievement of the norm that is calculated for each k of the object.

The application the proposed approach makes it possible to meet all the requirements for integrated assessment indicators.

The harmonic coefficient characterizing the balance of the system's functioning results can be determined by formula (Zhukov, 2016, 2017):

$$K_k = 1 - \frac{\sigma(\xi_{k,i})}{M(\xi_{k,i})}, \tag{10}$$

where $M(\xi_{k,i})$ is expected value, $\sigma(\xi_{k,i})$ is standard deviation. The closer to the one K_k, the more harmonic is the functioning of the object under research. This indicator does not characterize its specialization, but shows the degree of compliance of the indicators under consideration with the norms, taking into account specific conditions.

We introduce the notion of the effectiveness indicator. The effectiveness indicator is the ratio of the change in performance indicators to the change in the factors of the state (impact, generalized factor of state and impact) for the period under review.

A partial effectiveness indicator may be determined as:

$$Ef_{FPP,k}(t) = \frac{\Delta\xi_{k,i}(t)}{\Delta x_{k,i,j}(t)} = \frac{\xi_{k,i}(t) / \xi_{k,i}(t_o)}{x^0_{k,i,j}(t) / x^0_{k,i,j}(t_0)}, \tag{11}$$

where $\xi_{k,i}(t)$, $\xi_{k,i}(t_o)$ are determined by Formula (4) for current and base (previous) periods $x^0_{k,i,j}(t)$, $x^0_{k,i,j}(t_0)$ are normalized state factors correspondently; k is the region number; i is the index of partial performance indicators; j is index of state factor, FPP is Functioning, Partial indicator, Partial factor. Here and further "/" is the division operation, which is presented in this way for greater clarity of representation of the formula. The normalization procedure for $x^0_{k,i,j}(t)$, $x^0_{k,i,j}(t_0)$ is carried out by formula which is similar to Formula (5).

A partial effectiveness indicator of the functioning SEES by the generalized state factor is calculated as:

$$Ef_{FPG,k}(t) = \frac{\Delta\xi_{k,i}(t)}{\Delta x_{k,i,j}(t)} = \frac{\xi_{k,i}(t)/\xi_{k,i}(t_o)}{x^0_{k,i}(t/t_0)}, \tag{12}$$

where FPG is Functioning, Partial indicator, Generalized factor and $x^0_{k,i}(t/t_0)$ is determined by formula:

$$x^0_{k,i}(t/t_0) = \sqrt{\sum_{p=1}^{n}\sum_{q=1}^{n} r_{pq}\cdot\frac{x^0_{k,i,p}(t)}{x^0_{k,i,p}(t_0)}\cdot\frac{x^0_{k,i,q}(t)}{x^0_{k,i,q}(t_0)}}, \tag{13}$$

here r_{pq} is Pearson's correlation coefficient between p and q state factors, n is the amount of factors.

In case when state factors are independent then correlation matrix with r_{pq} are identity matrix and Formula (13) is simplified:

$$x^0_{k,i}(t) = \sqrt{\sum_{j=1}^{n}\left(\frac{x^0_{k,i,j}(t)}{x^0_{k,i,j}(t_0)}\right)^2}. \tag{14}$$

A generalized indicator of the effectiveness by partial and generalized state factors is determined as:

$$Ef_{FGP,k}(t) = \frac{\Delta\xi_{k}(t)}{\Delta x_{k,j}(t)} = \frac{\xi_{k}(t)/\xi_{k}(t_o)}{x^0_{k,j}(t)/x^0_{k,j}(t_0)}, \tag{15}$$

$$Ef_{FGG}(t) = \frac{\Delta\xi_{k}(t)}{\Delta x_{k}(t)} = \frac{\xi_{k}(t)/\xi_{k}(t_o)}{x^0_{k}(t/t_0)}, \tag{16}$$

where FGP is Functioning, Generalized indicator, Partial factor, FGG is Functioning, Generalized indicator, Generalized factor and $\xi_k(t)$, $\xi_k(t_o)$ are determined by Formula (9).

Similarly, an partial and generalized indicator of the effectiveness by partial and generalized impact factors is determined as:

$$Ef_{IPP,k}(t) = \frac{\Delta\xi_{k,i}(t)}{\Delta z_{k,i,s}(t)} = \frac{\xi_{k,i}(t)/\xi_{k,i}(t_o)}{z^0_{k,i,s}(t)/z^0_{k,i,s}(t_0)}, \tag{17}$$

$$Ef_{IPG,k}(t) = \frac{\Delta\xi_{k,i}(t)}{\Delta z_{k,i,j}(t)} = \frac{\xi_{k,i}(t)/\xi_{k,i}(t_o)}{z^0_{k,i}(t/t_0)}, \tag{18}$$

$$Ef_{IGP}(t) = \frac{\Delta\xi_{k}(t)}{\Delta z_{k,s}(t)} = \frac{\xi_{k}(t)/\xi_{k}(t_o)}{z^0_{k,s}(t)/z^0_{k,s}(t_0)}, \tag{19}$$

$$Ef_{IGG,k}(t) = \frac{\Delta\xi_{k}(t)}{\Delta z_{k}(t)} = \frac{\xi_{k}(t)/\xi_{k}(t_o)}{z^0_{k}(t/t_0)}. \tag{20}$$

Here s is the amount of impact factors, IP(G)P(G) is Impact, Particular (Generalized) indicator, Partial (Generalized) factor, $z^0_{k,i}(t/t_0)$ and $z^0_k(t/t_0)$ are defined similarly to Formula (13).

Partial and generalized indicators of management effectiveness are a generalization of performance and impact indicators, since they take into account the influence of both state and impact factors. The last of them change the state of the SEES, which is one of the management functions.

The main difference between the constructed indicators of SEES management effectiveness will be the presence in the denominator of the generalized state and impact factor:

$$Ef_{MPG}(t) = \frac{\Delta\xi_{k,i}(t)}{\{\Delta x_{k,i,j}(t), \Delta z_{k,i,j}(t)\}} = \frac{\xi_{k,i}(t)/\xi_{k,i}(t_o)}{\{x^0_{k,i}(t/t_0), z^0_{k,i}(t/t_0)\}}, \tag{21}$$

$$Ef_{MGG}(t) = \frac{\Delta\xi_k(t)}{\{\Delta x_{k,j}(t), \Delta z_{k,j}(t)\}} = \frac{\xi_k(t)/\xi_k(t_o)}{\{x^0_k(t/t_0), z^0_k(t/t_0)\}}, \tag{22}$$

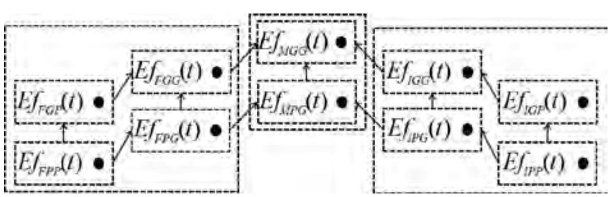
where M is Management and $\{x^0_k(t/t_0), z^0_k(t/t_0)\}$ is determined by the formula:

$$\{x^0_k(t/t_0), z^0_k(t/t_0)\} = \sqrt{\begin{aligned} &\sum_{p=1}^{n}\sum_{q=1}^{n} r_{pq}\cdot\frac{x^0_{k,i,p}(t)}{x^0_{k,i,p}(t_0)}\cdot\frac{x^0_{k,i,q}(t)}{x^0_{k,i,q}(t_0)} + \\ &+\sum_{p=1}^{n}\sum_{q=1}^{s} r_{pq}\cdot\frac{x^0_{k,i,p}(t)}{x^0_{k,i,p}(t_0)}\cdot\frac{z^0_{k,i,q}(t)}{z^0_{k,i,q}(t_0)} + \\ &+\sum_{p=1}^{s}\sum_{q=1}^{s} r_{pq}\cdot\frac{z^0_{k,i,p}(t)}{z^0_{k,i,p}(t_0)}\cdot\frac{z^0_{k,i,q}(t)}{z^0_{k,i,q}(t_0)}\end{aligned}}. \tag{23}$$

It should be noted that not all influencing factors determine the efficiency of management, but only those of them that are directly related to the purposeful influence on the system's functioning on the part of the subjects of management, for example, investments of various kinds, planned expenditures for the performance of an activity, etc.

The efficiency indicators formed this way represent an assessment of the functioning, impact and management of the SEES with various level of detail, and it allows to give a comprehensive assessment to the object of research (Figure 1).

Figure 1 Scheme of conformity of indicators of the effectiveness (functioning, impact and management) of SEES



Source: Own construction

Figure 1 shows that a set of efficiency assessment indicators represents step structure and provides a possibility of studying of SEES in various cuts.

The data of the research is represented by the Federal State Statistics Service of the Russian Federation (ROSSTAT) data for the regions of the Central Federal district in 2004–2015[2] (2004–2014 – social component, 2007–2014 – ecological component, 2007–2015 – economic component, number of observations for each variable made 187, 136, 153 for 17 CFD regions correspondently). The choice of different periods is related to the availability of open statistical data of ROSSTAT. The statistical data set has no outliers.

The description of the variables is represented in Table 1 to Table 8.

Table 1 The description of the variables of social assessment (generalized and private assessment indicators)

№	Variables	Description
1	$\hat{y}^*_{soc}$	Generalized social indicator
2	$\hat{y}^*_{2..9}$	Remaining life expectancy index
3	$\hat{y}^*_{5}$	Education index
4	$\hat{y}^*_{10.2/2}$	Per capita income

Note: Variable without any extra characters are variables in absolute units, * standardized variables, ^ model (calculated) variables, fact values for No. 2,3,4 were calculated using technique of UNDP.[3]
Source: ROSSTAT, own construction

Table 2 The description of the variables of social assessment (state factors)

№	Variables	Description
1	$x^*_{2.8}$	Natural increase per 1 000 people
2	$x^*_{2.12}$	Net migration per 1 000 people
3	$x^*_{4,1}$	Percentage rates for real disposable income in comparison with the corresponding period of the year 2004
4	$x^*_{4.10}$	Social transfers as per cent of total income levels
5	$x^*_{4.19,5/1}$	Education expenditures of the general population generating inflation-adjusted positive and with the corresponding period of the year 2004
6	$x^*_{4.19,6/1}$	Health expenditures of the general population generating inflation-adjusted positive and with the corresponding period of the year 2004
7	$x^*_{4.22}$	Per capita consumption of meat and meat products
8	$x^*_{8.1}$	Number of crimes reported per 100 000 people
9	$x^*_{9.1}$	Air pollutants from stationary sources
10	$x^*_{9.5}$	Discharges of polluted waste water into surface water bodies

Note: * standardized variables.
Source: ROSSTAT, own construction

[2] Federal State Statistics Service of the Russian Federation (ROSSTAT) [online]. [cit. 20.6.2017]. <*http://www.gks.ru*>.
[3] The United Nations development programme (UNDP) [online]. [cit. 18.9.2017]. <*http://www.undp.ru*>.

Table 3 The description of the variables of social assessment (impact factors)

№	Variables	Description
1	$z^*_{1/1}$	Investments in fixed capital per capita (total) adjusted for inflation
2	$z^*_{1,5/1}$	Investments in fixed capital by education adjusted for inflation
3	$z^*_{1/2}$	Per capita investments in fixed capital (total) to purchasing power parity (PPP) in US dollars
4	z^*_3	Consolidated budget expenditures (total)

Note: * standardized variables.
Source: ROSSTAT, own construction

Table 4 The description of the variables of ecological assessment (generalized and private assessment indicators)

№	Variables	Description
1	$\hat{y}^*_{ecol}$	Generalized ecological indicator
2	$\hat{y}^*_{9,1}$	Generalized performance indicator for the assessment of the air
3	$\hat{y}^{*-1}_{9.1,1}$	Air pollutants
4	$\hat{y}^*_{9.2,1}$	Capture of air pollutants from stationary sources
5	$\hat{y}^*_{9,2}$	Generalized performance indicator for the assessment of the water source
6	$\hat{y}^*_{9.3,2}$	The use of fresh water
7	$\hat{y}^*_{9.4,2}$	Volume of circulating and consistently used water
8	$\hat{y}^{*-1}_{9.5,2}$	Discharges of polluted waste water into surface water bodies
9	$\hat{y}^*_{9,3}$	Generalized performance indicator for the assessment of the generation, disposal and use of waste
10	$\hat{y}^{*-1}_{9.6,3}$	Waste generation of production and consumption
11	$\hat{y}^{*-1}_{9.7,3}$	Waste storage and disposal
12	$\hat{y}^*_{9.8,3}$	Waste use and decontamination
13	$\hat{y}^{*-1}_{9.9,3}$	Waste intensity

Note: * standardized variables, ^ model (calculated) variables, –1 shows that the indicator is negative.
Source: ROSSTAT, own construction

In Tables 6 to 8 the variables that we use are grouped according to the division offered by Kolesnikov et Tolstoguzov (2016) and performance indicators (private assessment indicators) that include GDP by regions by economic activities (Russian Classification of Economic Activities (NACE (OKVED)) was used in the Russian Federation till 2015)[4] grouped by aggregate sector of the economy.

[4] Russian Classification of Economic Activities (NACE (OKVED)) [online]. [cit. 18.6.2017]. *<http://www.gks.ru/bgd/free/b02_60/Main.htm>*.

Table 5 The description of the variables of ecological assessment (state and impact factors)

№	Variables	Description
		State factors
1	$x^*_{17.30}$	Passenger turnover (cars and public buses)
2	$x^*_{13.1}$	Sold goods of their own production and provided works and services by types of economic activity (total)
3	$x^*_{2.2}$	Average annual population
4	$x^*_{13.64}$	Electricity production
5	$x^*_{13.1,1.2}$	Sold goods of their own production and provided works and services by types of economic activity (mining and quarrying)
6	$x^*_{13.2}$	Industrial production index as % to the base year
		Impact factors
7	$z^*_{9.1,3}$	Environmental expenditure in 2007 prices (air)
8	$z^*_{9.3,3}$	Environmental expenditure in 2007 prices (waste)

Note: * standardized variables.
Source: ROSSTAT, own construction

Table 6 The description of the variables of economic assessment (private assessment indicators)

№	Variables	Description	№	Variables	Description
1	$\hat{y}^*_1$	Commodity aggregate sector	4.3	$\hat{y}^*_{4.3}$	Section H. Transporting and storage. Section J. Information and communication (I)
1.1	$\hat{y}^*_{1.1}$	Section A. Agriculture, forestry and fishing (A)	4.4	$\hat{y}^*_{4.4}$	Section K. Financial and insurance activities (J)
1.2	$\hat{y}^*_{1.2}$	Section B. Mining and quarrying (C)	4.5	$\hat{y}^*_{4.5}$	Section L. Real estate activities. Section M. Professional, scientific and technical activities. Section N. Administrative and support service activities (K)
2	$\hat{y}^*_2$	Manufacturing aggregate sector	5	$\hat{y}^*_5$	Aggregate sector of non-market services
2	$\hat{y}^*_2$	Section C. Manufacturing (D)	5.1	$\hat{y}^*_{5.1}$	Section D. Electricity, gas, steam and air conditioning supply. Section E. Water supply; sewerage; waste management and remediation activities (E)
3	$\hat{y}^*_3$	Construction aggregate sector	5.2	$\hat{y}^*_{5.2}$	Section O. Public administration and defence; compulsory social security (L)
3	$\hat{y}^*_3$	Section F. Construction (F)	5.3	$\hat{y}^*_{5.3}$	Section P. Education (M)
4	$\hat{y}^*_4$	Aggregate sector of market services	5.4	$\hat{y}^*_{5.4}$	Section Q. Human health and social work activities (N)
4.1	$\hat{y}^*_{4.1}$	Section G. Wholesale and retail trade; repair of motor vehicles and motorcycles (G)	5.5	$\hat{y}^*_{5.5}$	Section R. Arts, entertainment and recreation. Section S. Other services activities (O)
4.2	$\hat{y}^*_{4.2}$	Section I. Accommodation and food service activities (H)			

Note: * standardized variables, ^ model (calculated) variables, () NACE Rev. 1.1. sections.
Source: ROSSTAT, Zhukov (2018)

Table 7 The description of the variables of economic assessment (state factors)

№	Variables	Description	№	Variables	Description
6	The cost of fixed production assets at full accounting value at the end of the year by types of economic activity		7.4	$\overset{*}{x}_{3,2}$	Section F
6.1	$\overset{*}{x}_{1.1,1}$	Section A (A)	7.5	$\overset{*}{x}_{4.1,2}$	Section G (G)
6.2	$\overset{*}{x}_{1.2,1}$	Section B (C)	7.6	$\overset{*}{x}_{4.2,2}$	Section I (H)
6.3	$\overset{*}{x}_{2,1}$	Section C (D)	7.7	$\overset{*}{x}_{4.3,2}$	Sections H,J (I)
6.4	$\overset{*}{x}_{3,1}$	Section F (F)	7.8	$\overset{*}{x}_{4.5,2}$	Sections L, M, N (K)
6.5	$\overset{*}{x}_{4.1,1}$	Section G (G)	7.9	$\overset{*}{x}_{5.1,2}$	Sections D, E (E)
6.6	$\overset{*}{x}_{4.3,1}$	Sections H, J (I)	7.10	$\overset{*}{x}_{5.3,2}$	Section P (M)
6.7	$\overset{*}{x}_{5.1,1}$	Sections D,E (E)	8	$\overset{*}{x}_{2.2}$	Average annual population
7	Average annual number of persons employed by types of economic activities		9	Transport	
7.1	$\overset{*}{x}_{1.1,2}$	Section A (A)	9.1	$\overset{*}{x}_{4.1}$	Passenger turnover of public buses
7.2	$\overset{*}{x}_{1.2,2}$	Section B (C)	9.2	$\overset{*}{x}_{4.2}$	Departure of passengers by public railway transport
7.3	$\overset{*}{x}_{2,2}$	Section C (D)	10	$\overset{*}{x}_{5}$	Morbidity per 1 000 of population, registered diseases diagnosed in patients for the first time in life

Note: Variables without any extra characters are variables in absolute units, * standardized variables, ^ model (calculated) variables, () NACE Rev. 1.1. sections.
Source: ROSSTAT, Zhukov (2018)

Table 8 The description of the variables of economic assessment (impact factors)

№	Variables	Description	№	Variables	Description
11	Investments in fixed capital by kinds of economic activities		12	$\overset{*}{z}_{2}$	Organic fertilizers per 1 ha of agricultural crops (in terms of 100% nutrients)
11.1	$\overset{*}{z}_{2,1}$	Section D (D)	13	Consolidated budget expenditures (by object)	
11.2	$\overset{*}{z}_{4.1,1}$	Section G (G)	13.1	$\overset{*}{z}_{5.2,3}$	Social policy
11.3	$\overset{*}{z}_{4.2,1}$	Section I (H)	13.2	$\overset{*}{z}_{5.3,3}$	Education
11.4	$\overset{*}{z}_{4.5,1}$	Sections L, M, N (K)			

Note: Variable without any extra characters are variables in absolute units, * standardized variables, ^ model (calculated) variables, () NACE Rev. 1.1. sections.
Source: ROSSTAT, Zhukov (2018)

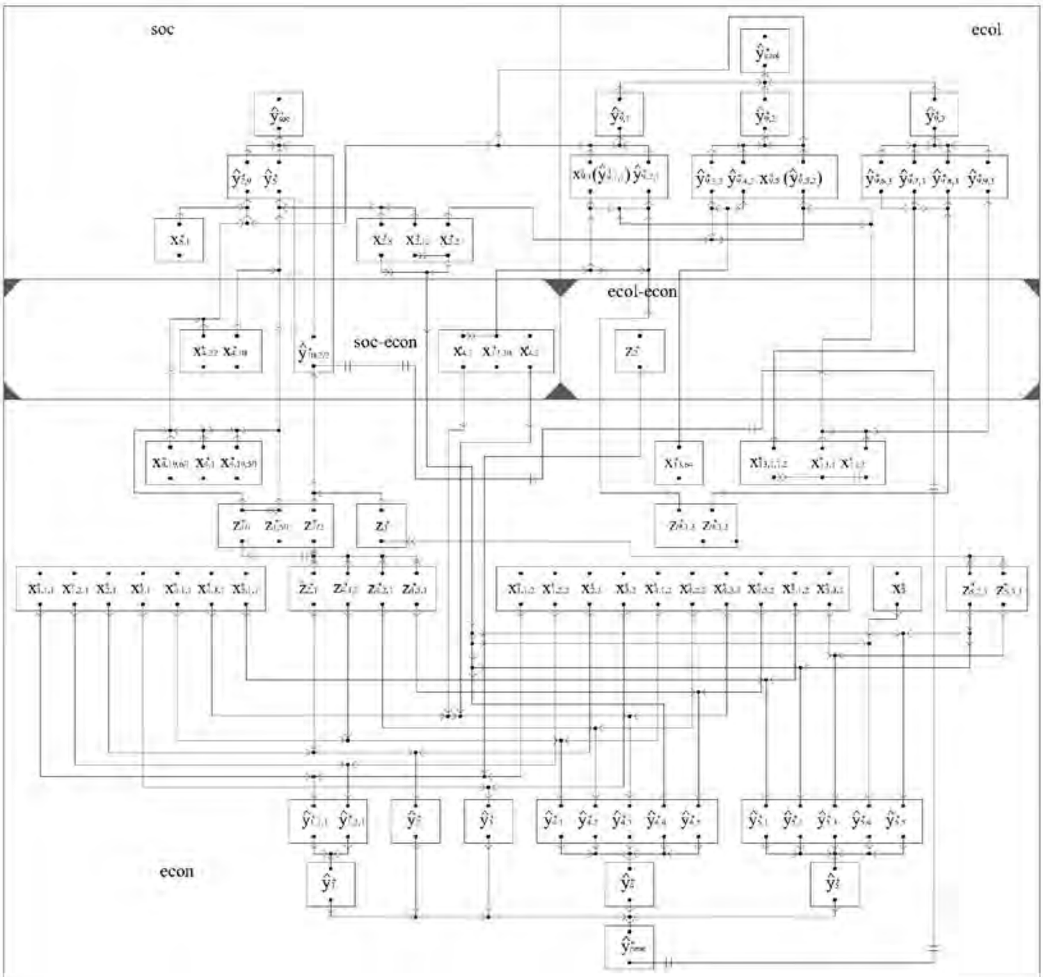
These tables include only substantial factors, significant at the level of no more than 5%. We used least square method (backward selection) to select included variables.

3 RESULTS AND DISCUSSION

The conducted research resulted in the corresponding models in forms (1), whose specification is represented in Tables 1 to 8 using author's expert system (beta version) (Zhukov, 2015). The corresponding formulas, which were reflected in the author's earlier investigations (Zhukov, 2016, 2017, 2018), are presented in the Annex.

The functioning model of SEES is included 26 equations with 49 state and impact factors. The visualization of the detected significant links is shown in Figure 2.

Figure 2 Model of socio-ecological and economic system for the CFD of the Russian Federation



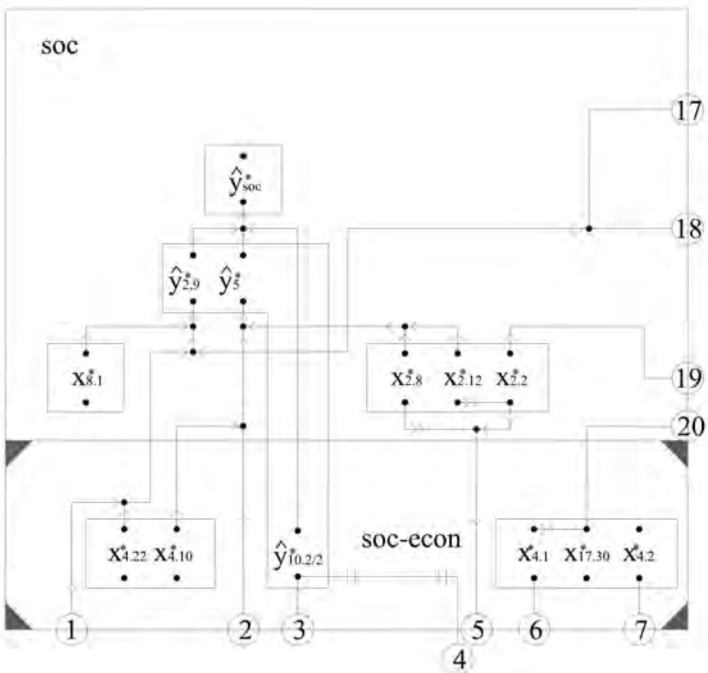
Note: Soc is social area; ecol is ecological area; econ is economic area; soc-econ is socioeconomic area; ecol-econ is ecological-economic area; > is direction of dependence of factors; >> shows that this indicator is part of the indicator with which it is associated (for example, the consolidated budget expenditures by education are part of the total consolidated budget expenditures); | | this indicator is derived from the associated indicator (for example, investments in fixed capital adjusted for inflation and investments in fixed capital (total) to purchasing power parity (PPP) in US dollars).

