COMENIUS UNIVERSITY IN BRATISLAVA
FACULTY OF MATHEMATICS, PHYSICS AND INFORMATICS

IDENTIFICATION OF POVERTY BY DEA

DIPLOMA THESIS

2015 **Kristína JABLONICKÁ**

COMENIUS UNIVERSITY IN BRATISLAVA
FACULTY OF MATHEMATICS, PHYSICS AND INFORMATICS

IDENTIFICATION OF POVERTY BY DEA

DIPLOMA THESIS

Study programme:	Economic and Financial Mathematics
Field of Study:	1114 Applied Mathematics
Department:	FMFI.KAMŠ -Department of Applied Mathematics and Statistics
Supervisor:	prof. RNDr. Pavel Brunovský, DrSc.

Bratislava 2015 **Kristína JABLONICKÁ**

UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY

IDENTIFIKÁCIA CHUDOBY POMOCOU METÓDY DEA

DIPLOMOVÁ PRÁCA

Študijný program:	Ekonomická a finančná matematika
Študijný odbor:	9.1.9. Aplikovaná matematika
Školiace pracovisko:	Katedra aplikovanej matematiky a štatistiky
Vedúci práce:	prof. RNDr. Pavel Brunovský, DrSc.

Bratislava 2015 **Bc. Kristína JABLONICKÁ**

74532867

Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta: Bc. Kristína Jablonická
Študijný program: ekonomická a finančná matematika (Jednoodborové štúdium, magisterský II. st., denná forma)
Študijný odbor: 9.1.9. aplikovaná matematika
Typ záverečnej práce: diplomová
Jazyk záverečnej práce: slovenský

Názov: Identifikácia chudoby pomocou metódy DEA / *Identification of poverty by DEA.*

Cieľ: Chudoba je vlastnosť, ktorá nie je dosť dobre charakterizovateľná hodnotou jednej číselnej premennej. Preto je prirodzené siahnuť po metóde DEA ako možnom nástroji hodnotenia viacerých číselných parametrov. Cieľom práce je preskúmať možnosti formulácie úlohy identifikácie chudoby pomocou metódy DEA a otestovanie prístupu na dostupných dátach SR.

Vedúci: prof. RNDr. Pavel Brunovský, DrSc.
Katedra: FMFI.KAMŠ - Katedra aplikovanej matematiky a štatistiky
Vedúci katedry: prof. RNDr. Daniel Ševčovič, CSc.

Dátum zadania: 29.01.2014

Dátum schválenia: 10.02.2014

prof. RNDr. Daniel Ševčovič, CSc.
garant študijného programu

....

študent vedúci práce

Acknowledgements I would like to thank my supervisor, Professor Pavel Brunovský for numerous consultations and support.

Abstract

JABLONICKÁ, Kristína: Identification of Poverty by DEA [Diploma Thesis], Comenius University in Bratislava, Faculty of Mathematics, Physics and Informatics, Department of Applied Mathematics and Statistics; Supervisor: prof. RNDr. Pavel Brunovský, DrSc., Bratislava, 2015, 57p.

We examine what DEA can contribute to identification of the poorest districts in a relative sense. If we consider poverty multidimensionally, it is not trivial to say which districts are the poorest because there are two or more indicators to consider. Rather than assigning arbitrary weights, DEA uses linear programming to find the envelopment of the dataset - a poverty frontier. For this examination we used a combination of publicly available data from 2009 and data an extensive survey of entrepreneurs from all 79 districts of Slovakia from the study of Hajko et al. (2011). We selected 9 indicators which reflected international concepts of poverty. We compare DEA with an arithmetic mean of indicators for several justifiable choices of indicators ranging from only 2 of them to all 9. We also calculate the Human Development Index using the life expectancy, average monthly wage and level of education and compare the results with DEA approach to the same parameters. Based on our results we discuss the pros and contras of DEA compared to alternative methods.

Keywords: Data Envelopment Analysis, Poverty Identification, Additive Model, Human Development Index,

Abstrakt

JABLONICKÁ, Kristína: Identifikácia chudoby pomocou metódy DEA [Diplomová práca], Univerzita Komenského v Bratislave, Fakulta matematiky, fyziky a informatiky, Katedra aplikovanej matematiky a štatistiky; školiteľ: prof. RNDr. Pavel Brunovský, DrSc., Bratislava, 2015, 57 s.

Skúmame, čo môže poskytnúť použitie DEA pri identifikácii najchudobnejších okresov v relatívnom ponímaní. Pri mnohorozmernom koncepte chudoby nie je jednoduché určiť, ktoré okresy sú najchudobnejšie, pretože máme k dispozícii viac ako dva indikátory. DEA je neparametrická metóda, ktorá namiesto toho, aby požadovala ako vstup váhy jednotlivých indikátorov, používa lineárne programovanie, aby našla obálku dát - v tomto prípade neefektívnu hranicu. Používame dáta zo štúdie Hajko, Klatik, Tunega (2010). Táto štúdia obsahuje verejne dostupné dáta z roku 2009 a dáta z prieskumu podnikateľov zo všetkých 79 okresov SR. Vybrali sme deväť indikátorov, ktoré odrážajú