Source: Own construction

Figure 2 shows that some factors are socioeconomic and ecological-economic factors. At the same time, the environmental partial performance indicator $\hat{y}^*_{9,1}$ is state factor for remaining life expectancy index $\hat{y}^*_{2..9}$.

The detail visualization of the detected significant links is shown in Figures 3 to 5.

Figure 3 Detailing the social and socio-economic area of SEES model



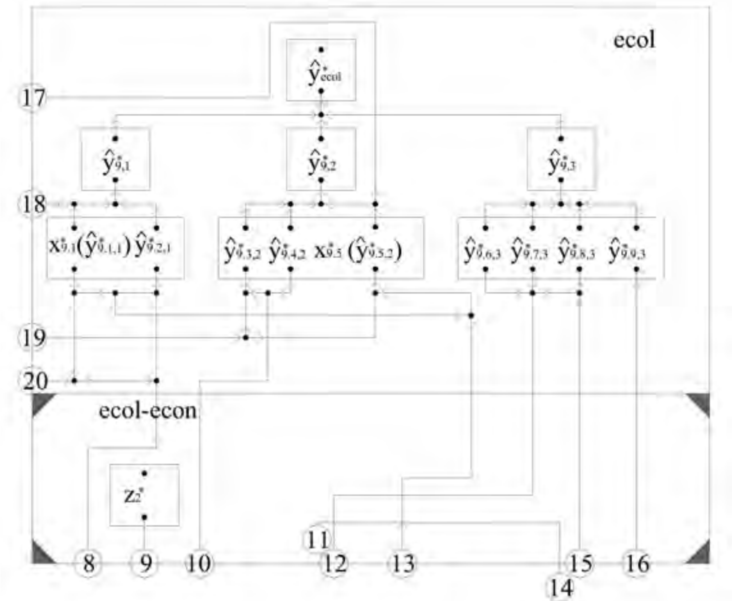
Source: Own construction

One of the performance indicators of ecological characteristic namely discharges polluted waste water into surface water bodies ($\hat{y}^*_{9.5,2}$) defines the state of social component; it also influences remaining life expectancy index ($\hat{y}^*_{2..9}$). The factors ($x^*_{2.8}$ and $x^*_{2.12}$) are components of the average annual population $x^*_{2.2}$ (Figure 3). The passenger turnover of public buses ($x^*_{4.1}$) are component passenger turnover (cars and public buses) ($x^*_{17.30}$). Also per capita income ($\hat{y}^*_{10.2/2}$) (Figure 3) belong to socioeconomic area and it is included in the generalized social indicator $\hat{y}^*_{econ}$ (Figure 5).

Figure 4 shows that organic fertilizers per 1 ha of agricultural crops (in terms of 100% nutrients) is included in ecological-economic area of SEES model.

The represented model that includes only the essential factors selected with the help of least square method (backward selection) is connected to the external environment through state and impact factors. Net migration per 1 000 people ($x^*_{2.12}$) and passenger turnover (cars and public buses) ($x^*_{17.30}$) are referred to such factors of state. Investments in fixed capital by kinds of economic activities ($z^*_{2,1}$) and other total (for example, $z^*_{1/1}$) can be referred to the factors of impact. All the rest factors of state and impact define the set of limitations. For example, regional budget revenues in the form of federal budget transfers impose restrictions to the structure and volume of consolidated budget expenses. All this requires a more detailed evaluation for a particular SEES. So the factors

Figure 4 Detailing the ecological and ecological-economic area of SEES model



Source: Own construction

of state and impact not included in the model can serve as the constraints imposed on the socio-ecological and economic system's functioning that allows to describe it within the framework of open systems.

To confirm the independence of the social, environmental and economic assessment, we will perform a factor analysis for the generalized SEES performance indicators. To do it we used the principal component analysis (PCA). The calculation was carried out by means of the analytical platform Deductor by BaseGroup Labs[5] (Table 9). The evaluation period was 2010–2014.

Table 9 shows that that the contribution of each component is significant so they can't be elicit from the model. It determines the possibility of constructing a relationship between them in the form of a linear or other communication model.

The first component gives the largest contribution to the result (46.25%).

We built three models showing the links between the social, ecological and economic performance indicators.

$$\xi_{\mathrm{soc}} = 0.940 + 0.071 \cdot \xi_{\mathrm{econ}}, \tag{24}$$

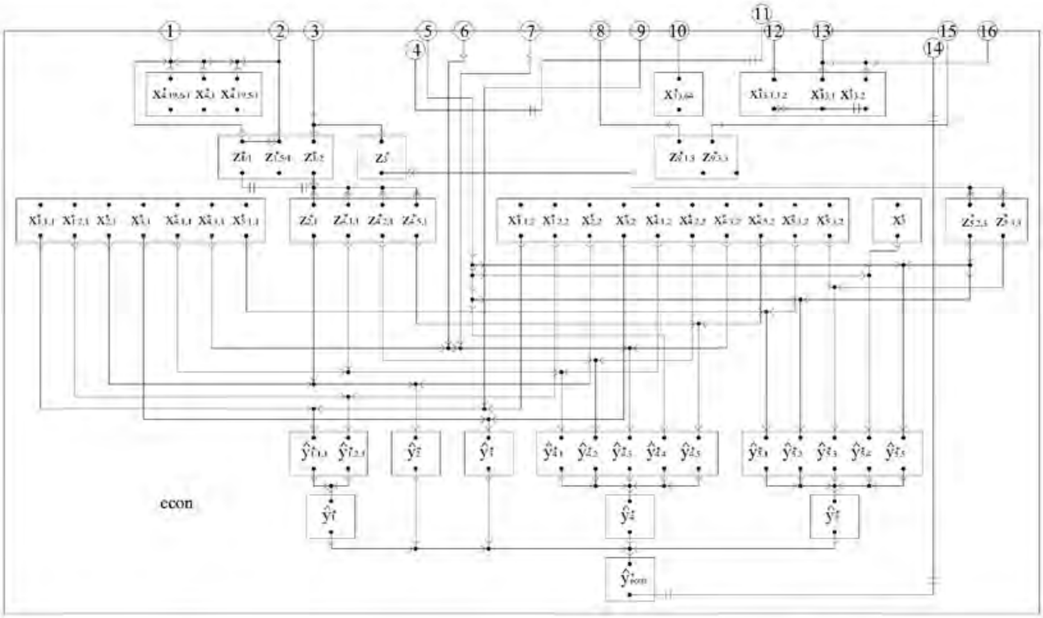
$$\xi_{\mathrm{ecol}} = 0.754 + 0.103 \cdot \xi_{econ}, \tag{25}$$

$$\xi_{\mathrm{econ}} = -1.115 + 1.677 \cdot \xi_{soc} + 0.590 \cdot \xi_{\mathrm{ecol}}, \tag{26}$$

where ξ_{econ}, ξ_{soc}, ξ_{ecol} are generalized social, ecological and economic performance indicators correspondently.

5 Analytical platform Deductor Studio [online]. [cit. 23.1.2018]. <*https:// basegroup.ru*>.

Figure 5 Detailing the ecological and ecological-economic area of SEES model



Source: Own construction

Table 9 The results of principal component analysis (PCA)

N⁰	Principal component	Eigenvalues	% of the variance	Cumulative weight, %
1	Value 1	1.388	46.25	46.25
2	Value 2	1.044	34.82	81.07
3	Value 1	0.568	18.93	100

Source: Own construction

The standardized equation has the following form:

$$\breve{\xi}_{\mathrm{soc}} = 0.339 \cdot \breve{\xi}_{econ}, \tag{27}$$

$$\breve{\xi}_{\mathrm{ecol}} = 0.238 \cdot \breve{\xi}_{econ}, \tag{28}$$

$$\breve{\xi}_{\mathrm{econ}} = 0.351 \cdot \breve{\xi}_{soc} + 0.255 \cdot \breve{\xi}_{\mathrm{ecol}}. \tag{29}$$

For these models (Formulas 24–26) the coefficients of determination are $R^2_1 = 0.115$, $R^2_2 = 0.057$, $R^2_3 = 0.179$; the coefficients of multiple correlation are $R_1 = 0.339$, $R_2 = 0.238$, $R_2 = 0.424$; the standard errors are $\sigma_1 = 0.085$, $\sigma_2 = 0.181$, $\sigma_3 = 0.394$; the calculated values of F-test are $F_{calc1} = 10.741$, $F_{calc2} = 4.984$, $F_{calc3} = 8.961$, the critical values are $F_{cr1} = 3.956$, $F_{cr2} = 3.956$, $F_{cr3} = 3.108$, the statistical significance at 5% level with $(1, 83)_1$, $(1, 83)_2$, $(1, 82)_3$ degrees of freedom correspondently.

For the parameters of the model one (Formula 24) t-statistic is calculated (t_{calc}, in parentheses is the standard error) and, correspondently, values for the coefficients are $a_0 = 20.093$ (0.047), $a_1 = 3.164$ (0.022), the critical values are $t_{cr} = 1.989$ the statistical significance at 5% level with 82 degrees of freedom. The average relative errors of this model is $E_{rel} = 6.320\%$. For the model two (Formula 25) these evaluation parameters are $a_0 = 13.837$ (0.055), $a_1 = 2.232$ (0.046), $E_{rel} = 16.525\%$. For the model three (Formula 26) there are $a_0 = 2.073$ (0.538), $a_1 = 3.502$ (0.479), $a_2 = 2.544$ (0.232), $E_{rel} = 20.526\%$.

The model linking the economic indicator with the rest indicators turned out the most qualitative but its accuracy is lower in comparison with other models.

In this case the system of Formulas (24), (25) or (27), (28) or just one Formula (26) or (29) can be used to describe the functioning of SEES.

To construct the generalized performance indicator (ξ) we took 26 partial performance indicators and 51 factor attributes (the factors of state and impact). The results of calculations are shown in Table 10.

Table 10 The values of the integral performance indicators $\xi(t)$ for the regions of the CFD in 2010–2014 years

Nº	Region/Year	2010	2011	2012	2013	2014
1	Belgorod	0.972	0.995	0.960	0.984	1.064
2	Bryansk	1.082	1.126	1.220	1.154	1.224
3	Vladimir	0.892	0.964	0.976	0.966	0.984
4	Voronezh	0.750	0.877	0.907	0.842	0.886
5	Ivanovo	0.852	0.923	0.974	0.931	0.965
6	Kaluga	0.924	0.938	0.912	0.907	0.927
7	Kostroma	0.937	1.023	1.110	1.119	1.158
8	Kursk	1.070	1.028	1.025	1.037	1.041
9	Lipetsk	0.786	0.810	1.032	1.025	1.083
10	Moscow	0.923	0.918	0.942	0.899	0.903
11	Orel	0.968	1.007	1.035	0.933	0.986
12	Ryazan	1.004	1.017	1.054	1.055	1.256
13	Smolensk	1.009	0.956	0.982	0.997	1.014
14	Tambov	1.150	1.215	1.022	0.996	1.016
15	Tver	0.838	0.857	0.844	0.846	0.876
16	Tula	0.809	0.909	0.896	0.905	0.949
17	Yaroslavl	0.861	0.980	1.019	1.073	1.109

Source: Own construction

Table 10 shows that the maximum value of the indicator is observed in 2014 compared to previous years for the almost all CFD regions except Voronezh, Kursk, Moscow, Orel and Tambov regions.

The performance uniformity analysis of SEES functioning showed the differentiation of the harmonic coefficient (see Formula 10) for the CFD regions in 2010–2014 years (see Table 11).

Table 11 shows that in comparison with the harmonic coefficient values of the individual components (economic and social) that were presented in previous studies (Zhukov, 2016, 2017, 2018), the coefficient value is lower, for the ecological component the values above are observed for most regions except for the Bryansk, Kostroma, Tambov and Tula regions. That is, for allocated SEES the imbalance is not compensated by inclusion of the ecological component when calculating the harmonic coefficient.

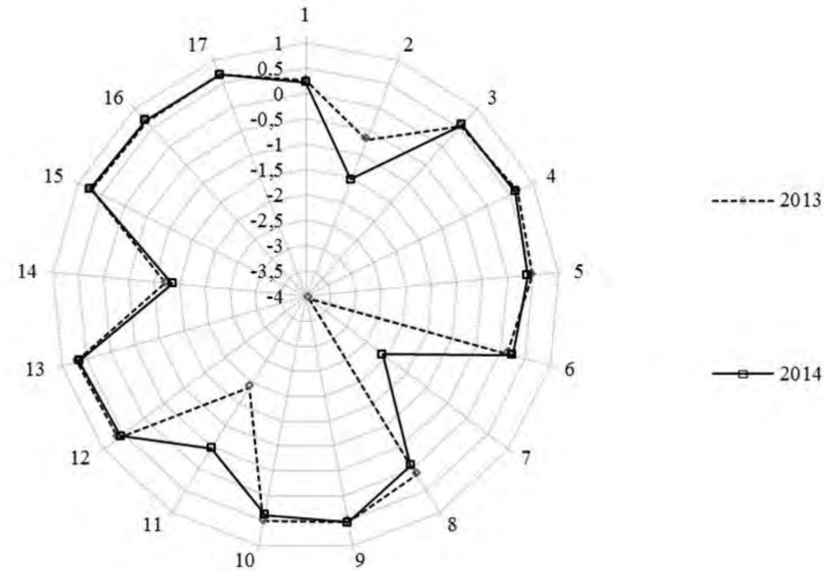
The visualization of the harmonic coefficient in 2013–2014 years is shown in Figure 6.

Table 11 The values of the harmonic coefficient for the regions of the CFD in 2010–2014 years

Nº	Region/Year	2010	2011	2012	2013	2014
1	Belgorod	0.179	0.193	0.212	0.252	0.217
2	Bryansk	–0.099	–0.687	–0.632	–0.695	–1.534
3	Vladimir	0.699	0.151	0.439	0.532	0.570
4	Voronezh	0.685	0.653	0.677	0.653	0.622
5	Ivanovo	0.420	0.303	0.545	0.486	0.393
6	Kaluga	–0.040	0.035	0.043	0.138	0.235
7	Kostroma	–0.269	–0.782	–0.192	–3.919	–2.104
8	Kursk	–0.238	0.552	0.567	0.136	–0.072
9	Lipetsk	0.508	0.413	0.580	0.528	0.534
10	Moscow	0.573	–0.374	0.426	0.511	0.381
11	Orel	–2.424	–0.802	–1.192	–1.893	–0.476
12	Ryazan	0.328	0.579	0.642	0.636	0.577
13	Smolensk	0.648	0.682	0.702	0.688	0.645
14	Tambov	–0.908	–0.833	–2.445	–1.222	–1.358
15	Tver	0.604	0.709	0.694	0.716	0.746
16	Tula	0.509	0.601	0.688	0.661	0.694
17	Yaroslavl	0.668	0.629	0.638	0.683	0.679

Source: Own construction

Figure 6 Harmonic coefficient for the CFD regions in 2013–2014



Note: 1 – Belgorod, 2 – Bryansk, 3 – Vladimir, 4 – Voronezh, 5 – Ivanovo, 6 – Kaluga, 7 – Kostroma, 8 – Kursk, 9 – Lipetsk, 10 – Moscow, 11 – Orel, 12 – Ryazan, 13 – Smolensk, 14 – Tambov, 15 – Tver, 16 – Tula, 17 – Yaroslavl.
Source: Own construction

We can conclude that the decomposition of the social and economic harmonic coefficients into individual components worsens the value of the harmonic coefficient, which includes all three components, and the ecological coefficient, on the contrary, increases its value.

Analyzing the efficiency of SEES management we chose the following main factors influencing the generalized indicator of SEES functioning assessment: investments in fixed capital per capita (total) adjusted for inflation ($\Delta z^0_{1/1}$); investments in fixed capital by education adjusted for inflation ($\Delta z^0_{1,5/1}$); consolidated budget expenditures (total; Δz^0_3); environmental expenditure in 2007 prices (air; $\Delta z^0_{9.1,3}$); environmental expenditure (waste; $\Delta z^0_{9.3,3}$); investments in fixed capital by kinds of economic activities (D(D), G(G), H(I), K(L), M, N; $\Delta z^0_{2,1}$, $\Delta z^0_{4.1,1}$, $\Delta z^0_{4.2,1}$, $\Delta z^0_{5.1,1}$ sections correspondently).

Indicators of management effectiveness ($\Delta\xi$ / Δz^0_i) is calculated by Formula (19), where t, t_0 are current and previous periods correspondently (see Table 12).

Thus, in the Tula region the values of the effective indicators are more (less) than one for 5 (4) factors correspondently.

The least value (0.653) of the indicator is observed for the consolidated budget expenditures factor.

High efficiency for the generalized indicator by the factor of investment in fixed-capital assets per capita (total), adjusted for the inflation rate of the corresponding year ($\Delta z^0_{1/1}$) was observed for the Orel (29.228) and Smolensk (16.338) regions, and its value is out of the general trend for the rest of the regions. It can be caused by the following factors:. As for the Orel region, its investments in fixed capital changed insignificantly. In 2014 they were equal to 21 552.20 rubles (0.006 in the standardized variables), in 2013 – 23 148.90 rubles (0.179), i. e. they decreased by 1 596.70 rubles, and their ratio in standardized variables was 0.036.

Table 12 The values of the effective indicator for the regions of the CFD in 2014

Region/Indicator	$\Delta\xi / \Delta z^0_{1/1}$	$\Delta\xi / \Delta z^0_{1,5/1}$	$\Delta\xi / \Delta z^0_{3}$	$\Delta\xi / \Delta z^0_{9.1,3}$	$\Delta\xi / \Delta z^0_{9.3,3}$	$\Delta\xi / \Delta z^0_{2,1}$	$\Delta\xi / \Delta z^0_{4.1,1}$	$\Delta\xi / \Delta z^0_{4.2,1}$	$\Delta\xi / \Delta z^0_{5.1,1}$
Belgorod	1.933	1.178	1.4	0.701	1.102	2.145	1.298	2.202	1.374
Bryansk	1.126	0.696	1.378	1.065	0.62	1.253	1.061	1.024	1.074
Vladimir	0.674	1.384	0.722	1.149	0.555	0.219	0.683	0.789	1.171
Voronezh	1.034	1.148	0.928	20.46	1.152	0.678	0.815	0.478	0.849
Ivanovo	1.245	2.502	0.933	1.051	1.079	1.028	0.892	1.02	0.903
Kaluga	1.035	1.491	0.589	1.343	1.198	0.926	1.78	1.037	1.024
Kostroma	0.819	1.145	0.965	0.963	0.985	0.596	1.019	1.051	1.033
Kursk	2.084	7.792	0.983	1.173	0.675	1.075	0.889	1.366	0.905
Lipetsk	1.112	1.004	0.225	1.512	0.85	1.833	0.513	1.853	0.588
Moscow	1.312	0.971	0.885	1.412	1.286	0.866	0.657	0.892	0.755
Orel	29.228	1.065	1.042	1.034	1.187	0.957	0.947	1.049	1.017
Ryazan	2.95	0.538	2.028	3.03	0.991	4.895	1.349	1.194	1.255
Smolensk	16.338	2.309	1.39	1.075	1.052	0.831	0.964	0.925	0.956
Tambov	0.936	1.356	0.691	1.085	0.909	0.917	1.067	1.007	0.3
Tver	5.863	1.375	1.024	0.826	1.012	1.683	0.444	0.039	0.969
Tula	1.681	4.652	0.653	1.312	1.393	0.7	1.707	0.881	0.803
Yaroslavl	3.869	5.639	0.863	0.903	0.807	0.104	1.211	0.048	1.831

Source: Own construction

Meanwhile, the generalized performance indicator increased and amounted to 0.986 in 2014 against 0.933 in 2013, i.e. 1.057 times. Their ratio gives the value of 29.228. It means that at the decrease of investments in fixed capital the growth of the generalized performance indicator of functioning SEES is observed. A similar situation is typical for the Smolensk region.

High value of the effective indicator (20.460) by the factor of environmental expenditure in 2007 prices (air; $\Delta z^0_{9.1,3}$) is observed in the Voronezh region. The maximum value by the factor of investments in fixed capital by kinds of economic activities (Section D (C) Manufacturing) was identified in the Ryazan region (4.895), and minimum value was found in the Yaroslavl region (0.104). The minimum value of all the efficiencies is observed in the Tver region by the factor of investments in fixed capital by kinds of economic activities (Section H (I) Hotels and restaurants), i.e. they do not lead to the significant increase of the generalized efficiency indicator, that speaks of the poor performance of their usage.

Complex (generalized) effective indicator (by a generalized factor) can serve as an assessment for the cumulative analysis of the use of operating conditions and the impact factors on the part of the managerial subjects.

The values of the generalized effective indicator of generalized factor using Formulas (20) and (14) is presented in Table 13.

In 2014, the greatest efficiency was observed in the Ryazan region (0.557), and the lowest in the Moscow region (0.076). It shows that the change of generalized indicator of efficiency by 0.201 and 0.004 for these regions, respectively, demanded minor expenses from the management entity aimed at the development of SEES. So, for example, investments in fixed-capital assets (at current prices) in Ryazan region amounted

Table 13 The values of the generalized effective indicator of generalized factor for the regions of the CFD in 2011–2014 years

N⁰	Region/Year	2010	2011	2012	2013	2014
1	Belgorod	-	0.218	0.291	0.263	0.345
2	Bryansk	-	0.292	0.338	0.358	0.341
3	Vladimir	-	0.314	0.310	0.304	0.322
4	Voronezh	-	0.121	0.098	0.284	0.376
5	Ivanovo	-	0.365	0.327	0.345	0.360
6	Kaluga	-	0.338	0.307	0.269	0.336
7	Kostroma	-	0.329	0.315	0.357	0.328
8	Kursk	-	0.243	0.379	0.364	0.281
9	Lipetsk	-	0.104	0.303	0.321	0.328
10	Moscow	-	0.301	0.256	0.294	0.076
11	Orel	-	0.131	0.345	0.276	0.325
12	Ryazan	-	0.133	0.238	0.088	0.557
13	Smolensk	-	0.298	0.226	0.286	0.354
14	Tambov	-	0.361	0.251	0.332	0.339
15	Tver	-	0.013	0.049	0.273	0.319
16	Tula	-	0.231	0.225	0.391	0.332
17	Yaroslavl	-	0.417	0.314	0.336	0.327

Source: Own construction

to 54 056 million rubles, and in Moscow region – to 640 320 million rubles, which is 11.8 times less, per capita – 47 720 rubles and 88 018 rubles respectively. So, it can be confirmed that the Moscow region was to expect even greater changes in Generalized effectiveness because the region's potential is much higher than the Ryazan region has. And, it indicates the significance of generalized efficiency method. I.e., despite the fact that Moscow is more attractive from the point of view of investments, the Ryazan region is ahead in relative terms. Drawing a parallel with economic effects, one can speak of lower return on investment. Similar arguments can be made for other regions.

CONCLUSION

The main result of the study is the model specification of functioning of socio-ecological and economic system for the regions of the Central Federal district using author's approach. The results of the generalized assessment of the CFD regions functioning, their degree of balance, as well as the efficiency of their management are presented. The functioning model of SEES included 26 equations with 49 state and impact factors from social, ecological and economic areas.

The results of the research may be of interest for specialists in the field of economics, ecology, sociology, state and municipal management, students and graduate students of the relevant fields, as well as for general readers studying the problems of sustainable development of the SEES, including the regional and international levels.

ACKNOWLEDGMENTS

The reported research was funded by the Russian Foundation for Basic Research and the government of the Tula region of the Russian Federation, grant № 18-410-710001 (r_a).

References

AIVAZIAN, S. A. Empiricheskiy analiz sinteticheskikh kategoriy kachestva zhizni naseleniya. In Russian [Empirical analysis of synthetic categories of the quality of life of the population]. *Ekonomika i matematicheskiye metody [Economics and mathematical methods]*, 2003, Vol. 39, No. 3, pp. 19–53.

AKERLOF, G., SPENCE, M., STIGLITZ, J. L'asymetrie au coeur de la nouvelle microeconomie. *Problemès econ. P.*, 2001, 2734, pp. 19–24.

ANOPCHENKO, T. YU. AND MURZIN, A. D. Struktura sotsial'no-ekonomicheskikh i ekologicheskikh sostavlyayushchikh sistemy kompleksnogo razvitiya territoriy. In Russian [Structure of socio-economic and ecological components of the system of integrated development of territories] [online]. *Nauka i obrazovaniye: khozyaystvo i ekonomika; predprinimatel'stvo; pravo i upravleniye [Science and education: economy and economics; entrepreneurship; law and management]*, 2012, No. 1. [cit. 18.12.2017]. <*http://www.journal-nio.com/index.php?option=com_content&view=article&id=1136&Itemid=102*>.

BARANOV, A. O., DONDOKOV, Z. B. D., SLEPENKOVA, YU. M. Postroyeniye i ispol'zovaniye regional'nykh mezhotraslevykh modeley dlya analiza i prognozirovaniya razvitiya ekonomiki regionov. In Russian [Construction and use of regional interbranch models for analysis and forecasting the development of regional economies]. *Idei i idealy* [Ideas and Ideals], 2016, Vol. 2, No. 4(30), pp. 66–85.

BASHIROVA, A. A. Formirovaniye strategii regional'nogo razvitiya s pozitsiy ekologo-ekonomicheskoy sbalansirovannosti. In Russian [Formation of the strategy of regional development from the position of ecological and economic balance]. *Ekonomika i upravleniye [Economy and management]*, 2010, No. 11(72), pp. 87–90.

BORODIN, A. I. Modelirovaniye ekologo-sotsial'no-ekonomicheskikh system. In Russian [Borodin AI Modeling of ecological-social-economic systems]. *Izvestiya Tomskogo politekhnicheskogo universiteta [Bulletin of the Tomsk Polytechnic University]*, 2006, No. 2, pp. 221–224.

CHARNES, A., COOPER, W. W., RHODES, E. Measuring the efficiency of decision making units. *European Journal of Operational Research*, 1978, Vol. 6, No. 2, pp. 429–444.

COBB, C. W. AND DOUGLAS, P. H. A Theory of Production. *American Economic Review*, 1928, Vol. 18, pp. 139–165.

GHARAJEDAGHI, J. *Systems Thinking: Managing Chaos and Complexity: A Platform for Designing Business Architecture.* Morgan Kaufmann of Elsevier 30 Corporate Drive, Suite 400, Burlington, MA 01803, USA, 2011.

GURMAN, V. I. *Modeli upravleniya prirodnymi resursami.* In Russian *[Models of natural resources management].* Moscow: Nauka, 1981.

GURMAN, V. I. AND RYUMINA, E. V. *Modelirovaniye sotsio-ekologo-ekonomicheskoy sistemy regiona.* In Russian *[Modeling of the socio-ecological and economic system of the region].* Moscow: Nauka, 2001.

JACOBSON, A. YA. AND JACOBSON, B. B. *Ekologo-ekonomicheskiye sistemy kak ob"yekt informatsii upravleniya i rayonirovaniya.* In Russian *[Ecological and economic systems as an object of information management and regionalization].* Novosibirsk: IEIOPP, 1987.

KANTOROVICH, L. V. *Matematicheskiye metody organizatsii i planirovaniya proizvodstva.* In Russian *[Mathematical methods of organization and production planning].* Leningrad: Izdatelstvo LGU [Leningrad State University Publishing House], 1939.

KOLESNIKOV, N. G. AND TOLSTOGUZOV, O. V. Strukturnye izmeneniya ekonomiki Severo-Zapada Rossii: prostranstvennyj aspekt. In Russian [Structural changes in the economy of the North–West of Russia: spatial aspect]. *Baltijskij region [The Baltic region]*, 2016, Vol. 8, No. 2, pp. 30–47.

KONDRATIEV, N. D. *Izbrannyye sochineniya.* In Russian *[Selected works].* Moscow: Ekonomika, 1993.

KRIVONOZHKO, V. E. AND LYCHEV, A. V. *Analiz deyatelnosti slozhnyh socialno-ekonomicheskih sistem.* In Russian *[Analysis of activity of complex socio-economic systems].* Moscow: Maks Press, 2010.

LARIONOV, I. K., SILVESTROV, S. N., ANTIPOV, K. V. et al. *Ekonomicheskaya teoriya. Ekonomicheskiye sistemy: formirovaniye i razvitiye: uchebnik.* In Russian *[Economic theory. Economic systems: formation and development: a textbook].* Moscow: Izdatel'sko-torgovaya korporatsiya «Dashkov i K°», 2017.

LATYPOVA, N. M. AND CHERTYKOVTSEV, V. K. *Ekonometricheskiye modeli ustoychivosti sotsial'no-ekonomicheskikh sistem. Statisticheskiye aspekty issledovaniya.* In Russian *[Econometric models of sustainability of socio-economic systems. Statistical Aspects of the Research].* Samara: Samarskiy gosudarstvennyy arkhitekturno-stroitel'nyy universitet [Samara State University of Architecture and Civil Engineering], 2008.

LEONTIEV, V. V. *Mezhotraslevaya ekonomika.* In Russian *[Intersectoral economy].* Moscow: Ekonomika [Economy], 1997.

MEADOWS, D. H, RANDERS, J., MEADOWS, D. L., BEHRENS, W. W. *The Limits to Growth: A Report for the Club of Rome's Project on the Predicament of Mankind.* NY: Universe Books, 1972.

MEADOWS, D. H. RANDERS, J., MEADOWS, D. L. *Predely rosta: 30 let spustya.* In Russian *[Limits of growth: 30 years later].* Moscow: BINOM, Laboratoriya znaniy, 2012.

MUHINA, L. I., PREOBRAZHENSKIY, V. S., RUNOVA, T. G., DOLGUSHIN, I. Y. Osobennosti sistemnogo podhoda k probleme ocenki vozdejstviya cheloveka na sredu. In Russian [Specific Features of the system approach to the problem of assessment of human impact on environment]. *Geograficheskie aspekty vzaimodejstviya v sisteme «chelovek-priroda»,* 1978, pp. 22–49.

NEUMANN, VON J. A model of general economic equilibrium. *Rev. Econ. Studies*, 1945–1946, Vol. 13, No. 33, pp. 1–9.

PEARCE, D. W. AND TURNER, R. K. *Economics of Natural Resources and the Environment.* New York: Harvester Wheatsheaf, 1990.

ROTHSCHILD, M. AND STIGLITZ, J. E. Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information. *Quarterly Journal of Economics*, 1976, Vol. 90, No. 4, pp. 629–649.

ROZANOVA, L. N. Territorial'nyye sotsio-ekologo-ekonomicheskiye sistemy: obshchiye podkhody k otsenke [online]. In Russian [Territorial socio-ecological and economic systems: general approaches to evaluation]. *Vestnik TISBI*, 2001, No. 3. [cit. 23.12.2017]. <*http://old.tisbi.ru/science/vestnik/2001/issue3/econom[5].html*>.

SCHUMPETER, J. A. A Theorist's Comment on the Current Business Cycle. *Journal of the American Statistical Association*, 1935, No. 30(189), pp. 167–168.

SOLOW, R. M. A Contribution to the Theory of Economic Growth. *The Quarterly Journal of Economics*, 1956, Vol. 70, No. 1, pp. 65–94.

TATARKIN, A. I., DAVANKOV, A. YU., PRYAKHIN, G. N., SEDOV, V. V., SHUMAKOV, A. YU. *Upravleniye sbalansirovannym razvitiyem territorial'nykh sistem: voprosy teorii i praktiki: monografiya.* In Russian *[Management of Balanced Development of Territorial Systems: Issues of Theory and Practice: Monograph].* Chelyabinsk: Chelyabinsk State University, 2016.

TOLSTOVA, YU. N. *Osnovy mnogomernogo shkalirovaniya.* In Russian *[Fundamentals of multidimensional scaling].* Moscow: KDU, 2006.

ZERKALOV, D. V. *Ekologicheskaya bezopasnost'. Monografiya.* In Russian *[Environmental Safety. Monograph].* Kiyev: Osnova, 2012.

ZHUKOV, R. A. Economic Assessment of the Development Level of the Central Federal District Regions of the Russian Federation: Econometric Approach [online]. *Statistika: Statistics and Economy Journal*, 2018, Vol. 98, No. 1, pp. 53–68.

ZHUKOV, R. A. Nekotoryye aspekty otsenki kachestva zhizni i upravleniya v sotsio-ekologo-ekonomicheskikh sistemakh: regiony Tsentral'nogo federal'nogo okruga. In Russian [Some aspects of the assessment of quality of life and management in socio-ecological-economic systems: The Central Federal District regions case study]. *Regional'naya ekonomika: teoriya i praktika [Regional Economics: Theory and Pactice]*, 2017, Vol. 15, No. 7(442), pp. 1261–1275.

ZHUKOV, R. A. Ocenka sostoyaniya slozhnyh sistem na primere regionov Centralnogo federalnogo okruga: ekologo-ekonomicheskiy aspekt. In Russian [Assessment of comlex systems of regions of the Central Federal district: ecological and economic aspect]. *Regionalnye issledovaniya [Regional studies]*, 2016, No. 4(54), pp. 81–89.

ZHUKOV, R. A. Primenenie fundamental'nogo podhoda k upravleniyu social'no-ekonomicheskim razvitiem territoriy. In Russian [Application of the fundamental approach to the management of socio-economic development of regions]. *Nauchnoe obozrenie [Scientific review]*, 2014, No. 4, pp. 272–274.

ZHUKOV, R. A. Vnedrenie programmnyh ekonomiko-matematicheskih kompleksov v praktiku deyatelnosti organov gosudarstvennogo upravleniya. In Russian [Implementation of the program of economic and mathematical systems in practice of activity of bodies of state management]. *Fundamentalnye issledovaniya [Fundamental research]*, 2015, Vol. 9, No. 3, pp. 555–559.

ZHULANOV, E. E. Ekonomiko-matematicheskoye modelirovaniye razvitiya vosproizvodstvennogo kontura regional'noy sotsial'no-ekonomicheskoy sistemy. In Russian [Economic-mathematical modeling of the development of the reproduction contour of the regional socio-economic system]. *Ekonomicheskiy analiz: teoriya i praktika [Economic analysis: theory and practice]*, 2016, No. 6(453), pp. 172–187.

ZHURAVLEV, S. D., ZHUKOV, R. A., MAMON, Y. I. Fundamental Research for Management Efficiency. *Applied and Fundamental Studies Proceedings of the 2nd International Academic Conference. Publishing House «Science and Innovation Center», and the International Journal of Advanced Studies*, 2013, pp. 228–334.

ANNEX

The equations for description of model of functioning of socio-ecological and economic systems by CFD regions:

$$\begin{array}{lllll} \hat{y}^*_{2..9} = & 0.336 \cdot x^*_{4,1} + & 0.146 \cdot x^*_{4.19,6/1} + & 0.300 \cdot x^*_{4.22} - & 0.251 \cdot x^*_{8.1} - \\ & (0.004) & (0.003) & (0.010) & (0.000) \\ & -0.099 \cdot x^*_{9.1} - & 0.230 \cdot x^*_{9.5} + & 0.115 \cdot z^*_{1/1}, & (R^2 = 0.740), \\ & (0.001) & (0.001) & (0.000) & \end{array} \tag{30}$$

$$\begin{array}{lllll} \hat{y}^*_{5} = & 0.732 \cdot x^*_{2.8} - & 0.536 \cdot x^*_{2.12} - & 0.201 \cdot x^*_{4,1} + & 0.122 \cdot x^*_{4.10} - \\ & (0.001) & (0.000) & (0.000) & (0.000) \\ & -0.159 \cdot x^*_{4.19,5/5} - & 0.288 \cdot z^*_{1,5/1}, & (R^2 = 0.623), & \\ & (0.000) & (0.001) & & \end{array} \tag{31}$$

$$\begin{array}{llll} \hat{y}^*_{10.2/2} = & 0.735 \cdot z^*_{1/2} + & 0.292 \cdot z^*_{3}, & (R^2 = 0.804) \\ & (0.107) & (0.002) & \end{array} , \tag{32}$$

$$\begin{array}{llll} \hat{y}^*_{9.1,1} = & 0.635 \cdot x^*_{17.30} + & 0.278 \cdot x^*_{13.1}, & (R^2 = 0.918) \\ & (0.000) & (0.000) & \end{array} \tag{33}$$

$$\begin{array}{lllll} \hat{y}^*_{9.2,1} = & -0.484 \cdot x^*_{17.30} + & 0.793 \cdot x^*_{13.1} + & 0.558 \cdot z^*_{9.1,3}, & (R^2 = 0.585) \\ & (0.000) & (0.004) & (0.626) & \end{array} \tag{34}$$

$$\begin{array}{llll} \hat{y}^*_{9.3,2} = & 0.417 \cdot x^*_{2.2} + & 0.466 \cdot x^*_{13.64}, & (R^2 = 0.537) \\ & (0.029) & (0.130) & \end{array} \tag{35}$$

$$\begin{array}{lll} \hat{y}^*_{9.4,2} = & 0.867 \cdot x^*_{13.64}, & (R^2 = 0.751) \\ & (0.340) & \end{array} \tag{36}$$

$$\hat{y}^*_{9.5,2} = \underset{(0.012)}{0.474} \cdot x^*_{2.2} + \underset{(0.000)}{0.483} \cdot x^*_{13.1}, \quad (R^2 = 0.878) \tag{37}$$

$$\hat{y}^*_{9.6,3} = \underset{(0.042)}{0.958} \cdot x^*_{13.1,1.2}, \quad (R^2 = 0.918) \tag{38}$$

$$\hat{y}^*_{9.7,3} = \underset{(0.046)}{0.921} \cdot x^*_{13.1,1.2}, \quad (R^2 = 0.849) \tag{39}$$

$$\hat{y}^*_{9.8,3} = \underset{(0.018)}{0.774} \cdot x^*_{13.1,1.2} + \underset{(1.202)}{0.172} \cdot z^*_{9.3,3}, \quad (R^2 = 0.736) \tag{40}$$

$$\hat{y}^*_{9.9,3} = \underset{(0.000)}{0.473} \cdot x^*_{13.1} + \underset{(0.003)}{0.322} \cdot x^*_{13.2}, \quad (R^2 = 0.378) \tag{41}$$

$$\hat{y}^*_{1.1} = \underset{(0.001)}{0.668} \cdot x^*_{1.1,1} + \underset{(0.001)}{0.259} \cdot x^*_{1.1,2} + \underset{(0.035)}{0.611} \cdot z^*_2, \ (R^2 = 0.874) \tag{42}$$

$$\hat{y}^*_{1.2} = \underset{(0.001)}{0.604} \cdot x^*_{1.2,1} + \underset{(0.025)}{0.346} \cdot x^*_{1.2,2}, \ (R^2 = 0.889) \tag{43}$$

$$\hat{y}^*_{2} = \underset{(0.001)}{0.396} \cdot x^*_{2,1} + \underset{(0.002)}{0.395} \cdot x^*_{2,2} + \underset{(0.001)}{0.226} \cdot z^*_{2,1}, \ (R^2 = 0.964) \tag{44}$$

$$\hat{y}^*_{3} = \underset{(0.001)}{0.170} \cdot x^*_{3,1} + \underset{(0.003)}{0.786} \cdot x^*_{3,2}, \ (R^2 = 0.888) \tag{45}$$

$$\hat{y}^*_{4.1} = \underset{(0.001)}{0.715} \cdot x^*_{4.1,1} + \underset{(0.003)}{0.168} \cdot x^*_{4.1,2} + \underset{(0.001)}{0.121} \cdot z^*_{4.1,1}, \ (R^2 = 0.966) \tag{46}$$

$$\hat{y}^*_{4.2} = \underset{(0.001)}{0.758} \cdot x^*_{4.2,2} + \underset{(0.001)}{0.229} \cdot z^*_{4.2,1}, \ (R^2 = 0.942) \tag{47}$$

$$\hat{y}^*_{4.3} = \underset{(0.001)}{0.171} \cdot x^*_{4.3,1} + \underset{(0.004)}{0.553} \cdot x^*_{4.3,2} + \underset{(0.001)}{0.138}{*} \cdot x^*_{4,1} + \underset{(0.001)}{0.141}{*} \cdot x^*_{4.2}, \ (R^2 = 0.968) \tag{48}$$

$$\underset{}{\hat{y}^*_{4.4}} = \underset{(0.001)}{0.969} \cdot x^*_{2.2}, (R^2 = 0.940) \tag{49}$$

$$\hat{y}^*_{4.5} = \underset{(0.004)}{0.576} \cdot x^*_{4.5,2} + \underset{(0.001)}{0.408} \cdot z^*_{4.5,1}, (R^2 = 0.959) \tag{50}$$

$$\hat{y}^*_{5.1} = \underset{(0.001)}{0.567} \cdot x^*_{5.1,1} + \underset{(0.004)}{0.418} \cdot x^*_{5.1,2}, (R^2 = 0.929) \tag{51}$$

$$\hat{y}^*_{5.2} = \underset{(0.001)}{0.250} \cdot x^*_{2.2} + \underset{(0.001)}{0.749} \cdot z^*_{5.2,3}, (R^2 = 0.979) \tag{52}$$

$$\hat{y}^*_{5.3} = \underset{(0.001)}{0.225} \cdot x^*_{5.3,2} + \underset{(0.001)}{0.781} \cdot z^*_{5.3,3}, (R^2 = 0.982) \tag{53}$$

$$\hat{y}^*_{5.4} = \underset{(0.001)}{0.993} \cdot x^*_{2.2} + \underset{(0.001)}{0.054} \cdot x^*_{5}, (R^2 = 0.962) \tag{54}$$

$$\hat{y}^*_{5.5} = \underset{(0.001)}{0.466} \cdot x^*_{2.2} + \underset{(0.001)}{0.518} \cdot z^*_{5.5,3}, (R^2 = 0.958) \tag{55}$$

Here is denoted: () standard errors, R^2determination coefficient.

For these models the coefficient of determination are statistical significant at 1% level. For assessment the F-test was used. For the assessment of coefficients of models we is used t-test. All coefficients are statistical significant at 5% level.

Methods for Risk Measurement of Start-Up Firms in the Conditions of Emerging Capital Markets

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

Abstract

The aim of this paper is to offer the know-how for quantifying risk, which may reflect the restrictions faced by investors in the conditions of emerging capital markets when they start up a new company. The theoretically suitable risk measurement techniques are subject to empirical testing with finding that methods on financial basis outperform those on the market basis and that the level of risk of the respective companies is particularly dependent on the combined level of operating and financial leverage. This result allowed for the construction of a new risk-quantifying technique for investors with low capital diversification, zero entrepreneurial history and access to capital market data with low information content.[3]

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

INTRODUCTION

The reason for choosing the subject of quantifying risk and the impact thereof on discount rates for start-up companies subsisted in the repeated occurrence of limits which confront users of traditional techniques for estimating the costs of capital of newly established businesses in the conditions of emerging capital markets. These specific conditions reduce the range of risk measurement techniques due to low information content of market data, zero entrepreneurial history of the new-born firms and low capital diversification of investors.

[1] Department of Regional and Business Economics, Faculty of Regional Development and International Studies, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic. Corresponding author: e-mail: gabriela.chmelikova@mendelu.cz, phone: (+420)731624694.
[2] Department of Applied Statistics, Faculty of Regional Development and International Studies, Mendel University in Brno, Czech Republic.
[3] This paper follows the article of the same authors published in the previous *Statistika: Statistics and Economy Journal* issue No. 2/2018.

Methods for quantifying risk can be divided into two groups, their main difference being the information that they are based on. Either is their basis a comprehensive capital market information, or they use the information from the companies' financial system. The flagship of capital market based techniques is Capital Asset Prising Model. Sharpe (1964) and Lintner (1965) with all coefficient beta modifications (beta on the historical basis, beta determined by analogy (Damodaran, 2009), beta determined by multicriterial correlation analysis (Fama and French, 2012; Womack and Zhang, 2003), followed by Arbitrage Pricing Theory (Ross, 1976), Derivation from the interest rate and Dividend model (Gordon, 1959). The financial techniques are represented especially by Modular model (e.g. Mařík, 2011), Method of Security Equivalents (Ballwieser, 2004), Capital Asset Pricing Model with beta on financial basis (again with different coefficient beta modifications – based on accounting return (e.g. Hill and Stone, 1980), or based on business risk fundamentals (e.g. Li and Henderson, 1991; or Toms, 2012).

The reliability of individual methods for risk quantification depends on factors that characterise the environment where these methods are used. The target group of firms for our research are start-up companies in the conditions of emerging capital markets. The character of the capital market, as well as the character of the newness of the studied sample of firms and undiversified investors, lead to the creation of several limiting factors that prevent the application of a large portion of commonly-used methods The restrictions ensuing from the character of the emerging capital markets and the character of start-up companies affecting the use of methods for quantifying risk can be summarised in the following points:

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

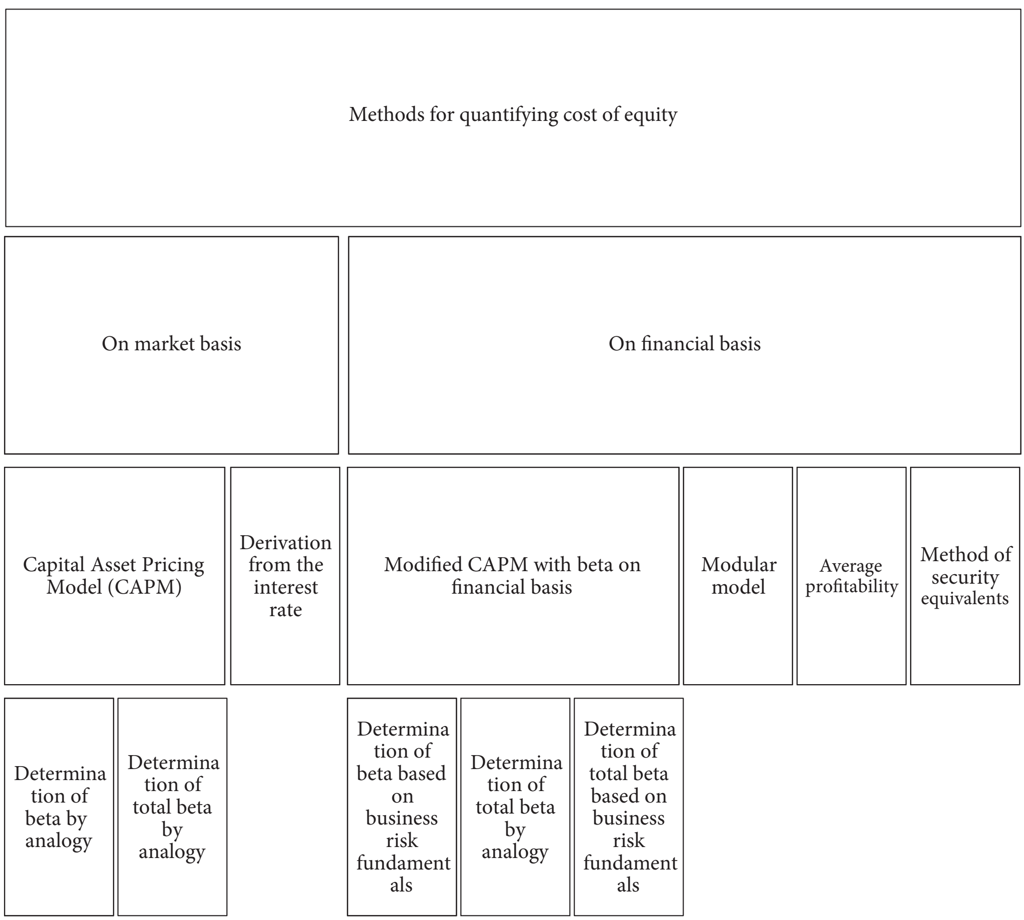
The conducted research on risk measurement techniques under these specific conditions is limited. There exist several studies focusing on businesses in particular specific conditions, but their mix is rare.

The most frequent limit, that bothers the risk measurement techniques´ applicants, is connected with the capital market imperfections and low data reliability. This is withal the area with the most intensive research. Harvey (1995) in an early research paper finds that in the emerging markets the betas are very low, which underpriceds required returns. Godfrey and Espinosa (1996) propose an adjusted CAPM, where the adjustment can be made by adjusting either beta or risk free rate. Bekaert and Harvey (1995) propose an alternative approach, where the cost of capital is allowed to vary, or to change over time in accordance with the level of market integration. Estrada (2000) proposes adjusting CAPM with downside risk methodology using the semi-standard deviation. Damodaran (2011) suggests to calculate beta using the global market index, which assumes fully integrated markets.

The risk measurement in the start-up phase of a business has attracted the interest of many researchers. Wuermseherb and Cattaneoc (2013) claim that limits resulting from the newness of these firms present an obstacle for the application of the basic forms the modified CAPM, constructed on both market and financial basis. Both forms are built on a regression analysis of return development in the studied company and relevant group of firms. Given that the profitability development of the given project during the last period is unknown, it is not possible to do a regression analysis and so identify the beta coefficient. Damodaran (2009) suggests allowing for a certain degree of generalisation, to eliminate this deficiency via the use of an analogous beta coefficient, i.e. the beta coefficient of an analogous. Baker and English (2011) report that the character of the newness also eliminates the Arbitrage Pricing Theory (APT), which is based on an analysis of the relationship between the profitability development of the given title and the relevant macro-economical quantities. Given the non-existent history of the assessed investment projects, it is impossible to apply this technique. Another method belonging to the market-based group, which is unable to reflect the limits of a non-existent history of this group of companies, is the dividend model. This model's construction is intuitively very straightforward, but it is very exacting on the quality of input data. Garrett and Priestley (2012) claim that discount rate quantification here is based on estimating the dividend development growth amount and rate, which, as a rule, is based on current company dividend policies. Given these companies' non-existent economic-activity history, this information is unknown when they start up in business and therefore a qualified estimate is impossible.

Another significant specific is the fact that start-up companies' capital is usually in the investors' hands, for whom this investment presents their only, or at least their predominant, personal investment. Analysis of capital structures showed (e.g. Gallo and Vilaseca, 1996) that start-up businesses have low debt-equity levels. This is also in accordance with findings of Chmelíková and Somerlíková (2014) who concluded that 90% of own capital is made up of internal sources. Only 10% comes into start-up companies from external investors, who are usually individual investors (business angels) or investment companies. Both these groups present a type of an investor whose investment capital is usually effectively diversified and it is mainly the systematic part of the risk that is relevant to his decision-making. However, for the prevalent type of investor shares in their own capital are complicated by a low, and in many cases non-existent, diversification of their capital sources. Seeing as a low diversification of their capital may

Figure 1 Methods usable for the quantification of discount rates for start-up companies after a reflection of market limits, a non-existent history and the specifics of non-diversified investors



Source: Own processing

be presumed with this type of an investor, it is necessary to look for methods which produce a quantification of the overall risk, not just of its systematic parts. According to McConaquahy (2008) when quantifying the capital costs of an investment that is not part of a perfectly diversified portfolio, or is in fact held independently, it is necessary to take into account the influence of company-specific factors and reflect them into the required capital costs.

Figure 1 shows an overview of suitable methods for the quantification of the discount rate for the target group of firms –companies who are starting up in the conditions of emerging capital markets.

The aim of this paper is therefore to offer a technique for quantifying risk which can reflect the above-mentioned limits of the target group of companies – the newly established businesses operating in the conditions of developing capital markets.

1 EMPIRICAL VERIFICATION OF RELEVANT TECHNIQUES

1.1 Methods and data

Partial aim of this part is to subject individual relevant (from a theoretical point of view) techniques to an empirical test, which is to verify their practical abilities in the specific conditions of the economy with emerging capital market. Verification is performed on a number of newly established companies in the Czech Republic. Individual techniques are applied retrospectively to a group of specific start-up companies in the Czech Republic resulting in the relevant risk scale at a given moment and using a given technique. This result is then confronted with the real development of the selected start-up companies after the risk evaluation date. The resulting confrontation between the real development after the chosen technique application date and the risk scale values discovered via chosen methods then offers an effective tool for evaluating the effectiveness of individual techniques.

Previous research (e.g. Chmelíková and Somerlíková, 2018) has identified that fluctuation in return to equity (Free Cash Flow to Equity – *FCFE*) is statically significantly associated with high probability of decline and hence serves as an appropriate measure of total riskiness. A retrospective approach based on the retrospective assessment of techniques has been chosen to evaluate the individual techniques devised to quantify risk. In view of the extent of databases available (especially considering the structure of the electronic database of financial statements for Czech companies), the development in *FCFE* can be observed and its fluctuation over a fixed time period in the past, for which the resulting figures are known for the degree of risk as measured by the individual techniques. Because of the mutual comparability of the observed companies and the ability to characterise the average fluctuation of a whole industry, for every company the standard deviation has been relativized by conversion to a coefficient of variation of *FCFE* variation in accordance with Formula (1).

Coefficient of variation of *FCFE* for firm *i*:

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

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

The reliability of the estimate of probable future risk can therefore be confronted with the actual development after a given point in time.

This empirical test will only be subject to methods which theoretically reflect the limits of the start-up company`s particular character. The theoretical discussion on the ability of individual techniques to incorporate the specifics of new firms in the conditions of emerging capital markets has already been covered in previous part of the paper, resulting in the methods shown in Figure 1. Its empirical evaluation will use the same structure as in Figure 1. The analysis of firm and sector specific variables is based on the data published by Bisnode in the corporate database Albertina – Gold Edition (Bisnode Czech Republic, 2012). 2008 has been chosen as the starting year for evaluating the individual techniques from when the development in *FCFE* has been monitored. The accounts data of all newly-established companies in 2008 to publish their financial statements until 2011 inclusive has been used to calculate the variation coefficient *FCFE* for the individual sections of NACE.[4] There were 2 546 companies incorporated into the researched sample that included newly established firms in the Czech Republic. In contrast to Chmelíková (2014), the weighted average of coefficients of variation for individual NACE sections was used. The weights for particular companies were calculated according to the following formula:

$$w_i = \frac{Total\ Assets_i}{Total\ Assets\ in\ the\ sector}, \qquad (2)$$

where *Total Assets*$_i$ stand for total assets of firm *i* in the year of inception and *Total Assets in the sector* represent the sum of *Total assets* of all firms in the respective NACE sector. The figures for the weighted average variation coefficients of *FCFE* for the individual sections of NACE are shown in Table 1.

The table includes the values for the total beta coefficient for CAPM models based on both the market and finance for the beginning of 2008. A mutation of the technique is being considered for both alternatives with an analogical beta coefficient. For the market-based CAPM model, this alternative is necessary especially in the environment of local capital markets, while for finance-based CAPM due to the newness of the company research sample. In order to evaluate the effectiveness of market-based CAPM, the risk coefficient for total beta has been chosen for the European capital market and was taken from the database: The Data Page, Damodaran Online (Damodaran, 2013). In order to evaluate finance-based CAPM, the coefficients of total beta were calculated in the population of existing companies for each company individually as per following formula:

$$Total\beta = \frac{\dfrac{\operatorname{cov}(ROE_j, ROE_m)}{\sigma^2(ROE_m)}}{\dfrac{\operatorname{cov}(ROE_j, ROE_m)}{\sigma(ROE_m)\sigma(ROEj)}} = \frac{\sigma(ROE_j)}{\sigma(ROE_m)}, \qquad (3)$$

where $cov(ROE_j, ROE_m)$ is the covariance between return on equity of the business j and average market return on equity in the 4-years period before 2008, $\sigma^2(ROE_m)$ is the dispersion of market returns on equity in the 4-years period before 2008, $\sigma(ROEm)$ is the standard deviation of market earnings

Table 1 The values of average total beta coefficients based on the market and finance for the individual sections of economic activity in 2007 (January 2008) and the fluctuation in weighted average coefficient of variation of FCFE for companies established in 2008 in the Czech Republic

NACE code	Description	Coefficient total market BETA – Europe 2008	Coefficient total financial BETA – Czech Republic 2008	Weighted average coefficient of variation of FCFE of businesses newly born in 2008
01	Crop and animal production, hunting and related service activities	3.49	3.586	0.456
02	Forestry and logging	3.27	4.164	0.166

[4] Nomenclature statistique des activités économiques dans la Communauté européenne.

Table 1 (continuation)

NACE code	Description	Coefficient total market BETA – Europe 2008	Coefficient total financial BETA – Czech Republic 2008	Weighted average coefficient of variation of FCFE of businesses newly born in 2008
03	Fishing and aquaculture	N/A	3.049	N/A
05	Mining of coal and lignite	3.11	4.294	0.347
06	Extraction of crude petroleum and natural gas	5.48	4.102	N/A
07	Mining of metal ores	4.48	3.656	N/A
08	Other mining and quarrying	4.67	4.739	0.412
09	Mining support service activities	N/A	4.721	0.376
10	Manufacture of food products	2.89	4.519	0.506
11	Manufacture of beverages	2.63	4.123	0.323
12	Manufacture of tobacco products	1.60	N/A	N/A
13	Manufacture of textiles	3.52	3.033	0.499
14	Manufacture of wearing apparel	3.60	3.870	0.313
15	Manufacture of leather and related products	2.24	3.245	0.167
16	Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials	3.13	2.447	0.294
17	Manufacture of paper and paper products	3.25	3.247	0.223
18	Printing and reproduction of recorded media	N/A	2.300	0.094

Table 1 (continuation)

NACE code	Description	Coefficient total market BETA – Europe 2008	Coefficient total financial BETA – Czech Republic 2008	Weighted average coefficient of variation of FCFE of businesses newly born in 2008
19	Manufacture of coke and refined petroleum products	2.20	4.250	N/A
20	Manufacture of chemicals and chemical products	3.40	4.998	0.203
21	Manufacture of basic pharmaceutical products and pharmaceutical preparations	2.78	3.711	N/A
22	Manufacture of rubber and plastic products	2.79	4.678	0.463
23	Manufacture of other non-metallic mineral products	N/A	4.934	0.485
24	Manufacture of basic metals	4,72	4.171	0.318
25	Manufacture of fabricated metal products, except machinery and equipment	4.39	3.649	0.131
26	Manufacture of computer, electronic and optical products	4.00	3.716	0.186
27	Manufacture of electrical equipment	2.16	3.105	0.218
28	Manufacture of machinery and equipment n.e.c.	3.08	3.871	0.291
29	Manufacture of motor vehicles, trailers and semi-trailers	3.51	4.147	0.428
30	Manufacture of other transport equipment	3.82	4.399	0.360
31	Manufacture of furniture	4.33	3.532	0.258
32	Other manufacturing	N/A	4.295	0.332
33	Repair and installation of machinery and equipment	N/A	1.982	0.010
35	Electricity, gas, steam and air conditioning supply	2.84	2.851	0.201
36	Water collection, treatment and supply	N/A	4.255	0.327

Table 1 (continuation)

NACE code	Description	Coefficient total market BETA – Europe 2008	Coefficient total financial BETA – Czech Republic 2008	Weighted average coefficient of variation of FCFE of businesses newly born in 2008
37	Sewerage	3.46	3.647	0.097
38	Waste collection, treatment and disposal activities; materials recovery	3.33	4.255	0.251
39	Remediation activities and other waste management services	3.55	4.180	0.326
41	Construction of buildings	3.93	4.057	0.336
42	Civil engineering	2.20	2.497	0.111
43	Specialised construction activities	2.94	3.800	0.312
45	Wholesale and retail trade and repair of motor vehicles and motorcycles	3.53	3.248	0.201
46	Wholesale trade, except of motor vehicles and motorcycles	3.43	3.347	0.191
47	Retail trade, except of motor vehicles and motorcycles	2.97	3.809	0.389
49	Land transport and transport via pipelines	3.32	3.702	0.302
50	Water transport	2.48	2.178	0.026
51	Air transport	2.67	3.321	0.365
52	Warehousing and support activities for transportation	3.09	3.826	0.411
53	Postal and courier activities	N/A	3.752	0.359
55	Accommodation	3.31	3.122	0.244
56	Food and beverage service activities	2.95	3.642	0.398
58	Publishing activities	3.42	2.464	0.124

Table 1 (continuation)

NACE code	Description	Coefficient total market BETA – Europe 2008	Coefficient total financial BETA – Czech Republic 2008	Weighted average coefficient of variation of FCFE of businesses newly born in 2008
59	Motion picture, video and television programme production, sound recording and music publishing activities	3.64	3.236	N/A
60	Programming and broadcasting activities	3.58	3.966	N/A
61	Telecommunications	2.14	3.173	0.314
62	Computer programming, consultancy and related activities	3.99	3.407	0.319
63	Information service activities	3.20	3.061	0.204
64	Financial service activities, except insurance and pension funding	3.41	4.819	0.483
65	Insurance, reinsurance and pension funding, except compulsory social security	3.38	N/A	N/A
66	Activities auxiliary to financial services and insurance activities	4.62	3.430	0.402
68	Real estate activities	3.76	2.909	0.337
69	Legal and accounting activities	3.03	4.037	0.397
70	Activities of head offices; management consultancy activities	3.21	3.399	0.132
71	Architectural and engineering activities; technical testing and analysis	3.40	2.061	0.091
72	Scientific research and development	2.76	3.660	0.459
73	Advertising and market research	3.93	3.809	0.399
74	Other professional, scientific and technical activities	N/A	2.763	0.116
75	Veterinary activities	N/A	2.563	0.090
77	Rental and leasing activities	N/A	3.979	0.285

Table 1 (continuation)

NACE code	Description	Coefficient total market BETA – Europe 2008	Coefficient total financial BETA – Czech Republic 2008	Weighted average coefficient of variation of FCFE of businesses newly born in 2008
78	Employment activities	N/A	3.542	0.276
79	Travel agency, tour operator and other reservation service and related activities	N/A	3.146	0.262
80	Security and investigation activities	N/A	2.651	0.148
81	Services to buildings and landscape activities	2.97	4.954	0.167
82	Office administrative, office support and other business support activities	3.12	4.534	0.054
84	Public administration and defence; compulsory social security	N/A	2.706	N/A
85	Education	3.91	4.649	0.301
86	Human health activities	3.44	2.561	0.440
87	Residential care activities	2.99	2.852	N/A
88	Social work activities without accommodation	3.27	3.051	0.164
90	Creative, arts and entertainment activities	3.59	1.972	0.095
91	Libraries, archives, museums and other cultural activities	N/A		0.000
92	Gambling and betting activities	3.34	1.896	0.086
93	Sports activities and amusement and recreation activities	3.23	2.019	0.159
94	Activities of membership organisations	N/A	2.810	0.315
95	Repair of computers and personal and household goods	2.69	4.447	0.316
96	Other personal service activities	N/A	3.588	0.376

Table 1 (continuation)

NACE code	Description	Coefficient total market BETA – Europe 2008	Coefficient total financial BETA – Czech Republic 2008	Weighted average coefficient of variation of FCFE of businesses newly born in 2008
97	Activities of households as employers of domestic personnel	N/A	N/A	N/A
98	Undifferentiated goods- and services-producing activities of private households for own use	3.27	N/A	N/A
99	Activities of extraterritorial organisations and bodies	N/A	N/A	N/A

Note: N/A – not available data.
Source: Own calculation based on data from: <*http://pages.stern.nyu.edu/~adamodar*> (Damodaran, 2013) and Albertina (Bisnode, 2012)

and $\sigma(ROEj)$ is the standard deviation of business´s return on equity both in the 4-years period before 2008. The individual sections are then characterised by the simple arithmetic mean for all the total beta coefficients. The figures shown in the Table 1 represent the average figures of beta coefficients for the individual sections of NACE in the population of all companies in the Czech Republic which published their financial statements in 2004, 2005, 2006 and 2007 in a row.

1.2 Results

A regression analysis has been used as the method to analyse the relationship between FCFE fluctuation for newly-established companies in 2008 and the risk scales for market- and finance-based total beta. First, the normality of the individual files of data was verified, both by using the Kolmogorov-Smirnov normality test and based on a normal probability plot. This was then subject to two regression analyses on the following variables:

- average coefficient of variation of *FCFE* (*VCoFCEF*) companies established in 2004 as independent variable and
- dependant variable *Total Market Beta* (*TMB*).

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

Table 2 Variable description and summary statistics for Total Market Beta analysis

Variable	Abbreviation	Mean	SD	Min	Max	N
Dependent Variable						
Total Market Beta	*TMB*	3.3567	0.5884	2.14	4.89	58
Independent Variable						
Variation Coefficient of Free Cash Flow to Equity	*VCoFCEF*	0.2787	0.1254	0.0255	0.5059	58

Source: Own calculation based on data from: <*http://pages.stern.nyu.edu/~adamodar*> and Albertina

And

- average coefficient of variation of *FCFE* (*VCoFCEF*) companies established in 2004 as independent variable and
- dependant variable *Total Financial Beta* (*TFB*).

All variables, including a description of the measures used and their descriptive statistics, are summarized in Table 3.

Table 3 Variable description and summary statistics for Total Financial Beta analysis

Variable	Abbreviation	Mean	SD	Min	Max	N
Dependent Variable						
Total Financial Beta	*TFB*	3.5426	0.8115	1.896	4.998	73
Independent Variable						
Variation Coefficient of Free Cash Flow to Equity	*VCoFCEF*	0.2742	0.1266	0.0102	0.5059	73

Source: Own calculation based on data from: <*http://pages.stern.nyu.edu/~adamodar*> and Albertina

The function showing the dependence of the total market beta coefficient on the FCFE average variation coefficient for companies established in 2008 takes the form $TMB' = 2.9145 + 1.5866 \cdot VCoFCEF$, with the correlation coefficient $r = 0.3381$ showing a mostly lower dependence. Therefore, the function $TFB' = 2.5784 + 3.5166 \cdot VCoFCEF$ shows the dependence of the total finance beta coefficient on the average variation coefficient for companies established in 2008. The correlation coefficient $r = 0.5488$ shows medium dependence between the variables monitored. Using the weighted average of FCFE coefficient of variation led to slightly decreased resultant values of the coefficient of correlation than in case of simple average (Chmelíková, 2014). The summary results of statistical analysis are presented in the Table 4.

Table 4 Regression analysis

Independent Variable VCoFCEF	Dependent Variable (Coefficients)	Independent Variable VCoFCEF	Dependent Variable (Coefficients)
Intercept	2.9145 ***	Intercept	2.5784 ***
	(0.0000)		(0.0000)
TMB	1.5866 **	TFB	3.5166 ***
	(0.0094)		(0.0000)
R^2	0.1143	R^2	0.3012
F-test	7.2273	F-test	30.5959
p-value	0.0094	p-value	0.0000
	< 0.001		< 0.001

Note: Standard errors in parentheses ***p<0.001, **p<0.05.
Source: Own calculations (processed in software Unistat)

The results of the correlation analysis for the total beta coefficients connected to the risk criteria chosen pointed to the closer relationship between the fluctuation in free cash flow to equity for start-up companies and the total beta for financial basis than for the market basis. These findings can be considered significant since the criterium for assessing how good is the forecasting of the individual risk indicators is connected very closely with how probable bankruptcy is for newly-established economic subjects. This is also in line with some empirical tests covering the relationship between finance and market betas. The connection in the figures of both indicators were tested several times in the past, with the individual studies mostly confirming a close interdependence between their figures (e.g. Kulkarni, Powers and Shanon, 1991; or Karels and Sackley, 1993). However, number of results which failed to confirm this close interdependence (e.g. Beaver and Manegold, 1975; or Gonedes, 1973) was presented. The degree of association fluctuated depending on the way the accounting beta indicator was calculated, as well as the length of the trial period.

The discount rates deduced from the costs of debt, the modular model or average profitability, belongs to other methods which are theoretically suitable with the limits drawn from the character of the target group of companies. The idea of a method transferring the costs to own capital from the costs of debt comes from the fact that the owners carry a higher degree of investor risk than the creditors due to the residual requirements when the company is wound up. It is therefore obvious that the earnings demanded by them should be higher than the creditors'. The difference in rates is the subject of an expert estimate. Due to this, it is not possible to verify this method at a common level among the various methods, so it is based on fixing the discount rate for investing in a share of own capital in the company in question based on the average profitability of own capital in the company`s own field. The model faces too big degree of generalising to the level of the industry`s average, which actually presents the same handicap as the other methods built on searching for an analogous firm.

The modular model is a method built on the individual preferences of the investor in the same way as the method of certainty equivalents. The method of certainty equivalents is less demanding on the investor`s knowledge of the risk factors than the modular module. When using the certainty equivalents techniques, the potential investor will 'make do with' the forecast for turnover for various world situations, their probability and the knowledge of his own attitude to risk. The quality of turnover forecast for various circumstances is critical for both determining the discount rate and also evaluating the whole project.

2 LIMITATIONS OF EXISTING TECHNIQUES AND CONSTRUCTING A MODEL SUITABLE FOR RISK ASSESSMENT WITH START-UP COMPANIES

In addition to methods emanating from reflections on the limits of the level of accessibility for information on the capital markets, it is possible to synthesise chosen unsuitable methods with their later amendments on the limiting elements to propose other techniques which will fully reflect the needs of start-up companies in the Czech environment.

The above discussion has shown that the primary requirements for a model suitable for quantifying risk and the following calculation of costs for the capital with start-up companies in transitional economies are as follows:

- an easily-predictable fluctuation rate for future earnings without being tied to the company's past and without using data from the capital market,
- the ability to reflect not only the systematic part of risk, but also its company-specific factors and to honour them accordingly.

Such requirements led to the synthesis of some of the above-mentioned principles. The first of them can be met by using the knowledge of business risk fundamentals. As mentioned above, the primary determinants of business risk include the fluctuations in demand for a company's products, the fluctuation in the end product price, the fluctuation in input prices, the ability to adapt the output price to the varying input price, the ratio of fixed operating costs in the overall cost structure – the operating leverage. Last but not least, from the business owners´ angle this includes the loading of fixed financial payments on own capital when using foreign capital – the financial leverage.

In short, it can be said that a higher share of fixed costs leads to a higher business risk. The previous empirical test (Chmelíková and Somerlíková, 2018) confirmed this supposition since it was shown that the fluctuation of future free cash flow for the owners is mostly actually explained by the level of operating and financial leverage.

Incorporating these results into CAPM principles makes it possible to produce a method which is not dissimilar to ABRM (Accounting Based Risk Management) models (Toms, 2012). Nevertheless, it will honour the overall risk faced by the owners of the separate investment in shares in own capital of start-up types of companies. As Damodaran (2009) mentioned the absence of diversification could be shown by expanding the scale of systematic risk beta by its specific part by recalculating to the so-called overall beta. In the terminology of market risk scales, the process of transfer is accompanied by separating the market beta by the correlation coefficient of historical earnings of the company and market in question. After applying the relevant mathematical operations, this results in construction of beta which is only dependent on the standard deviations of historical earnings of the market and company independent of their interdependence – see Formula (3). Therefore, the total beta is generally constructed as a ratio of the fluctuation of earnings of the investment in a share of own capital to the fluctuation in earnings of the reference group (the fluctuation in this case is expressed as a standard deviation).

Analogically, a scale for the total investment risk can be devised in compliance with this idea. This scale is connected to the fluctuation of future return for funds invested in own capital. Previous empirical research (Chmelíková and Somerlíková, 2018) showed that the fluctuation in future free cash flows for the owners is very closely connected to the starting burden of company processes by fixed payments. The higher the level of fixed liabilities (whether in the form of past investment, contracts with suppliers or creditors), the lower the ability of the company to react flexibly to changes in demand (real and nominal) and, therefore, in changes in the level of business costs (again real and nominal). The fluctuation in the future return on investment in start-up companies can therefore be simplified into the level of the degree of operating leverage (DOL) and the degree of financial leverage (DFL). The influence of both risks can also be expressed as the degree of combined risk, or degree of leverage (*DL*), which can be characterised as follows:

$$DL = \frac{Net\ profit_t \,/\, Net\ profit_{(t-1)}}{Sales_t \,/\, Sales_{(t-1)}} . \tag{4}$$

Using the construction from CAPM model for total risk beta, the coefficient of total risk based on the business risk fundamentals can be described as follows:

$$\beta_{DL} = \frac{DL_c}{DL_m} , \tag{5}$$

where β_{DL} represents the coefficient of total risk DL_c refers to the degree of leverage on the intended investment and DL_m to the degree of leverage on the reference group. The idea of constructing this model is similar for the ABRM model, only with the exception that all cases of fixed payments are reflected here (including financial).

In order to identify more clearly the level, the two forms of leverage can be combined, a more detailed description can be used for the degree of leverage, as shown by Grünwald and Holečková (2006):

$$degree\ of\ combined\ leverage = \frac{Q\ (P - VC)}{Q\ (P - VC) - FC\ (including\ interests)} , \tag{6}$$

where Q represents the quantity of production, P is the price of one unit produced, VC are the variable costs and FC are the fixed costs including interest payments. The whole equation can be rewritten in the following form using accounting value added:

$$\textit{degree of combined leverage} = \frac{\textit{accounting value added}}{\textit{accounting value added} - FC\ (\textit{including interests})}, \tag{7}$$

where FC are again fixed costs. To identify the coefficient β_{DL}, the formula for the degree of leverage can be used, while the requirements for the data entered in the model are limited by the very variables entered into the calculation. To calculate the coefficient β_{DL}, it is necessary to have the information available on the company`s planned accounting value added and the planned value of fixed costs for the investment concerned. Moreover, it is important to have the information of the average accounting added value, as well as the level of fixed costs for the reference group of companies. The total result β_{DL} for the investment concerned can then be found using the following formula:

$$\beta_{DL} = \frac{\dfrac{\textit{accounting value added}_c}{\textit{accounting value added}_c - FC_c\ (\textit{including interests})}}{\dfrac{\textit{accounting value added}_m}{\textit{accounting value added}_m - FC_m\ (\textit{including interests})}}, \tag{8}$$

where the lower index c represents data for the investment concerned and the lower index m covers the data from the reference group of companies.

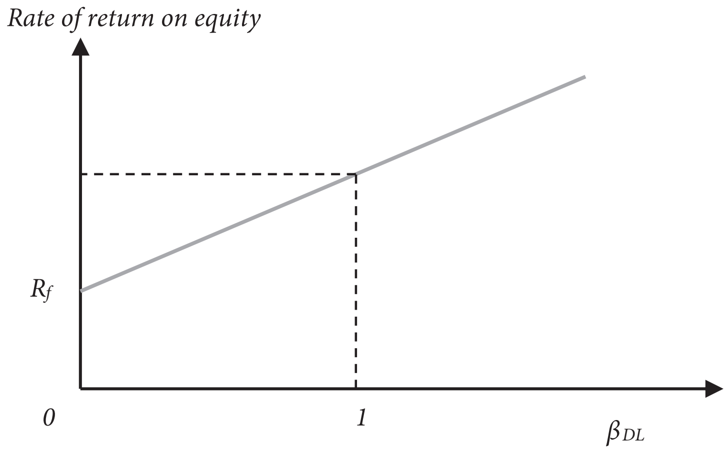
In the analogy with CAPM (the later modified version for total risk) and later with ABRM, the relationship between the total costs for own capital and the degree of combined risk can be written as follows:

$$E(Tr_c) = R_f + \beta_{DL}\,(E(p_m) - R_f), \tag{9}$$

where $E(T_{rc})$ is the total level of return-on-investment required for a share of own capital in the company in question, Rf is the risk-free rate, $E(pm)$ is the expected average profitability of the reference group of companies. The symbol β_{DL} represents the scale of the total risk of the investment.

The mechanism for estimating the total costs for own capital is similar as for the CAPM model. The risk-free rate of return corresponds to the earnings for postponing consumption and only reflects the time value of the money. The average profitability of the reference group is deduced from the nearest superior group of companies (industry, competition, national economy) as an accounting rate of return on the investment into a share for own capital. The difference between the average level of return-on-investment of the reference group of companies and the risk-free rate can be analogically termed in CAPM as the group risk premium. If the figure for the average profitability of the reference group of companies was equal the total of risk-free rate of return and the risk premium, the total risk coefficient β_{DL} for the reference group of companies must be 1. The company to reach a higher level of combined risk measured by financial and operating leverage than is usual in the reference group will look a more risky investment to investors, who can therefore expect a higher return on the money they have invested in the company. The company to reach a lower level of combined risk measured by financial and operating leverage than is usual in the reference group will look a less risky investment to investors, who can therefore expect a lower return on the money they have invested in the company than the average return in the field. The connection can be illustrated (cf. Figure 2) in almost the same way as the market line for securities, where the expected total return-on-investment rate is directly proportional to the total business risk measured by the coefficient β_{DL}.

Figure 2 The dependence of the rate of return on equity on the combination of financial and operating leverage

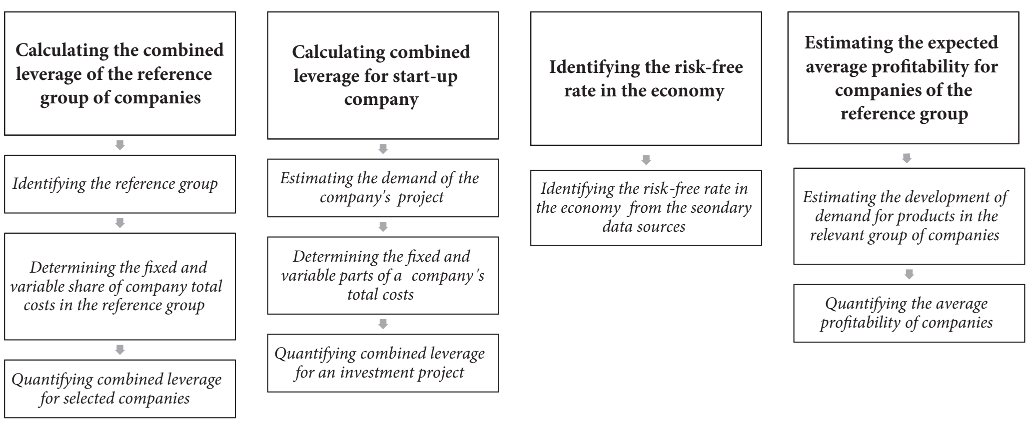


Source: Own processing based on CAPM model

The principle for assessing risk when related to other companies is very similar to the case with CAPM. The relationship between the risk-free rate of return and the average profitability of the reference group shows in the angle of the line on the graph the determining dependence on the changes in the company cost structures. The logic of the model is therefore built on the primary determinants of company risk given from the firm`s technical and financial base.

The above-mentioned connection shows that if a firm had only variable costs, thus perfectly correlated to developments in turnover, the owners would not be subjected to any proper risk with the loss of the funds invested in the company. The company would in extreme cases reach a coefficient β_{DL} of 0 and the owners would have to be content with the reward of delayed consumption.

Figure 3 The process of calculating total costs for the own capital of start-up companies



Source: Own processing based on Chmelíková (2014)

The advantage of such a formulated model for quantifying risk is its relative simplicity for entered data. Investors who use it for determining the discount rate of a start-up company can apply it based on the information they receive when the business plan is created and from publicly-available information about companies in the selected reference group. The process of calculating the discount rate can be sketched in Figure 3.

The proposed technique for quantifying capital costs requires the steps described above to be completed and the figured determined in Formula (9) to be reached. The complicated nature of the data in the individual steps of the above-mentioned technique is in line with common business practice and with the business plan for start-up companies. The individual steps of the proposed technique can be characterised in more detail in the following way.

I Calculating the combined leverage of the reference group of companies

The aim of this step is to identify the level of combined risk for the reference group of companies. This step can be completed using the following particular steps:

a) Identifying the reference group

A benchmark of average risk and profitability should be identified analogically to the CAPM model. Such a group of companies can be made up of direct competitors, similar industries or the whole national economy. The advantage of choosing a reference group extending throughout the national economy is the relative ease at acquiring data since the information on the average degree of combined leverage is available from secondary sources. Even if it is missing, the relative consistency of these figures can be assumed, which implies that it can be used reliably of some of the number of methods predicting future development.

b) Determining the fixed and variable share of company total costs in the reference group

The aim of this step is to identify the future average ratio of the variable to fixed element of total costs for companies in the reference group. The information will be used to quantify the average combined leverage for the reference group and the following calculation of the βDL coefficient. Fixed costs incorporate costs that remain constant during the monitored period despite changes in production levels. An expert analysis can be used to estimate fixed elements, as well as analysing historical costs data and even deducting from the figures for operating and financial leverage. The expert analysis is based on a specific knowledge of operating processes of the firms being researched and is therefore probably not suited to being applied to a large group of companies. An alternative to the expert analysis is to analyse the accounting records of companies, identifying their variable and fixed elements and, consequently, their forecast for the future. The problem with this approach is the difficulties on the border of the impossible to obtain the data for the financial statements, thus making it unsuited for identifying the combined leverage for a large group of firms. The last alternative is to calculate the ratio of the variable element (with up to 100% of total costs added to their fixed element) in line with the model taken from the equation (Grünwald and Holečková, 2006):

$$\frac{Net\ profit_t / Net\ profit_{(t-1)}}{Sales_t / Sales_{(t-1)}} = \frac{Q\ (P - VC)}{Q\ (P - VC) - FC\ (including\ interest\ rates)} = DL\ , \tag{10}$$

where Q represents the amount of production, P is the price for one unit of product, VCs are the individual variable costs and FC are the fixed costs including financial payments and DL is the symbol for the level of combined risk. After mathematical adjustments, the level of variable costs can be expressed in the following formula:

$$VC = Sales - (DL \cdot Net\ profit)\ . \tag{11}$$

The variables entered in the calculation for variable costs are data freely accessible for external users of accounting statements. Therefore, this approach can be applied to identify the variable element both for individual companies and for companies in the reference group.

The future level of combined risk can then be estimated either from past data on splitting costs into variable and fixed, or directly from the level of average past rate of combined risk *DL*, whose calculation is accessible directly in the reports of the accounting statements.

c) Quantifying combined leverage for selected companies

The average level of combined risk for reference group companies can be quantified by using Formula (11) and data received from points Ia and Ib.

II Calculating combined leverage for start-up companies

The aim of this step is to identify the average level of combined risk for the company in question. This can be done with the following particular steps:

a) Estimating the demand of the (company's) investment project

The starting point for quantifying the level of combined risk is a precise estimation on the future sales. A number of approaches can be used to estimate sales. These approaches will not be specified since the issue of forecasting demand and deriving an estimate of earnings from it is an extremely broad issue and its solution lies outside the scope of this article. Nevertheless, it should be mentioned that, from a practical standpoint, it is not an additional task for the owner to apply this method to start-up companies since such information should be part of the firm`s business plan.

b) Determining the fixed and variable parts of a project's total costs

As when estimating the development of sales, the development of planned costs should be part of any well-prepared business plan. To this end, this step for calculating the discount rate using the suggested method should not trouble a start-up businessman with extra data collection.

Cost classification for the volume of outputs performed is usually divided into two cost categories – variable and fixed. According to Popesko (2009), fixed costs can incorporate whatever remains unchanged with a changing amount of production during the time period. These are not only the costs connected with acquiring long-term assets, but also fixed payments connected with contracts with third parties such as creditors, employees and business partners. An example of this type of costs can include depreciation, managers' salaries, interest payments or leasing repayments. Costs which change with a change in output volume can then be termed as variable. Variable costs can include piece-work payments to blue-collar workers, consumption of material or the energy required to operate machinery.

c) Quantifying combined leverage for an investment project

The average level of combined risk for the company in question can be quantified by using Formula (12) and data received from points IIa and IIb.

III Identifying the risk-free rate in the economy

Analogically to the CAPM model, the value of risk-free rate of return enters into the calculation of total capital costs using the suggested technique. According to Mařík (2011), it can be generally said that there is no completely risk-free rate since there are no assets whose earnings would not be subject to risk. Governmental Treasury Bonds are considered to be extremely low risk in the USA in the time period related to the assessed investment.

IV Estimating the expected average profitability for companies of the reference group

a) Estimating the development of demand for products in the relevant group of companies

The performance of the whole reference group of companies should be assessed in order to estimate the development of return on investment in own capital in the reference group. For a short list of samples (branch, sector), an estimate should be made using methods which are usually applied for this process (expert estimate, trend analysis, etc.). If the whole national economy is included, the macro-economic estimates can be used for aggregate demand and then the whole economic output.

b) Quantifying the average profitability of companies

The last step required for calculating average rates of return on investment in own capital for reference group companies is to estimate future average returns on investment into a share of own capital using predicted figures for turnover. A regression analysis could be a suitable tool to analyse the relationship of the two quantities. This can be used to estimate the figures for average return in the future.

By completing all steps using the procedure recommended and introducing them into the following formula:

$$E(Tr_c) = R_f + \beta_{DL}\,(E(p_m) - R_f), \qquad (12)$$

the investor should be provided with reliable information on the level of return on investment required in the company, or for a project with zero history and an undiversified capital base. The difficulties of the input data are limited to information sources from the publicly-accessible secondary data and the business aims of the protected being assessed. Information from business plans does not present an added burden as far as the difficulties of collating data is concerned since a high-quality business aim is one of the starting points of a start-up company.

CONCLUSIONS

The aim of this paper was to offer a way of quantifying risk for new companies based on conditions in economy with emerging capital market. The point of measuring investment risk is its reflection on required rate of return of the company in question. There are two alternatives how to reflect the level of risk in the evaluation process of investments. The most often-used approach is to incorporate the level of risk in the discount rate, which includes benefits for delayed consumption and the risk undertaken. The second alternative for reflecting the risk level of investment is to adjust earnings by recalculating to the so-called security equivalents. These should then be discounted only by the risk-free part of the discount rate.

In the introduction, the circumstances limiting the application of commonly-used methods for risk quantification by newly established firms in the conditions of emerging capital markets were established. An evaluation was then made of the individual approaches with the attempt to reflect the limitations mentioned from the character of the target group of companies. The most suitable method from the theoretical point of view was discovered the CAPM with analogical total beta on financial basis, however, its application can be complicated from both a data and an algebra point of view.

Another way of applying the total beta coefficient in the real decision-making process of investing in new companies is to understand how it behaves in connection with risk fundamentals. The research question was therefore formulated as to what extend the volatility of returns of start-up companies is caused by the risk fundamentals – operating and financial leverages. Using the data of start-up companies in the Czech Republic we found statistically significant evidence for the dependence of the fluctuation of free cash flow on the combined level of risk – operating and financial leverage. This could have been expected intuitively since both forms of leverage are among the primary determinants of company risk. The factors for the fluctuation in future earnings can actually be divided into two groups. On the one hand, there are factors which affect the level of future profits, such as the level of demand and the development of input prices. On the other hand, there is the company`s ability to adapt to these changes. The ability of the company to adapt to exogenous changes is then determined by the amount it is burdened by fixed payments (operating and also financial from the owners` point of view). The business risk is therefore partly dependent on the burden of the cost structure with its fixed elements. If there is a high level of fixed costs, even a small fall in demand can cause of large drop in the return-on-investment.

By verifying the dependence of risk on the burdening of the cost structure with fixed elements, it was possible to suggest constructing a model for quantifying the discount rate for start-up companies in the conditions of an economy with emerging capital markets. The construction of this model was described in detail in the final part of this article and offers a benefit in the form of providing a new technique with relatively low demands on input data. Since new companies have an important role in the national economy, this makes it a useful tool enriching the theory which can be used in practice in real-life decision-making.

This paper presents evidence on the link between fixed payments burdening and fluctuations in returns for the owners. Knowledge of this relationship enabled construction of risk measurement technique for specific conditions of new firms in economics with emerging capital market. However, further research is also needed to examine whether or not this technique can be operationalised in real decision making processes.

ACKOWLEDGMENT
This work was supported by the IGA Mendelu under Grant No. FRRMS – IGA – 2018/006 “Role mikrofinančního sektoru v rozvoji rurálních oblastí” [Role of Microfinance sector in the development of rural areas].

References

BAKER, H. K. AND ENGLISH, P. *Using the Capital Asset Pricing Model and Arbitrage Pricing Theory in Capital Budgeting, in Capital Budgeting Valuation: Financial Analysis for Today's Investment Projects.* Hoboken, NJ, USA: John Wiley & Sons, Inc., 2011.

BALLWIESER, W. *Unternehmensbewerung – Prozesse, Methode und Probleme.* Stuttgart: Schäffer-Poeschel Verlag, 2004.

BEAVER, B. AND MANEGOLD, J. The Association Between Market-Determined and Accounting-Determined Measures of Systematic Risk: Some Further Evidence. *Journal of Financial and Quantitative Analysis*, 1975, 10, pp. 213–284.

BEKAERT, G. AND HARVEY, C. Time-Varying World Market Integration. *Journal of Finance*, 1995, 50(2), pp. 403–444.

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

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

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

CHMELÍKOVÁ, G. AND SOMERLÍKOVÁ, K. Risk Substance of Newly Established Businesses [online]. S*tatistika: Statistics and Economy Journal*, 2018, Vol. 98, No. 2, pp. 135–149.

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

DAMODARAN, A. *Equity Risk Premiums (ERP): Determinants, Estimation and Implications – the 2011th Ed.* [online]. 2011. *<http://ssrn.com/ abstract=1769064>*.

DAMODARAN, A. *The Data Page. Damodaran Online* [online]. New York: Stern New York University, 2013. [cit. 28.7.2013]. *<http://pages.stern.nyu.edu/~adamodar>*.

ESTRADA, J. *The Cost of Equity in Emerging Markets: A Downside Risk Approach (August 2000)* [online]. EFMA 2000 Athens, EFA 2000 London, FMA 2000 Edinburgh. *<http://ssrn.com/abstract=170748 orhttp://dx.doi.org/10.2139/ssrn.170748>*.

FAMA, E. F. AND FRENCH, K. R. Size, Value, and Momentum in International Stock Returns. *Journal of Financial Economics*, 2012, 105, pp. 457–472.

GALLO, M. A. AND VILASECA, A. Finance in Family Business. *Family Business Review*, 1996, 9, pp. 387–401.

GARRETT, I. AND PRIESTLEY, R. Dividend Growth, Cash Flow, and Discount Rate News. *Journal of Financial and Quantitative Analysis*, 2012, 47, pp. 1003–1028.

GODFREY, S. AND ESPINOSA, R. Practical Approach to Calculating Costs of Equity for Investment in Emerging Markets. *Journal of Applied Corporate Finance*, 1996, 9(3), pp. 80–89.

GONEDES, N. Evidence on the Information Content of Accounting Massages: Accounting-Based and Market-Based Estimate of Systematic Risk. *Journal of Financial and Quantitative*, 1973, pp. 407–444.
GORDON MYRON J. Dividends, Earnings and Stock Prices. *Review of Economics and Statistics (The MIT Press)*, 1959, 41(2), pp. 99–105.
GRÜNWALD, R. AND HOLEČKOVÁ, J. *Finanční analýza a plánování.* University of Economics in Prague: Oeconomica, 2006.
HARVEY, C. Predictable Risk and Returns in Emerging Markets. *Review of Financial Studies*, 1995, 8(3), pp. 773–816.
HILL, N. AND STONE, B. Accounting Beta, Systematic Operating Risk, and Financial Leverage: A Risk-Composition Approach to the Determinants of Systematic Risk. *Journal of Financial and Quantitative Analysis*, 1980, 15, pp. 595–637.
KARELS, G. V. AND SACKLEY, W. H. The Relationship Between Market and Accounting Betas for Commercial Banks. *Review of Financial Economics*, 1993, 2, pp. 59–72.
KULKARNI, M., POWERS, M., SHANNON, D. The Use of Segment Earnings Betas in the Formation of Divisional Hurdle Rates. *Journal of Business Finance and Accounting*, 1991, 18, pp. 497–512.
LI, R.-J. AND HENDERSON, G. Combined Leverage and Stock Risk. *Quarterly Journal of Business and Economics*, 1991, Vol. 30, No. 1, pp.18–40.
LINTNER, J. The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets. *Review of Economics and Statistics*, 1965, 47, pp. 13–37.
MAŘÍK, M. et al. *Metody oceňování podniku. Proces ocenění, základní metody a postupy.* Ekopress, 2011.
MCCONAQUAHY, D. L. *The Cost of Capital for the Closely-held, FamilyControlled Firm.* Presented at the Family Firm Institute London By Daniel L. McConaughy, PhD., Northridge: California State University, October 29, 2008.
POPESKO, B. *Moderní metody řízení nákladů.* Prague: Grada Publishing, 2009.
ROSS, S. A. The Arbitrage Theory of Capital Asset Pricing. *Journal of Economic Theory*, 1976, 13(3), pp. 341–360.
SHARPE, W. F. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, 1964, 19(3), pp. 425–442.
TOMS, S. *Accounting based risk measurement: An alternative to CAPM derived discount factors.* Working Paper No. 68, University of York, The York Management School, 2012,
WOMACK, K. L. AND ZHANG, Y. *Understanding Risk and Return, the CAPM, and the Fama-French Three-Factor Model.* Working paper 03-111, Dartmouth: Tuck School of Business, 2003.
WUERMSEHERB, M. AND CATTANEOC, G. Valuation of Early Stage High-tech Start-up Companies. *International Journal of Business*, 2013, 18(3), pp. 216–231.

The Priestley-Chao Estimator of Conditional Density with Uniformly Distributed Random Design

Kateřina Konečná [1] | *Brno University of Technology, Brno, Czech Republic*

Abstract

The present paper is focused on non-parametric estimation of conditional density. Conditional density can be regarded as a generalization of regression thus the kernel estimator of conditional density can be derived from the kernel estimator of the regression function. We concentrate on the Priestley-Chao estimator of conditional density with a random design presented by a uniformly distributed unconditional variable. The statistical properties of such an estimator are given. As the smoothing parameters have the most significant influence on the quality of the final estimate, the leave-one-out maximum likelihood method is proposed for their detection. Its performance is compared with the cross-validation method and with two alternatives of the reference rule method. The theoretical part is complemented by a simulation study.[2]

Keywords	***JEL code***
Priestley-Chao estimator of conditional density, random design, uniform marginal density, bandwidth selection, maximum likelihood method, reference rule method	*C14*

INTRODUCTION

Kernel smoothing is still a popular non-parametric procedure, in theory as well as in practice. There are numerous monographs concerned with the kernel smoothing approach, e.g., Wand and Jones (1994). Computational implementations in MATLAB were developed by Horová et al. (2012). The present paper focuses on the kernel conditional density estimation. Several estimator types can be found in the literature with the Nadaraya-Watson one being probably best known (see Rosenblatt, 1969). The local linear estimator of conditional density was suggested by Fan et al. (1996) for its better statistical properties and boundary effects.

Conditional density can be regarded as a generalization of regression, which models the conditional mean while conditional density models the whole distribution. This is the reason why a kernel regression estimator can be generalized to a kernel conditional density estimator. The present paper extends the

[1] Institute of Mathematics and Descriptive Geometry, Faculty of Civil Engineering, Žižkova 17, 602 00 Brno, Czech Republic. E-mail: konecna.k@fce.vutbr.cz.

[2] This article is based on contribution at the conference *Robust 2018*.

Priestley-Chao regression estimator (for detailed information see Priestley and Chao, 1972) to estimate even conditional densities.

Each kernel estimator depends on the smoothing parameters called bandwidths, values which significantly influence the final estimation. This is the reason why so much importance is given to their selection. There are many methods discussed in the literature, most of them suggested for the Nadaraya-Watson estimator, with only a few of them for the local linear estimator.

Introduced by Fan and Yim (2004), Hansen (2004) and Hall et al. (2004) and based on minimizing the Integrated Squared Error, cross-validation is a method typical of bandwidth selection. Bashtannyk and Hyndman (2001) suggested a reference rule method for normal underlying conditional density and for two marginal density choices – normal and uniform. Some methods extend the methods suggested for kernel regression. The iterative method proposed by Konečná and Horová (2014), for one, is motivated by the iterative method developed for kernel density estimation and for kernel regression (for detailed information, see Horová and Zelinka (2007), Horová et al. (2012), Koláček and Horová (2012)). Other examples include the bootstrap method by Bashtannyk and Hyndman (2001) and Fan and Yim (2004) as well as the fast dual-tree based algorithms using a maximum likelihood criterion (see Holmes et al., 2012).

Kernel conditional density estimation is still employed in practice: Takeuchi et al. (2009) show its application in medicine (the relative change in spinal bone mineral density is explored as a function of the age of adolescents), Jeon and Taylor (2012) are interested in 1-to-72-hours ahead wind-power prediction from which the management of wind farms and electricity systems can profit. Another application, forecasting electricity smart meter data, helping consumers to analyze and to minimize their electricity consumption and enabling new pricing strategies for suppliers, is introduced by Arora and Taylor (2016).

As mentioned above, papers are focused primarily on the Nadaraya-Watson or the local linear estimator. The present paper suggests the Priestley-Chao estimator for the uniformly distributed design, based on the estimator suggested for the equally spaced design (see Konečná, 2017). The leave-one-out maximum likelihood method follows the one proposed by Konečná (2018).

The paper is organized as follows: Section 1 deals with the Priestley-Chao estimator of conditional density and its statistical properties. The optimal values of the smoothing parameters are derived, and the leave-one-out maximum likelihood method for their practical estimation proposed in Section 2. This method is complemented by the cross-validation method and by two alternatives of the reference rule method. A simulation study in Section 3 then presents the performance of the methods by a simulation study. The proofs of the statistical properties can be found in the Appendix.

1 THE PRIESTLEY-CHAO ESTIMATOR OF CONDITIONAL DENSITY

The conditional density $f(y|x)$ models the probability of a random variable Y given a random variable X, represented by a fixed observation $X = x$. Let $\{(X_i,Y_i), i = 1,\ldots,n\}$ be an observed data sample of a pair of real random variables (X,Y). The kernel estimate of conditional density generally takes the form:

$$\hat{f}\left(y\middle|x; h_x, h_y\right) = \sum_{i=1}^{n} w_i(x) K_{h_y}(y - Y_i), \tag{1}$$

where $w_i(x)$ is a weight function, and K is a real, symmetric, nonnegative kernel function satisfying:

$$\int_R K(x)\,dx = 1, \int_R xK(x)\,dx = 0, \int_R x^2K(x)\,dx = \beta_2(K) \neq 0. \tag{2}$$

The present paper uses the Gaussian kernel. The smoothing parameters $h_x > 0$, $h_y > 0$ control the smoothness of the estimate. The estimate of conditional density is also influenced by the estimator type (1).

Our focus is on the Priestley-Chao estimator, originally proposed for the kernel regression estimation (Priestley and Chao, 1972). Konečná (2017) dealt with the Priestley-Chao estimator for conditional density with the fixed design, i.e., the fixed values $x_i = \frac{i}{n}, i = 1, \dots, n$ of the design variable X were assumed.

Next, we are concerned with the estimator for the random design specified by a uniformly distributed variable X on the interval [0,1]. The estimator can easily be extended for the design variable X on the interval $[a, b]$, $a < b$. The statistical properties of the estimator will be given and methods for bandwidth detection proposed.

Let X be a uniformly distributed random variable with the marginal density function:

$$g(x) = \begin{cases} 1, & x \in [0,1], \\ 0, & \text{otherwise.} \end{cases}$$

As the focused weight function in Formula (1) is $w_i^{PC}(x) = \frac{1}{n} K_{h_x}(x - X_i)$, the Priestley-Chao estimator takes the form:

$$\hat{f}_{PC}(y|x; h_x, h_y) = \frac{1}{n} \sum_{i=1}^{n} K_{h_x}(x - X_i) K_{h_y}(y - Y_i). \tag{3}$$

The Priestley-Chao estimator of the regression function is expressed by the conditional mean of the Formula (3):

$$\widehat{m}_{PC}(x; h_x) = \frac{1}{n} \sum_{i=1}^{n} K_{h_x}(x - X_i) Y_i.$$

Theorem 1 *Let X be a uniformly distributed random variable on the interval* [0,1], *Y be a random variable with density $f(y|x)$ being at least twice continuously differentiable, and $K(x)$ be a real, symmetric, nonnegative kernel function satisfying (2). For $x \in [h_x, 1 - h_x]$, $h_x \to 0$, $h_y \to 0$ and $nh_x h_y \to \infty$ as $n \to \infty$, the asymptotic bias* (AB) *and the asymptotic variance* (AV) *are given by:*

$$\mathrm{AB}\{\hat{f}_{PC}(y|x; h_x, h_y)\} = \frac{1}{2} h_x^2 \beta_2(K) \frac{\partial^2 f(y|x)}{\partial x^2} + \frac{1}{2} h_y^2 \beta_2(K) \frac{\partial^2 f(y|x)}{\partial y^2},$$

$$\mathrm{AV}\{\hat{f}_{PC}(y|x; h_x, h_y)\} = \frac{1}{nh_x h_y} R^2(K) f(y|x),$$

where: $R_2(K) = \int_R K_2(u)\, du$.

Proof. The proof can be found in the Appendix.

The local quality of the estimate at the point $[x,y]$ is given by the mean squared error (MSE) which is the simple decomposition to the variance (V) and the squared bias (SB). Considering the main terms only, the asymptotic MSE (AMSE) is obtained as:

$$\begin{aligned} &\mathrm{AMSE}\{\hat{f}_{PC}(y|x; h_x, h_y)\} = AV\{\hat{f}_{PC}(y|x; h_x, h_y)\} + ASB\{\hat{f}_{PC}(y|x; h_x, h_y)\} \\ &= \frac{1}{nh_x h_y} R^2(K) f(y|x) + \left(\frac{1}{2} h_x^2 \beta_2(K) \frac{\partial^2 f(y|x)}{\partial x^2} + \frac{1}{2} h_y^2 \beta_2(K) \frac{\partial^2 f(y|x)}{\partial y^2} \right)^2. \end{aligned} \tag{4}$$

The statistical properties of the Formula (3), particularly the global quality measure expressed by the asymptotic mean integrated squared error (AMISE), are necessary for assessing the quality of the estimate and the theoretical values of the smoothing parameters. AMISE is obtained by integrating (4) weighted by the marginal density $g(x)$ as:

$$\mathrm{AMISE}\{\hat{f}_{PC}(\cdot\,|\,\cdot;h_x,h_y)\} = \iint \mathrm{AMISE}\{\hat{f}_{PC}(y|x;h_x,h_y)\}g(x)\,dx\,dy.$$

The following form of the AMISE is more succinct for further processing:

$$\mathrm{AMISE}\{\hat{f}(\cdot\,|\,\cdot;h_x,h_y)\} = \frac{1}{nh_xh_y}c_1 + c_2h_x^4 + c_3h_y^4 + c_4h_x^2h_y^2, \tag{5}$$

where the constants c_1, c_2, c_3, c_4 are given by:

$$\begin{aligned} c_1 &= R^2(K),\\ c_2 &= \frac{1}{4}\beta_2^2(K)\iint\left(\frac{\partial^2 f(y|x)}{\partial x^2}\right)^2 dx\,dy\,,\\ c_3 &= \frac{1}{4}\beta_2^2(K)\iint\left(\frac{\partial^2 f(y|x)}{\partial y^2}\right)^2 dx\,dy\,,\\ c_4 &= \frac{1}{2}\beta_2^2(K)\iint\frac{\partial^2 f(y|x)}{\partial x^2}\,\frac{\partial^2 f(y|x)}{\partial y^2}\,dx\,dy. \end{aligned} \tag{6}$$

Remark. Note that all the integrals with respect to x are computed over the support of the X variable, i.e., over the interval [0,1]. The integrals with respect to y are considered over R.

2 METHODS FOR BANDWIDTH SELECTION

The values of the smoothing parameters have an essential significance for the final estimate of conditional density. First, the optimal widths of the smoothing parameters are derived as the values minimizing the AMISE. As the optimal bandwidths depend on the true conditional and marginal density function, it is necessary to develop a data-driven method for their estimation. In this section, the leave-one-out maximum likelihood method, the cross-validation method, and two alternatives of the reference rule method are suggested for their detection.

2.1 Optimal values of the smoothing parameters

The optimal values of the smoothing parameters are given as the values which minimize AMISE given by (5). By differentiating (5) with respect to h_x and h_y and setting the derivatives to 0, we obtain the following system of non-linear equations:

$$\begin{aligned} -\frac{1}{nh_x^2h_y}c_1 + 4c_2h_x^3 + 2c_4h_xh_y^2 &= 0,\\ -\frac{1}{nh_xh_y^2}c_1 + 4c_3h_y^3 + 2c_4h_x^2h_y &= 0. \end{aligned} \tag{7}$$

Solving system (7), the optimal bandwidths are given by:

$$\begin{aligned} h_x^* &= n^{-1/6}c_1^{1/6}\left(4\left(\frac{c_2^5}{c_3}\right)^{1/4} + 2c_4\left(\frac{c_2}{c_3}\right)^{3/4}\right)^{-1/6},\\ h_y^* &= \left(\frac{c_2}{c_3}\right)^{1/4}h_x^*. \end{aligned} \tag{8}$$

Both h_x^* and h_y^* are of order $n^{-1/6}$ while the order of AMISE is $n^{-2/3}$.

2.2 The leave-one-out maximum likelihood method

As mentioned above, with a real dataset, a data-driven method is needed for bandwidth selection. We will modify the maximum likelihood method, which is a standard statistical procedure for estimating unknown parameters. This method was originally proposed for kernel density estimation by Leiva-Murillo and Artes-Rodriguez (2012), and their approach is generalized to include the Priestley-Chao estimator of conditional density.

Since the objective function:

$$L(h_x, h_y) = \prod_{j=1}^{n} \frac{1}{n} \sum_{i=1}^{n} K_{h_x}(X_j - X_i) K_{h_y}(Y_j - Y_i) \tag{9}$$

is considered for all n observations, the optimization problem $L \to \max$ has a trivial solution. If $i = j$ in (9), the objective function (9) increases to infinity for $h_x \to 0$ and $h_y \to 0$. Of course, this is not the desired behaviour because, with very small values of bandwidths, the final estimate tends to be undersmoothed.

This problem can be solved by leaving out one observation and employing the modified objective function:

$$L^*(h_x, h_y) = \prod_{j=1}^{n} \frac{1}{n} \sum_{i=1, i \neq j}^{n} K_{h_x}(X_j - X_i) K_{h_y}(Y_j - Y_i)\,. \tag{10}$$

If the natural logarithm of the likelihood function L^* given by (10) is taken into account, the values of the smoothing parameters maximize:

$$l^*(h_x, h_y) = \sum_{j=1}^{n} \ln\left(\frac{1}{n} \sum_{i=1, i \neq j}^{n} K_{h_x}(X_j - X_i) K_{h_y}(Y_j - Y_i) \right),$$

and are developed as:

$$(\hat{h}_x, \hat{h}_y) = \arg \max_{(h_x, h_y)} l^*(h_x, h_y).$$

2.3 The leave-one-out cross-validation method

The cross-validation method is a standard procedure for bandwidth selection in kernel smoothing. Introduced by Fan and Yim (2004), Hansen (2004) and Hall et al. (2004), the method is associated with the global quality measure of the estimator, with the integrated squared error (ISE). With $\hat{f}_{PC,-i}(Y_i | X_i; h_x, h_y)$, being the estimate at the point (X_i, Y_i) using the points $\{(X_j, Y_j), j \neq i\}$, the cross-validation function:

$$\mathrm{CV}(h_x, h_y) = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{\substack{j=1, \\ j \neq i}}^{n} K_{h_x\sqrt{2}}(X_i - X_j) K_{h_y\sqrt{2}}(Y_i - Y_j) - \frac{2}{n} \sum_{i=1}^{n} \hat{f}_{PC,-i}(Y_i | X_i; h_x, h_y),$$

is the proper estimator of the ISE.

The values of the smoothing parameters are given by:

$$(\hat{h}_x, \hat{h}_y) = \arg \min_{(h_x, h_y)} \mathrm{CV}(h_x, h_y).$$

2.4 The reference rule method

The reference rule method was originally proposed for the Nadaraya-Watson estimator by Bashtannyk and Hyndman (2001). They assumed a normally distributed random variable $Y|(X = x)$ with linear conditional mean and constant or linear standard deviation. Additionally, they distinguished two possibilities for the marginal distribution, considering uniform and truncated normal marginal densities.

Our approach via the Priestley-Chao estimator corresponds to the choice of a uniform marginal density. We assume that the conditional distribution is normal with the mean $m(x)$ and standard deviation $\sigma(x)$. Hence, the conditional density of $Y|(X = x)$ is:

$$f(y|x) = \frac{1}{\sigma(x)\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{y-m(x)}{\sigma(x)}\right)^2\right\}. \tag{11}$$

Two different situations are considered:
(a) According to Bashtannyk and Hyndman (2001), the model with the linear conditional mean $m(x) = p_0 + p_1 x$ and the linear standard deviation $\sigma(x) = q_0 + q_1 x$ is assumed. The values of the constants $c_1, \dots, c_4$ are given by the expressions:

$$\begin{aligned} c_1 &= R^2(K), \\ c_2 &= \frac{3}{512}\beta_2^2(K)\frac{wz}{q_1\sqrt{\pi}}, \\ c_3 &= \frac{3}{128}\beta_2^2(K)\frac{z}{q_1\sqrt{\pi}}, \\ c_4 &= \frac{3}{128}\beta_2^2(K)\frac{z(2p_1 - 3q_1^2)}{q_1\sqrt{\pi}}, \end{aligned} \tag{12}$$

where $z = \frac{(q_0+q_1)^4 - q_0^4}{(q_0+q_1)^4 q_0^4}$ and $w = 19q_1^4 + 4p_1^4 + 28p_1^2q_1^2, p_1 \neq 0.$
The values of the smoothing parameters are obtained by substituting (12) into (8).
(b) The model with the quadratic conditional mean $m(x) = p_0 + p_1 x + p_2 x^2$ and the constant standard deviation σ is suggested. The constants $c_1, \dots, c_4$ are given by:

$$\begin{aligned} c_1 &= R^2(K), \\ c_2 &= \frac{3}{160}\ \beta_2^2(K)\frac{1}{\sigma^5\sqrt{\pi}}(16p_2^4 + 40p_1p_2^3 + 40p_1^2p_2^2 + 20p_1^3p_2 + 5p_1^4 + 10p_2^2\sigma^2), \\ c_3 &= \frac{3}{32}\beta_2^2(K)\frac{1}{\sigma^5\sqrt{\pi}}, \\ c_4 &= \frac{1}{16}\beta_2^2(K)\frac{1}{\sigma^5\sqrt{\pi}}(4p_2 + 6p_1p_2 + 3p_1^2). \end{aligned} \tag{13}$$

The values of the smoothing parameters are obtained by substituting the terms (13) into (8).
Remark. The expressions (13) are obtained by differentiating (11) twice and substituting them into (6). As the computations of the integrals in (6) include many auxiliary derivations, only a sketch of them is presented.

A conditional random variable $Y|(X = x) \sim N(m(x), \sigma^2)$ with the density function $f(y|x)$ is assumed. The following equality:

$$f^2(y|x) = \frac{1}{2\pi\sigma^2}\exp\left\{-\left(\frac{y-m(x)}{\sigma}\right)^2\right\} = \frac{1}{2\sigma\sqrt{\pi}}\frac{1}{\frac{\sigma}{\sqrt{\pi}}\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\left(\frac{y-m(x)}{\frac{\sigma}{\sqrt{2}}}\right)^2\right\}$$

$$= \frac{1}{2\sigma\sqrt{\pi}}f_1(y|x),$$

where $f_1(y|x)$ is a density function of a conditional random variable $Y_1|(X = x) \sim N(m(x), \frac{1}{2}\sigma^2)$, was used. For evaluating the integrals in (6), the below auxiliary expression was derived:

$$\iint_A x^k y^n f_1(y|x)\,dx\,dy = K_{0,n} + \sum_{i=1}^{2n+1}\frac{K_{i,n}}{k+i}, k = 0, \dots, 4,$$

where $A = R \cdot [0,1]$ is the domain of integration,

$$K_{0,n} = \begin{cases} 0 & n = 0,1, \\ \frac{1}{2}\sigma^2 \frac{1}{k+1} & n = 2, \\ \frac{3}{2}\sigma^2 \sum_{i=1}^{n} \frac{p_{i-1}}{k+i} & n = 3, \\ \frac{3}{4}\sigma^4 \frac{1}{k+1} + 3\sigma^2 \sum_{i=1}^{n+1} \frac{K_{i,n-2}}{k+i} & n = 4, \end{cases}$$

$$K_{i,n} = \begin{cases} \sum_{j=1}^{\min\left\{n+2-\left\lceil\frac{i}{2}\right\rceil,\left\lceil\frac{i}{2}\right\rceil\right\}} \binom{n}{\left\lceil\frac{i}{2}\right\rceil + j - 2} \binom{\left\lceil\frac{i}{2}\right\rceil + j - 2}{\left\lceil\frac{i}{2}\right\rceil - j} p_0^{n-\left\lceil\frac{i}{2}\right\rceil - j + 2} p_1^{2j-2} p_2^{\left\lceil\frac{i}{2}\right\rceil - j} & \text{for } i \text{ odd}, \\ \sum_{j=1}^{\min\left\{n+1-\frac{i}{2},\frac{i}{2}\right\}} \binom{n}{\frac{i}{2} + j - 1} \binom{\frac{i}{2} + j - 1}{\frac{i}{2} - j} p_0^{n-\frac{i}{2}-j+1} p_1^{2j-1} p_2^{\frac{i}{2}-j} & \text{for } i \text{ even}, \end{cases}$$

and $\lceil \cdot \rceil$ denotes the ceiling function.

3 SIMULATION STUDY

In this section, a simulation study comparing four methods for bandwidth estimation is conducted. The considered methods are the maximum likelihood method (ML), the cross-validation method (CV), the reference rule method with linear conditional mean and linear standard deviation (REF1), and the reference rule method with quadratic conditional mean and constant standard deviation (REF2). Two models are involved in the simulation study. To demonstrate the adaptability of the methods to various shapes of the regression function or conditional density, a changing shape of the conditional mean is presented in the first model. In the second model, a bimodal and non-symmetric conditional distribution is chosen as a mixture of two normal densities. The models are defined as:

$$M_1: Y_i = \sin(\pi X_i^2) + X_i + \varepsilon_i, X_i \sim \text{unif}(0,2), \ \varepsilon_i \sim \text{N}(0, 0.5^2), \text{i} = 1, \dots, \text{n},$$

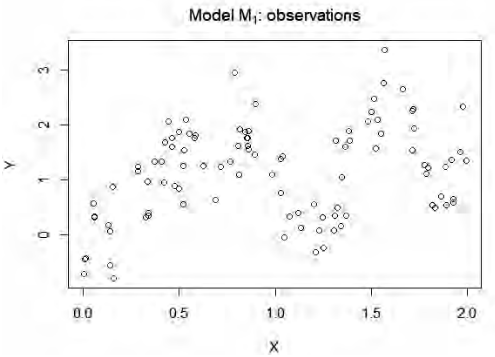
$$M_2: Y_i = \exp\{X_i\} + \varepsilon_i, X_i \sim \text{unif}(0,3), \varepsilon_i | X_i \sim W_i T_i + (1 - W_i) U_i, T_i \sim N(2X_i, 1),$$

$$U_i \sim N(-3X_i, 2), P(W_i = 0) = P(W_i = 1) = 0.5, i = 1, \dots, n.$$

In both simulation studies, one hundred observations ($n = 100$) were generated. The described methods for bandwidth selection are compared from several points of view – the estimates of the smoothing parameters and the quality measure of the estimate. The quality of the final estimate was measured by an estimate of the integrated squared error (ISE) given by:

$$\widehat{\text{ISE}}\{\hat{f}_{PC}(\cdot \mid \cdot; h_x, h_y)\} = \frac{\Delta}{n} \sum_{j=1}^{N} \sum_{i=1}^{n} \left(\hat{f}_{PC}(y_j | X_i; h_x, h_y) - f(y_j | X_i)\right)^2,$$

Figure 1 A scatterplot of one hundred observations of model M_1



Source: The author's own construction

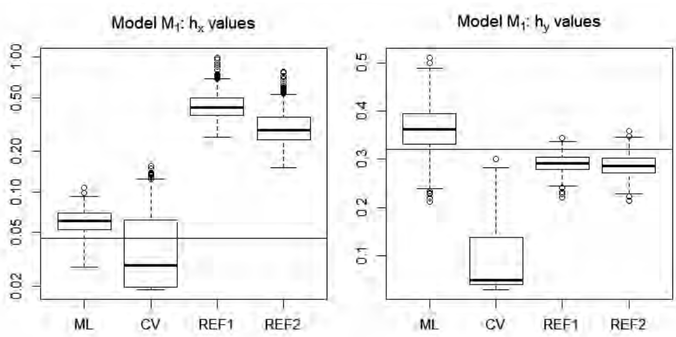
where $\boldsymbol{y} = (y_1, \dots , y_N)$ is a vector of equally spaced values over the sample space of Y and Δ is the distance between two consecutive values of y, i.e. $\Delta = y_{j+1} - y_j$, $j = 1, \dots , N - 1$. In both simulation studies, the number of $\boldsymbol{y}$ values was set to $N = 100$.

For both models, five hundred repetitions have been made to obtain the described characteristics. The results are displayed in boxplots and supplemented by numerical values in the text.

First, the results for the model M_1 are summarized. A scatterplot of one set of the sample values of model M_1 is displayed in Figure 1.

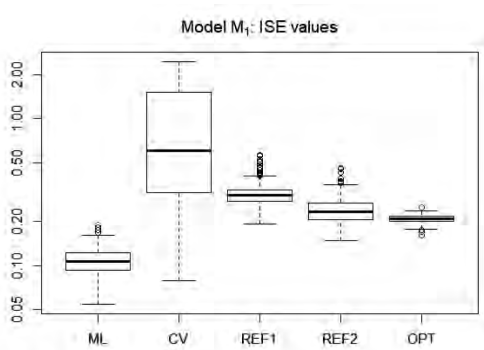
Boxplots of the estimates of the smoothing parameters h_x and h_y are displayed in Figure 2. It can be seen, that the ML and CV methods lead to good values for h_x (medians are 0.0613 for ML and 0.0288 for CV), the REF1 and REF2 methods produce highly variable bandwidths exceeding the optimal value $h_x^* = 0.0455$.

Figure 2 Estimates of the smoothing parameters h_x and h_y along with the optimal values (horizontal lines) for the ML, CV, REF1, and REF2 methods in model M_1



Note: The log-scale of the vertical axis in the left-hand-side panel.
Source: The author's own construction

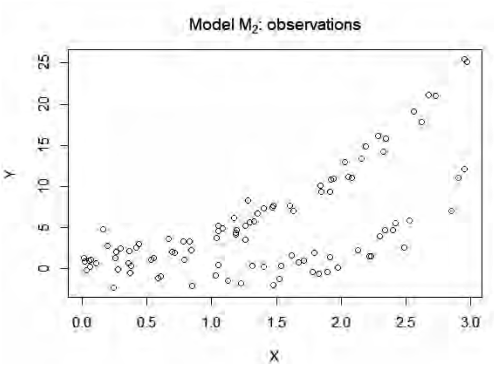
Figure 3 Estimates of the ISE values (expressed in the log-scale) for the ML, CV, REF1, and REF2 method and for the optimal bandwidth choice (OPT) in model M_1



Source: The author's own construction

Considering the estimates of the smoothing parameter h_y, the values estimated using REF1 resemble those with REF2. Their medians 0.2927 and 0.2863 (in this order) are close to the optimal value $h_y^* = 0.3213$, and the estimates are characterized by a low variability (their standard deviations are close to 0.02). The ML method gives slightly higher values than the optimum h_y^*, the median of the values being 0.3631. The CV method tends to produce values well under the optimal value which results in a much undersmoothed estimate of conditional density.

Figure 4 A scatterplot of 100 observations by model M_2



Source: The author's own construction

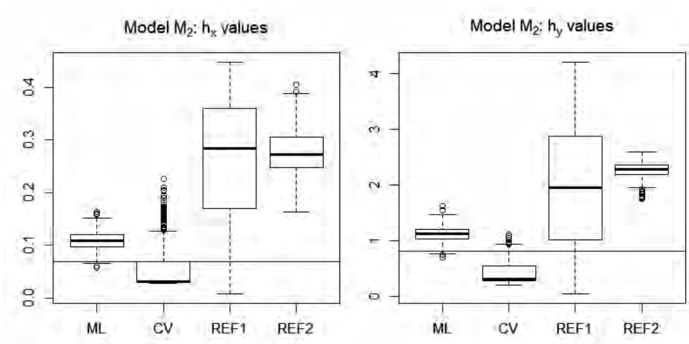
The ML method results in an ISE that is smaller than any other methods considered as well as the optimal bandwidth choice (OPT). Both reference rule methods produce ISE values slightly higher than OPT while the values of the CV method are well above the optimal bandwidth choice.

Now, we focus on the results of model M_2. The scatterplot of a sample generated by model M_2 is shown in Figure 4.

Boxplots of bandwidth estimates are shown in Figure 5. Both reference rule methods produce values of h_x and h_y well above the optimal values, which results in an overvalued final estimate with worse capability to adapt to bimodal conditional density. On the other hand, the low values of both smoothing parameters obtained by the CV method lead to the undersmoothing of conditional density and abundance of useless information in the data. The ML method performs the best in this simulation study.

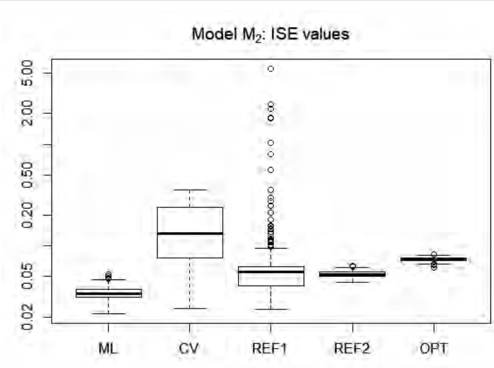
The ML method is also suitable in terms of the ISE (see Figure 6). The medians of the ISE values obtained by the reference rule methods (0.0555 for REF1 and 0.0527 for REF2) do not reach the median (0.0742) of the ISE values for optimal bandwidths, but the REF1 method suffers from the large variability (standard deviation is 0.3143). The CV method provides highly variable ISE values exceeding the OPT values.

Figure 5 Estimates of the smoothing parameters h_x and h_y and the optimal values (horizontal lines) for the ML, CV, REF1, and REF2 method in model M_2



Source: The author's own construction

Figure 6 Estimates of the ISE values (expressed in the log-scale) for the ML, CV, REF1, and REF2 methods and for the optimal bandwidths (OPT) in model M_2



Source: The author's own construction

CONCLUSION

The presented paper generalizes the Priestley-Chao estimator from the restrictive fixed design to the uniformly distributed random design variable *X*. The statistical properties of this estimator are derived, and the methods for bandwidth selection are suggested.

The leave-one-out maximum likelihood method, a modification of the classical likelihood approach, is proposed for bandwidth detection. This method is complemented by the cross-validation method and the reference rule method. The original approach of the reference rule method was extended to a normally distributed conditional variable with quadratic mean and constant standard deviation.

The performance of the suggested methods is presented using two simulation studies focusing on the bandwidth estimates and the quality measure estimates. The results show that the cross-validation method tends to undersmooth significantly. The reference rule method assuming the quadratic conditional mean produces results similar to or better than the reference rule with a linear conditional mean and linear standard deviation, but none of these two methods outperforms the ML method. On the other hand, the results of these two references are better than those of the CV method, even in the cases of the underlying conditional density not resembling the conditional density assumed by the reference model.

The ML method can adapt well not only to the changing shape of the conditional mean and conditional normal distribution but also to a bimodal or an asymmetric distribution. The method always results in an ISE that is smaller than the optimal bandwidth choice. It also detects the bandwidths which decrease ISE estimates, but does not underestimate the parameter h_x as the optimal case usually does. The simulation study shows that the proposed maximum likelihood method is a reasonable tool for bandwidth selection.

ACKNOWLEDGMENT

The research was supported by the project of specific university research at Brno University of Technology FAST-S-18-5184.

References

ARORA, S. AND TAYLOR, J. W. Forecasting electricity smart meter data using conditional kernel density estimation. *Omega*, 2016, 59 (Part A), pp. 47–59.

BASHTANNYK, D. M. AND HYNDMAN R. J. Bandwidth selection for kernel conditional density estimation. *Computational Statistics & Data Analysis*, 2001, 36(3), pp. 279–298.

FAN, J., YAO Q., TONG H. Estimation of conditional densities and sensitivity measures in nonlinear dynamical systems. *Biometrika*, 1996, 83(1), pp. 189–206.

FAN, J. AND YIM T. H. A crossvalidation method for estimating conditional densities. *Biometrika*, 2004, 91(4), pp. 819–834.

HALL, P., RACINE, J., LI, Q. Cross-validation and the estimation of conditional probability densities. *Journal of the American Statistical Association*, 2004, 99(468), pp. 1015–1026.

HANSEN, B. E. *Nonparametric conditional density estimation*. Unpublished manuscript, 2004.

HOLMES, M. P., GRAY, A. G., ISBELL, C. L. Fast Nonparametric Conditional Density Estimation. *ArXiv e-prints*, 2012.

HOROVÁ, I., KOLÁČEK, J., ZELINKA, J. *Kernel Smoothing in MATLAB: Theory and Practice of Kernel Smoothing*. Singapore: World Scientific Publishing Co. Pte. Ltd., 2012.

HOROVÁ, I. AND ZELINKA, J. Contribution to the bandwidth choice for kernel density estimates. *Computational Statistics*, 2007, 22(1), pp. 31–47.

JEON, J. AND TAYLOR, J. W. Using conditional kernel density estimation for wind power density forecasting. *Journal of the American Statistical Association*, 2012, 107(497), pp. 66–79.

KOLÁČEK, J. AND HOROVÁ, I. Iterative bandwidth method for kernel regression. *Journal of Statistics: Advances in Theory and Applications*, 2012, 8, pp. 93–101.

KONEČNÁ, K. Priestley-Chao estimator of conditional density. *Mathematics, Information Technologies and Applied Sciences 2017, post-conference proceedings of extended versions of selected papers*, 2017, pp. 151–163.

KONEČNÁ, K. The leave-one-out maximum likelihood method for the Priestley-Chao estimator of conditional density. *Proceedings, 17th Conference on Applied Mathematics – APLIMAT 2018*, 2018, pp. 577–589.

KONEČNÁ, K., HOROVÁ, I., KOLÁČEK, J. Conditional Density Estimations. In: SKIADAS, C. H. *Theoretical and Applied Issues in Statistics and Demography*, Athens: International Society for the Advancement of Science and Technology (ISAST), 2014, pp. 15–31.
LEIVA-MURILLO, J. M. AND ARTES-RODRIGUEZ, A. Algorithms for maximum-likelihood bandwidth selection in kernel density estimators. *Pattern Recognition Letters*, 2012, 33(13), pp. 1717–1724.
PRIESTLEY, M. B. AND CHAO, M. T. Non-parametric function fitting. *Journal of the Royal Statistical Society. Series B (Methodological)*, 1972, 34(3), pp. 385–392.
ROSENBLATT, M. Conditional probability density and regression estimators. *Multivariate analysis II*, 1969, 25, pp. 25–31.
TAKEUCHI, I., NOMURA, K., KANAMORI, T. Nonparametric conditional density estimation using piecewise-linear solution path of kernel quantile regression. *Neural Computation*, 2009, 21(2), pp. 533–559.
WAND, M. P. AND JONES M. C. *Kernel smoothing.* 1st Ed. London: Crc Press, 1994.

APPENDIX

Here, a detailed proof of **Theorem 1** can be found.
Proof. All the computations are based on Taylor expansions with higher-order terms ignored. First, the expectation (E) of the Formula (3) is derived:

$$\mathrm{E}\{\hat{f}_{PC}(y|x;h_x,h_y)\} = n\frac{1}{n}\mathrm{E}\left\{\mathrm{K}_{\mathrm{h_x}}(\mathrm{x}-\mathrm{X_i})\mathrm{K}_{\mathrm{h_y}}(\mathrm{y}-\mathrm{Y_i})\right\}$$
$$= \iint K_{h_x}(x-u)K_{h_y}(y-v)f(v|u)g(u)\,du\,dv$$
$$= f(y|x) + \frac{1}{2}h_x^2\beta_2(K)\frac{\partial^2 f(y|x)}{\partial x^2} + \frac{1}{2}h_y^2\beta_2(K)\frac{\partial^2 f(y|x)}{\partial y^2} + O(h_x^4) + O(h_y^4) + O(h_x^2h_y^2).$$

Then, the asymptotic bias is given as:

$$\mathrm{AB}\{\hat{f}_{PC}(y|x;h_x,h_y)\} = \frac{1}{2}h_x^2\beta_2(K)\frac{\partial^2 f(y|x)}{\partial x^2} + \frac{1}{2}h_y^2\beta_2(K)\frac{\partial^2 f(y|x)}{\partial y^2}.$$

The variance of the estimator is derived by the well-known law of total variance. Let X and Y be random variables. Then, the following equality holds:

$$\mathrm{var}\{Y\} = \mathrm{E}\left\{\mathrm{var}_{Y|X}\{Y|X\}\right\} + \mathrm{var}\left\{\mathrm{E}_{Y|X}\{Y|X\}\right\}. \tag{14}$$

First, by (14), the variance of the i-th term of the Formula (3) is derived.
The conditional expectation of the estimator's i-th term $\hat{f}_{PC}(y|x;h_x,h_y)$ can be written as:

$$\begin{aligned} &E_{f(y|X_i)}\left\{\frac{1}{n}K_{h_x}(x-X_i)K_{h_y}(y-Y_i)|X_i\right\} \\ &= \frac{1}{n}K_{h_x}(x-X_i)\left(f(y|X_i) + \frac{1}{2}h_y^2\beta_2(K)\frac{\partial^2 f(y|X_i)}{\partial y^2} + O(h_y^4)\right). \end{aligned} \tag{15}$$

The conditional expectation of the squared i-th term in (3) is given by:

$$\begin{aligned} &\mathrm{E}_{f(y|X_i)}\left\{\frac{1}{n^2}K_{h_x}^2(x-X_i)K_{h_y}^2(y-Y_i)|X_i\right\} \\ &= \frac{1}{n^2h_y}K_{h_x}^2(x-X_i)\left(R(K)f(y|X_i) + \frac{1}{2}h_y^2G(K)\frac{\partial^2 f(y|X_i)}{\partial y^2} + O(h_y^4)\right), \end{aligned} \tag{16}$$

where $\mathrm{G(K)} = \int u^2 K^2(u)du$. By subtracting the second power of (15) from (16), we have:

$$\operatorname{var}_{f(y|X_i)}\left\{\frac{1}{n}K_{h_x}(x-X_i)K_{h_y}(y-Y_i)|X_i\right\}$$
$$=\frac{1}{n^2}K_{h_x}^2(x-X_i)\left(\frac{1}{h_y}R(K)f(y|X_i)-f^2(y|X_i)+\frac{1}{2}h_yG(K)\frac{\partial^2 f(y|X_i)}{\partial y^2}+O\left(h_y^2\right)\right). \tag{17}$$

Finally, by applying the expectation to (17), the first term of (14) is acquired:

$$\mathrm{E}\left\{\operatorname{var}_{f(y|X_i)}\left\{\frac{1}{n}K_{h_x}(x-X_i)K_{h_y}(y-Y_i)\right\}\right\}$$
$$=\frac{1}{n^2h_xh_y}R^2(K)f(y|x)-\frac{1}{n^2h_x}R(K)f^2(y|x)+\frac{h_y}{2n^2h_x}R(K)G(K)\frac{\partial^2 f(y|x)}{\partial y^2}+O\left(h_y^2\right). \tag{18}$$

The expected value of (16) and its square are used to derive the variance of the conditional Formula (16).

$$\mathrm{E}\left\{\mathrm{E}_{f(y|X_i)}\left\{\frac{1}{n}K_{h_x}(x-X_i)K_{h_y}(y-Y_i)\right\}\right\}$$
$$=\frac{1}{n}f(y|x)+\frac{h_x^2}{2n}\beta_2(K)\frac{\partial^2 f(y|x)}{\partial x^2}+\frac{h_y^2}{2n}\beta_2(K)\frac{\partial^2 f(y|x)}{\partial y^2}+O\left(h_y^4\right), \tag{19}$$

$$\mathrm{E}\left\{\mathrm{E}_{f(y|X_i)}^2\left\{\frac{1}{n}K_{h_x}(x-X_i)K_{h_y}(y-Y_i)\right\}\right\}=\frac{1}{n^2h_x}R(K)f^2(y|x)+O\left(\frac{h_x}{n^2}\right)+O\left(\frac{h_y}{n^2}\right). \tag{20}$$

Thus, the variance of (16) is obtained by subtracting (19) squared from (20):

$$\operatorname{var}\left\{\mathrm{E}_{f(y|X_i)}\left\{\frac{1}{n}K_{h_x}(x-X_i)K_{h_y}(y-Y_i)\right\}\right\}$$
$$=\frac{1}{n^2h_x}R(K)f^2(y|x)-\frac{1}{n^2}f^2(y|x)+O\left(\frac{h_x}{n^2}\right)+O\left(\frac{h_y^2}{n^2}\right). \tag{21}$$

As the expression (21) is the desired second term of (14), the variance of the i-th term of the Priestley-Chao estimator is obtained by summing up (18) and (21):

$$\operatorname{var}\left\{\frac{1}{n}K_{h_x}(x-X_i)K_{h_y}(y-Y_i)\right\}=\frac{1}{n^2h_xh_y}R^2(K)f(y|x)+O\left(\frac{1}{n^2}\right)+O\left(\frac{h_x}{n^2}\right)+O\left(\frac{h_y}{n^2}\right).$$

It can be easily shown that the equation:

$$\operatorname{cov}\left\{\frac{1}{n}K_{h_x}(x-X_1)K_{h_y}(y-Y_1),\frac{1}{n}K_{h_x}(x-X_2)K_{h_y}(y-Y_2)\right\}=0$$

holds, i.e., the two terms of the Priestley-Chao estimator are not correlated. Then, the variance of the Formula (3) is given by:

$$\operatorname{var}\{\hat{f}_{PC}(y|x;h_x,h_y)\}=\frac{1}{nh_xh_y}R^2(K)f(y|x)+O\left(\frac{1}{n}\right)+O\left(\frac{h_x}{n}\right)+O\left(\frac{h_y}{n}\right).$$

By taking into account only the leading terms of bias and variance, the theorem is proven.

Bayesian Geographical Profiling in Terrorism Revealing

Jana Svobodová [1] | *Masaryk University, Brno, Czech Republic*
Jan Koláček [2] | *Masaryk University, Brno, Czech Republic*

Abstract

A significant part of research in terrorism studies focuses on the analysis of terrorist groups. An important issue for this type of research is that a large number of attacks are not attributed to a specific group. As an appropriate approach to solve the problem of attributing group responsibility we applied the geographic profiling theory. We analyzed several terrorist organizations which typically commit attacks far away from their headquarters. We proposed an innovative method based on Bayesian approach to find the organization's base and to attribute responsibility to perpetrators of terrorist attacks. We compared the results with classical techniques used in criminology. The real data analysis shows rationale for the proposed approach. Analyzed data comes from the Global Terrorism Database which is currently the most extensive database on terrorism ever collected.[3]

Keywords	***JEL code***
Bayesian data analysis, geographic profiling, Global Terrorism Database, anchor point	*C11, C14*

INTRODUCTION

Terrorism is usually understood as the use or threat of violence to further a political cause. Since acts of terrorism across the globe have increased notably in recent decades, this area plays an important role in sociological and political science research. A significant part of research in terrorism studies focuses on the analysis of terrorist groups. Studies of this type explore group attributes, e.g., ideology, size, and state sponsorship, in order to determine their impact on phenomenon such as the number of attacks conducted, their location or the targets of attacks (Asal and Rethemeyer, 2008; Carter, 2012). An important issue for this type of research is, that a large number of attacks are not attributed to a specific group (Arva and Beieler, 2014). Although many terrorist organizations actively seek publicity for their attacks, it is sometimes difficult to attribute responsibility to perpetrators of terrorist attacks. A comprehensive empirical overview of these uncertainties is given in Lafree et al. (2014) in context with the Global Terrorism Database (National Consortium for the Study of Terrorism and Responses to Terrorism (START), 2017) which is currently the most extensive database on terrorism ever collected.

As an appropriate approach to solving the problem of attributing group responsibility could be applying the geographic profiling theory. Geographic profiling is extensively used for finding criminals

[1] Department of Mathematics and Statistics, Faculty of Science, Masaryk University, Kotlářská 2, 611 37 Brno, Czech Republic. Email: janyfa@mail.muni.cz.

[2] Department of Mathematics and Statistics, Faculty of Science, Masaryk University, Kotlářská 2, 611 37 Brno, Czech Republic. Email: kolacek@math.muni.cz.

[3] This article is based on contribution at the conference *Robust 2018*.

such as thieves, robbers, burglars, rapists or sexual assailants. According to (Canter and Youngs, 2008) these types of perpetrators usually commit offenses at a distance between 0.89 and 3.87 kilometer from their base. This is the reason why the methods of geographic profiling are especially focused on offenders committing crimes near their anchor points.

However, we can find a number of criminal groups in various countries where a large percentage of offenders commute long distances to perpetrate a crime. These include American rapists, Canadian sexual assailants, Finnish thieves or Australian robbers where about 50% crimes are committed up to tens of kilometers away from the perpetrators' anchor point (Lundrigan and Cantter, 2001). According to distance where the offender perpetrates his crime, we distinguish two types of criminals – residents and non-residents (Svobodová, 2018). Terrorists represent a very specific group and a majority of them is an example of the latter. For non-local organizations (non-residents), the usual methods and approaches of geographic profiling are not applicable or do not bring such satisfactory results as for residents.

Bayesian approach (O'Leary, 2009) offers a very strong and useful tool for finding an anchor point of all types of criminals. Applying prior knowledge and a suitable likelihood model, we obtain a posterior function that can be a powerful source of information about the anchor point of both residents and non-residents.

The terrorist attack often occurs hundreds to thousands of kilometers away from the headquarters. The attack sites of one terrorist organization are also more distant from each another than it is usual for other above-mentioned crimes. Thus, we take all considerations in the units of hundreds of kilometers.

1 METHODS

1.1 Global Terrorism Dataset

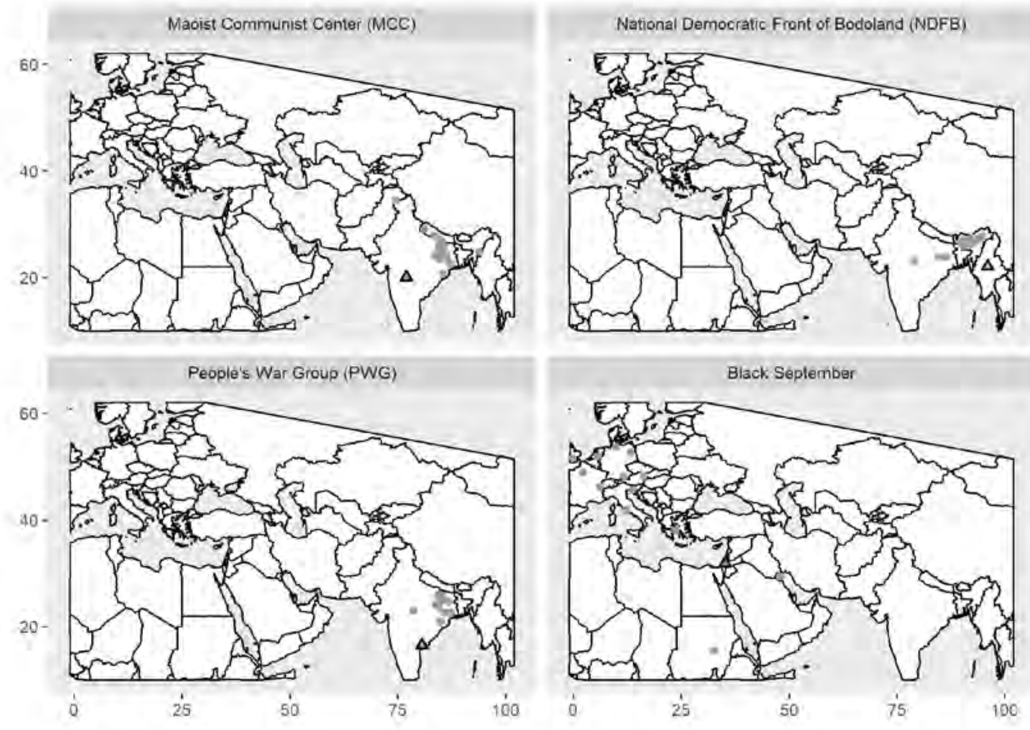
Data comes from the Global Terrorism Database (National Consortium for the Study of Terrorism and Responses to Terrorism (START), 2017). The Global Terrorism Database (GTD) is an open-source database containing the information on domestic and international terrorist attacks that have taken place around the world since 1970. It contains data on place, time and manner for more than 170 000 terrorist attacks. The database is updated annually adding new records of events from the previous calendar year. For each GTD incident, the information is available on the date and location of the incident, the weapons used and nature of the target, the number of casualties, and – when identifiable – the group or individual who is responsible. For our purposes, we turned our attention to the incident location and the perpetrator group name. The location details are specified by the longitude and latitude (based on WGS1984 standards) of the city in which the event took place. In order to ensure consistency in the usage of group names in the database, the GTD database uses a standardized list of group names that have been established by project staff to serve as a reference for all subsequent entries.

In the recent paper, we restricted ourselves to data with known perpetrator group name. We chose 10 well-known terrorist organizations from the region of the middle, south and south-east Asia. Unfortunately, in the GTD database, there is no information on the perpetrator group anchor point. Thus, we needed to study selected organizations from public sources, as Wikipedia or Country Reports on Terrorism 2016, and added their headquarters (anchor point) coordinates manually. The distribution of incidents for considered organizations and their anchor points are depicted in Figure 1.

1.2 Representation of data in UTM

In the original GTD dataset, the incident location was defined by the longitude and latitude based on World Geodetic System standard (WGS, 1984). For this coordinate system, we should consider the orthodromic distance between two points, i.e. the shortest distance between two points on the surface of a sphere. However, the proposed method is based on the Euclidian distance in a plane. Thus a data projection to the Cartesian coordinates was necessary. We used the projection to the Universal Transverse Mercator

Figure 1 The distribution of selected terrorist incidents (points) and their headquarters (triangles)



Source: Own construction

(UTM) coordinate system, which is a global system of grid-based mapping references. The position on Earth is given by the UTM zone number and the easting and northing planar coordinate pair in that zone. The point of origin for each UTM zone is the intersection of the equator and the zone's central meridian. The main idea of the UTM projection is projecting each of the 60 zones onto a plane separately instead of projecting the complete globe into a flat surface. This leads to a minimal scale distortion within each zone. On the other side, the UTM is not suitable for areas that span more than a few zones since distortion and error increase when moving farther from the zone for which the projection is defined.

After transforming the data to the UTM coordinates we chose several points with the biggest distance for each group and compared their Euclidian distances with their original orthodromic distances to supervise the maximal distortion. The considered incident locations are spread over quite a large area. It includes 18 UTM zones starting in zone 30 and ending in zone 47. As the reference zone for projecting we chose the central zone of this area, i.e. zone 38.

The group named "Kurdistan Workers' Party" (PKK) is an organization based in Qandil Mountains, i.e. its anchor point coordinates are 36°N, 44°E (UTM zone 38). Incident locations for PKK are spread through the largest area in comparison to the other terrorist groups, see Figure 1. Thus we chose PKK to demonstrate maximal scale distortion. The most distant PKK incident is located in London with coordinates 51.5°N, 0.12°W (UTM zone 30). The orthodromic distance between the anchor point and the incident location is 3 942 km. After UTM projection (with the reference zone 38), the distance is 4 087 km, i.e., the absolute difference is 145 km which results in 3.7% relative distance distortion. We can supervise the maximal distance distortion for other cases in a similar way and conclude that the relative distance distortion is less than 5% in general. This upper bound is acceptable for our further probability modeling.

1.3 Procedures and models

Methods of geographic profiling are based on the construction of the probability distribution that indicates which areas of the investigation region contain the anchor point with the highest probability. A lot of approaches apply a hit score function to find a prioritized search area.

However, although these procedures are very popular, they do not provide a probability distribution in the true sense. Moreover, there is no option to incorporate geographic features and other background information into the model. Geography of the region may have a great effect on the choice of the crime location (Brantingham and Brantingham, 1993; Canter et al., 2000; Rossmo, 2000). These two aspects are the main reason for criticism of the hit score function methods (Mohler and Short, 2012) and lead to search for other approaches.

Bayesian approach is a very useful and appropriate tool that meets all requirements for the methods of geographic profiling and offers possibility how to implement geography and other important features into the model (O'Leary, 2009). For our case of terrorism, let us denote by $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ the known sites of attacks of one terrorist organization where $\boldsymbol{x}_\mathrm{i} = (x_i^{(1)}, x_i^{(2)})$ for $i = 1,2, \dots, n$. We assume that the choice of the attack location is influenced by the headquarters $z = (z^{(1)}, z^{(2)})$ and by other k parameters $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_k)$. Then, we can describe the way how the investigated organization chooses a site of its attacks by a function $p(\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}|\mathbf{z}, \boldsymbol{\theta})$. In the terms of Bayesian method, this function is called likelihood function.

Using Bayes rule we obtain:

$$p(\mathbf{z}, \boldsymbol{\theta}\,|\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}) = \frac{p(\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}|\mathbf{z}, \boldsymbol{\theta}) \cdot p(\mathbf{z}, \boldsymbol{\theta})}{p(\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\})}, \tag{1}$$

where $p(\mathbf{z}, \boldsymbol{\theta}\,|\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\})$ denotes a posterior distribution, $p(\mathbf{z}, \boldsymbol{\theta})$ contains information that is available before data analyzing and, therefore, it is called a prior distribution and the denominator $p(\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\})$ is referred to as an evidence. This part of Bayes rule is very important for comparing different models, for our purpose, it plays a role of normalization constant. Therefore we can replace equality (=) by proportionality (∝) and the denominator can be omitted.

To find the probability distribution of the headquarters $\mathbf{z}$, we get rid of unnecessary parameters by integrating over all possible values of $\boldsymbol{\theta}$. When considering the independence between the headquarters $\mathbf{z}$ and the parameters $\boldsymbol{\theta}$, we can simply write:

$$p(\mathbf{z}\,|\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}) \propto \int \dots \int_{M_\theta} p(\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}|\mathbf{z}, \boldsymbol{\theta}) \cdot h(\mathbf{z}) \cdot g(\boldsymbol{\theta})\, \mathrm{d}\theta_1 \dots \mathrm{d}\theta_k, \tag{2}$$

where M_θ indicates the region of integration and functions h and g denote prior distributions for headquarters $\mathbf{z}$ and for parameters $\boldsymbol{\theta}$.

There are a lot of possibilities on how to construct a likelihood function $p(\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}|\mathbf{z}, \boldsymbol{\theta})$. In a large number of papers about the Bayesian approach to the geographic profiling (O'Leary, 2009; O'Leary, 2010), we can find the assumption of the independence between offender's crime sites, thus we could write:

$$p(\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}|\mathbf{z}, \boldsymbol{\theta}) = \prod_{i=1}^{n} p_0(\mathbf{x}_i|\mathbf{z}, \boldsymbol{\theta}).$$

When dealing with terrorist organizations, the independence of attack locations cannot be assumed. There is usually a link between a series of attacks. Therefore, there is a need to proceed in a different way.

We use all attack sites to estimate the most likely one that we denote by $\mathbf{x}_H$ and for it we apply the model $p_0(\mathbf{x}_H|\mathbf{z}, \boldsymbol{\theta})$. We can then write the relationship as:

$$p(\{\mathbf{x}_1, \mathbf{x}_2, \dots , \mathbf{x}_n\}|\mathbf{z}, \boldsymbol{\theta}) = p_0(\mathbf{x}_H|\mathbf{z}, \boldsymbol{\theta}).$$

Another issue is how to model $p_0(\mathbf{x}_H|\mathbf{z}, \boldsymbol{\theta})$. In this paper, we only deal with non-local terrorist groups. Some authors point out (Levine, 2009; Mohler and Short, 2012), that the choice of attack site is, for this type of commuting offenders, influenced not only by the distance between the headquarters and the attack location but also by the angle at which the attack is committed. In (Mohler and Short, 2012), the designed kinetic model with the suitable choice of parameters can be applied to offenders committing crimes at great distances. It has been proved that after some assumptions it can be approximated by the product of a function of the distance and a function of the angle. This result inspires us to solve the problem of commuting offenders by combination of two suitable functions – first, a function that affects the probability of distance in which perpetrator commits a crime and, second, a function that influences probability of the corresponding angle. Construction of this model is presented in (Svobodová, 2018) as follows:

$$p_0(\mathbf{x}_H|\mathbf{z}, \alpha, \vartheta, \sigma_1, \sigma_2) = \frac{1}{N(\alpha, \vartheta, \sigma_1, \sigma_2)} \cdot q_1(\mathbf{x}_H|\mathbf{z}, \alpha, \sigma_1) \cdot q_2(\mathbf{x}_H|\mathbf{z}, \vartheta, \sigma_2), \tag{3}$$

where:

$$q_1(\mathbf{x}_H|\mathbf{z}, \alpha, \sigma_1) = \exp\left(-\frac{1}{2\sigma_1^2}\left[\sqrt{(x_H^{(1)} - z^{(1)})^2 + (x_H^{(2)} - z^{(2)})^2} - \alpha\right]^2\right),$$

and:

$$q_2(\mathbf{x}_H|\mathbf{z}, \vartheta, \sigma_2) = \exp\left(-\frac{1}{2\sigma_2^2}\left[\operatorname{atan2}(x_H^{(2)} - z^{(2)}, x_H^{(1)} - z^{(1)}) - \vartheta\right]^2\right).$$

The median of distance for committing terrorist attacks is denoted by α, σ_1 is the standard deviation corresponding to the function q_1. The average angle from the headquarters to the attack site measured from the horizontal axis with the origin at the headquarters $\mathbf{z}$ is expressed by the ϑ (the function q_2 achieves the highest values at the angle ϑ) and q_2 corresponds to the standard deviation of the function q_2. The functional values of q_2 around the angle ϑ decrease at a rate that is influenced by q_2.

The denominator of (3) is a normalization factor that ensures that the likelihood function $p_0(\mathbf{x}_i|\mathbf{z}, \alpha, \vartheta, \sigma_1, \sigma_2)$ is a probability distribution. If ϕ represents the distribution function of the standard normal distribution, the normalization factor has the form:

$$N(\alpha, \vartheta, \sigma_1, \sigma_2) = N_1(\alpha, \sigma_1) \cdot N_2(\vartheta, \sigma_2),$$

where:

$$N_1(\alpha, \sigma_1) = \sigma_1^2 \cdot \exp\left(-\frac{\alpha^2}{2\sigma_1^2}\right) + \sqrt{2\pi}\alpha\sigma_1\left(1 - \phi\left(-\frac{\alpha}{\sigma_1}\right)\right),$$

and:

$$N_2(\alpha, \sigma_2) = \sigma_2\sqrt{2\pi}\alpha \cdot \exp\left(\phi\left(\frac{2\pi - \vartheta}{\sigma_2}\right) - \phi\left(-\frac{\vartheta}{\sigma_2}\right)\right).$$

We can see in the relationship (2), that in addition to the likelihood function, we need to determine the prior functions for the headquarters $\mathbf{z}$ and for all other parameters – in our case α and ϑ.

The parameters σ_1 and σ_2 are estimated by the sample standard deviation using the known data about other offenders.

The most popular method of the geographic profiling is Rossmo's approach (Rossmo, 2000). In this paper, we use it as benchmark to compare its efficiency with the efficiency of our method. Rossmo uses the hit score function:

$$S(\mathbf{y}) = \sum_{i=1}^{n} f(d(\mathbf{x}_i,\mathbf{y})),$$

where the distance decay function f has the following form:

$$d(\mathbf{x}_i,\mathbf{y})) = \begin{cases} \dfrac{k}{(d(\mathbf{x}_i,\mathbf{y}))^h} & \text{for } d(\mathbf{x}_i,\mathbf{y}) > b, \\ \dfrac{kb^{g-h}}{(2b - d(\mathbf{x}_i,\mathbf{y}))^g} & \text{for } d(\mathbf{x}_i,\mathbf{y}) \leq b. \end{cases} \tag{4}$$

The distance between the crime site $\mathbf{x}_i$ and any place $\mathbf{y}$ is determined by the Manhattan distance, the parameter b denotes the radius of the buffer zone and is set to one half of the average distance of the nearest neighbour between terrorist attacks of the examined terrorist organization. The exponents g and h are recommended by Rossmo to be 1.2.

2 RESULTS AND DISCUSSION

In this section, we present results obtained by using the proposed method on the dataset described above. For all calculations and graph creations we used the software R (R Core Team, 2013). The analysis was performed on selected 10 terrorist organizations.

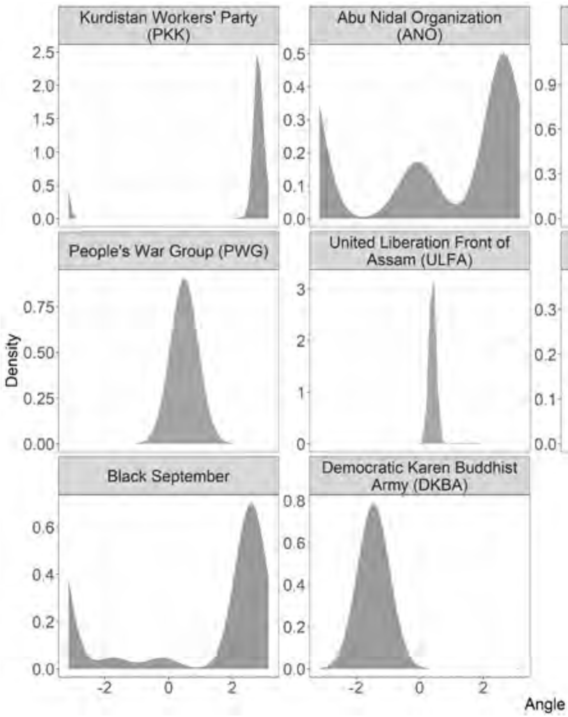
Our aim was to examine the accuracy of the estimate where the headquarters of the non-local terrorist organization is located. For each organization, we chose just the incidents with distance from its headquarters greater than a minimal value. Based on (Canter and Youngs, 2008), this value was set first to 4 units – in our case to 400 kilometers, then to 600, 800 and, finally, to 1 000 kilometers. Accuracy of the estimate was very similar in all cases. This study presents the results for attacks at a distance greater than 800 kilometers from the headquarters of the terrorist organizations.

Firstly, we inspected the angles at which the attacks were committed. The probability distribution of all angles is depicted in Figure 2. It is evident that for many organizations, the angle plays an important role in the incident location selection.

For construction of the prior functions for the parameters $\mathbf{z}$ and ϑ, we used kernel smoothing techniques, and for the parameter α, we applied the logspline density estimation. It allows to limit the range only to non-negative values. When estimating the prior functions, we used all available data. We always excluded only the information about the examined offender. Figure 3 shows the estimated prior functions for the average angle ϑ and the median distance α over all organizations.

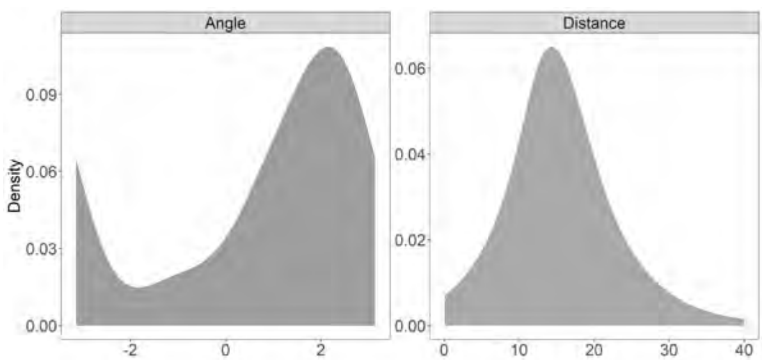
Further, we chose the investigated area to include all attacks and headquarters and increased it by 1 500 km approximately. This space was divided into a grid with the cell dimension of 100 km × 100 km . We evaluated our proposed posterior function (2) and the Rossmo's method (4) in all cells. For each approach, we ordered all cells based upon the score value, from the highest to the lowest, i.e. from the cell that includes the headquarters with the highest probability to the cell that includes it with the lowest probability. The efficiency of the method is given by the number of cells that we had to examine until we found the headquarters of the investigated terrorist organization. The proportion of this number to all cells points out how successful each method is, i.e. a lower proportion (percentage) indicates higher efficiency. The models performance for all organizations is given in Table 1, graphical representation in Figure 4.

Figure 2 The probability distribution of incident directions



Source: Own construction

Figure 3 Estimated prior distributions of angle and distance



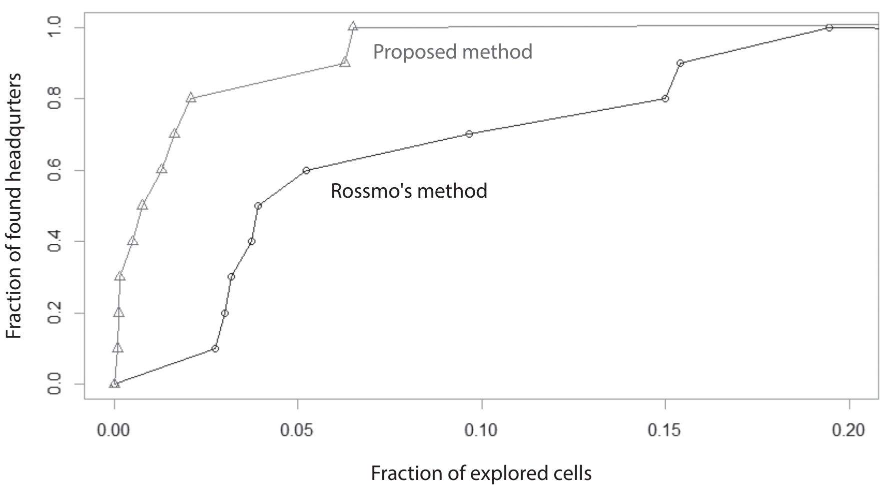
Source: Own construction

Table 1 Comparison of model performance (percentage of the number of cells that we had to examine until we found the headquarters of the investigated terrorist organization)

Organization	Rossmo's model	Proposed model
Kurdistan Workers' Party (PKK)	9.67	1.65
Abu Nidal Organization (ANO)	15.40	1.29
Maoist Communist Center (MCC)	5.24	6.50
National Democratic Front of Bodoland (NDFB)	3.00	0.09
People's War Group (PWG)	3.93	6.28
United Liberation Front of Assam (ULFA)	2.76	0.16
Mujahedin-e Khalq (MEK)	3.75	0.12
Armenian Secret Army for the Liberation of Armenia	14.99	0.51
Black September	19.45	0.76
Democratic Karen Buddhist Army (DKBA)	3.19	2.09

Source: Own construction

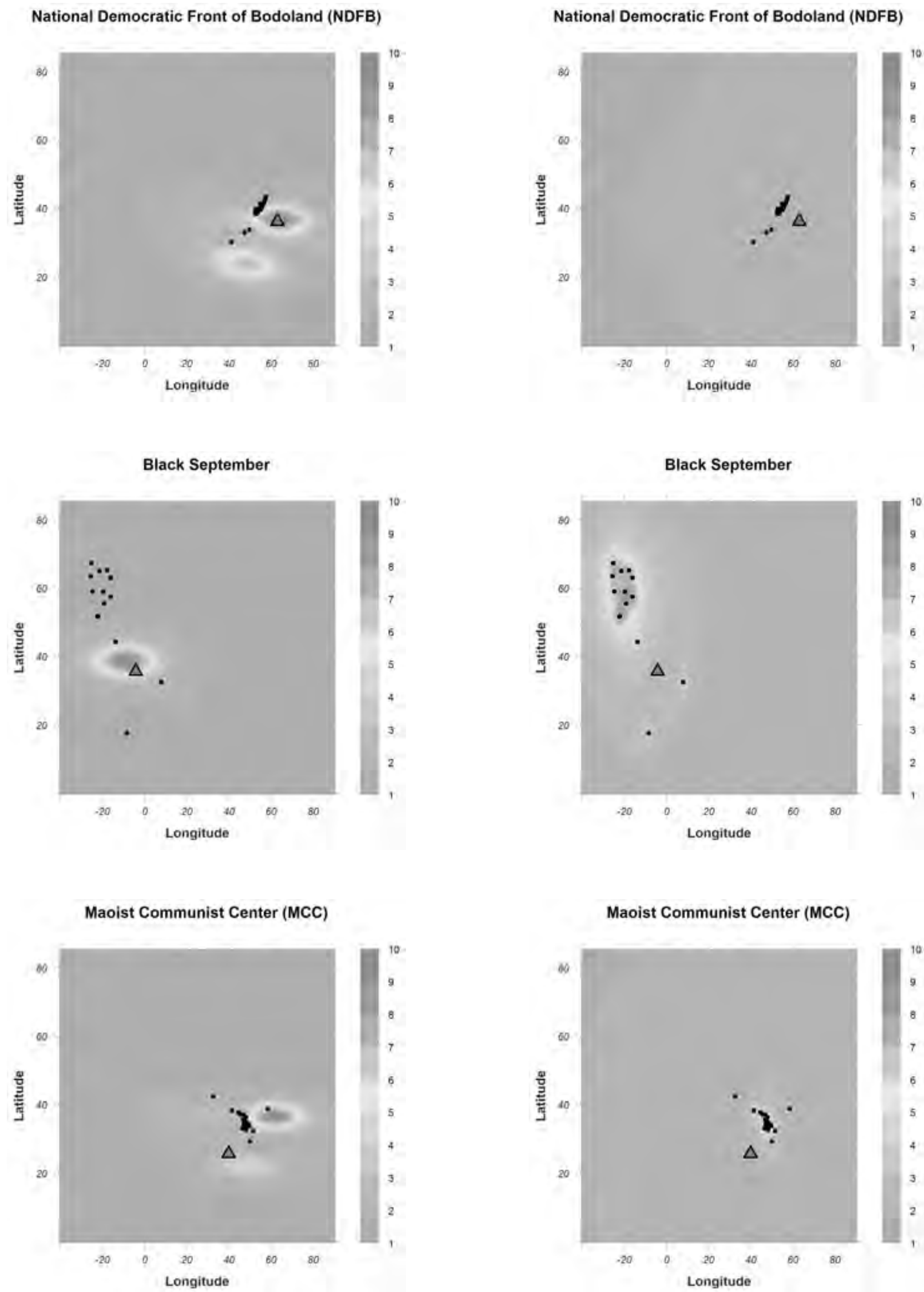
Figure 4 Graphical comparison of the model performance (triangles for the proposed method, circles for Rossmo's approach)



Source: Own construction

In all cases, the proposed method's rate was under 7%. Moreover, if two organizations with highest rate (MCC and PWG) are excluded, the rate was approximately 2% and less. This fact means, that the proposed model was efficient and the unknown headquarters was found quite quickly. It is not surprising that the Rossmo's model was not as efficient as the proposed model. It was suggested for residents, i.e.

Figure 5 Level plots for the proposed method (on the left) and Rossmo's approach (on the right) indicating how likely is that the area contains the offenders' headquarters (the black circles indicate attack sites, the triangle denotes real headquarters)



Source: Own construction

local perpetrators, and it cannot cover the behavior of another type of offenders. On the other side, in two mentioned cases (MCC and PWG), the efficiency of the proposed model was smaller than the efficiency of the Rossmo's model. The reason for it subsists in the incidents angle distribution. Figure 1 shows the two considered perpetrators preferred angles between 0 and $\pi/3$. However, the prior probability distribution (see Figure 3) estimates relatively small probabilities for these angles. In this sense, the prior angle estimate is not sufficient for these two organizations and thus the model results are biased.

In Figure 5, there are some examples of estimates of the terrorist headquarters. We can see that the hit score function with Rossmo's distance decay function assumes that the headquarters lies close to any of the attack sites. Our method admits the possibility that the headquarters is located at a greater distance from the attack sites. It is obvious from Table 1 that the proposed method is less accurate than Rossmo's approach for MCC. However, also in this case, the real headquarters lies very close to the second most probable region in the whole investigated area.

CONCLUSION

In previous works on geographic profiling, several types of offenders were analyzed to detect their anchor point. Perpetrators usually commit offenses at a shorter distance from their base. For these types of offenders the Rossmo's approach is the most popular and used in criminology.

In contrast to previous works, the analysis of offenders commuting long distances to perpetrate a crime seems to be helpful. We analyzed several terrorist organizations which typically commit attacks hundreds to thousands of kilometers away from their headquarters. We proposed an innovative method based on Bayesian approach to find the organization's base and to attribute responsibility to perpetrators of terrorist attacks. The real data analysis shows rationale for the proposed approach. The method is more flexible by covering the perpetrator's preference of incident's angles and distances from its headquarters. On the other side, the Bayesian approach is more sensitive to a quality of corresponding prior distributions estimates. It could cause slightly biased results in some cases. The complexity of the method brings a practical issue as it is more time consuming than the Rossmo's approach.

We see further challenges in the extension of the presented study in following ways. A more detailed criterion of "non-locality" of a perpetrator could be helpful. It would allow setting some weight parameters in construction of a more general model for any kind of perpetrator and it could lead to the development of an automated data-processing algorithm. The determination of a prior for perpetrator's travel direction with assumption of a type of dependency on its starting point could give more accurate results in modelling. Some offenders prefer several locations of attacks with specific angle and distance. Thus the assumption of dependency between them seems to be important in prior estimation.

ACKNOWLEDGEMENT

This research was supported by Masaryk University, project MUNI/A/1204/2017.

References

ARVA, B. AND BEIELER, J. Dealing with missing data in group-level studies of terrorism. In: *APSA Annual Meeting*, 2014, pp. 28–31.

ASAL, V. AND RETHEMEYER, R. K. The nature of the beast: Organizational structures and the lethality of terrorist attacks. *The Journal of Politics*, 2008, 70.2, pp. 437–449.

BRANTINGHAM, P. L. AND BRANTINGHAM, P. J. Nodes, paths and edges. *Journal of Environmental Psychology*, 1993, 13(1), pp. 3–28.

CANTER, D., COFFEY, T., HUNTLEY, M., MISSEN, C. Predicting serial killers' home base using a decision support system. *Journal of Quantitative Criminology*, 2000, 16(4), pp. 457–478.

CANTER, D. V. AND YOUNDS, D. *Applications of geographical offender profiling.* Burlington. VT: Ashage Pub, 2008.

CARTER, D. B. A blessing or a curse? State support for terrorist groups. *International Organization*, 2012, 66.1, pp. 129–151.

LAFREE, G., DUGAN, L., MILLER, E. *Putting terrorism in context: Lessons from the Global Terrorism Database.* Routledge, 2014.

LEVINE, N. Introduction to the Special Issue on Bayesian journey-to-crime modeling. *Journal of Investigative Psychology and Offender Profiling*, 2009, 6, pp. 162–167.

LUNDRIGAN, S. AND CANTER, D. A multivariate analysis of serial murderers' disposal site location choice. *Journal of Environmental Psychology*, 2001, 21(4), pp. 423–432.

MOHLER, G. O. AND SHORT, M. B. Geographic profiling from kinetic models of criminal behavior. *SIAM Journal on Applied Mathematics*, 2012, 72(1), pp. 163–180.

NATIONAL CONSORTIUM FOR THE STUDY OF TERRORISM AND RESPONSES TO TERRORISM (START). *Global Terrorism Database* [globalterrorismdb_0617dist.csv], 2007. <https://www.start.umd.edu/gtd>.

O'LEARY, M. The mathematics of geographic profiling. *Journal of Investigative Psychology and Offender Profiling*, 2009, 6(3), pp. 253–265.

O'LEARY, M. Implementing a bayesian approach to criminal geographic profiling. In: *Proceedings of the 1st International Conference and Exhibition on Computing for Geospatial Research & Application*, 2010, New York, NY, USA, pp. 59:1–59:8.

R CORE TEAM. R: A *Language and Environment for Statistical Computing* [online]. R Foundation for Statistical Computing, Vienna, Austria, 2013. <http://www.R-project.org/>.

ROSSMO, D. *Geographic profiling.* CRC Press, Boca Raton, Fla, 2000.

SVOBODOVÁ, J. *Bayesian models in geographic profiling.* ArXiv e-prints, 2018.

UNITED STATES DEPARTMENT OF STATE. *Country Reports on Terrorism 2016* [online]. 2017. <https://www.state.gov/j/ct/rls/crt/2016>.

WGS – World Geodetic System website of the NGA [online]. National Geospatial-Intelligence Agency, 1984. [archived from the original on 2.4.2012].

Recent Publications and Events

New publications of the Czech Statistical Office

Household Income and Living Conditions in 2017. Prague: CZSO, 2018.
Indicators of Social and Economic Development of the Czech Republic 2000–1st quarter 2018. Prague: CZSO, 2018.
Information Society in Figures 2018. Prague: CZSO, 2018.
Inovační aktivity podniků v letech 2014–2016. Prague: CZSO, 2018.
Population of Municipalities 1st January 2018. Prague: CZSO, 2018.

Other selected publications

Conference of European Statisticians. Road for Sustainable Development Goals. New York and Geneva: United Nations, 2017.
The Atlas of Religions in Czechia. Prague: Karolinum, 2017.

Conferences

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

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

The ***12th International Days of Statistics and Economics (MSED 2018 Conference)*** took place in the **University of Economics, Prague, Czech Republic, from 6th to 8th September 2018**. The conference is jointly organized by the Department of Statistics and Probability and the Department of Microeconomics, University of Economics, Prague, Czech Republic; Faculty of Economics, the Technical University of Košice, Slovakia; and the Ton Duc Thang University, Vietnam. The aim of the conference is to present and discuss current problems of statistics, demography, economics and management and their mutual interconnection. More information available at: <*https://msed.vse.cz*>.

The ***36th International Conference on Mathematical Methods in Economics (MME 2018)*** was held in **Jindřichův Hradec, Czech Republic, during 12–14 September 2018**. The conference is a traditional meeting of professionals from universities and businesses interested in the theory and applications of operations research and econometrics. More information available at: <*http://mme2018.fm.vse.cz*>.

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

The journal of *Statistika* has the following sections:

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

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

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

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

In the *Information* section we publish informative (descriptive) texts. The maximum range of information is 6 000 words or up to 10 1.5-spaced pages.

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

Recommended Paper Structure
Title (e.g. On Laconic and Informative Titles) — Authors and Contacts — Abstract (max. 160 words) — Keywords (max. 6 words / phrases) — JEL classification code — Introduction — … — Conclusion — Annex — Acknowledgments — References — Tables and Figures

Authors and Contacts
Rudolf Novak*, Institution Name, Street, City, Country
Jonathan Davis, Institution Name, Street, City, Country
* Corresponding author: e-mail: rudolf.novak@domainname.cz, phone: (+420) 111 222 333

Main Text Format
Times 12 (main text), 1.5 spacing between lines. Page numbers in the lower right-hand corner. *Italics* can be used in the text if necessary. *Do not* use **bold** or <u>underline</u> in the text. Paper parts numbering: 1, 1.1, 1.2, etc.

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

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

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

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

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

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

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

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

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

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

Subscription price (4 issues yearly)
CZK 372 (incl. postage) for the Czech Republic,
EUR 117 or USD 174 (incl. postage) for other countries.
Printed copies can be bought at the Publications Shop of the Czech Statistical Office (CZK 66 per copy).
address: Na padesátém 81 | 100 82 Prague 10 | Czech Republic

Subscriptions and orders
MYRIS TRADE, s. r. o.
P. O. BOX 2 | 142 01 Prague 4 | Czech Republic
phone: (+420) 234 035 200,
fax: (+420) 234 035 207
e-mail: myris@myris.cz

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

98th year of the series of professional statistics and economy journals of the State Statistical Service in the Czech Republic: ***Statistika*** (since 1964), ***Statistika a kontrola*** (1962–1963), ***Statistický obzor*** (1931–1961) and ***Československý statistický věstník*** (1920–1930).

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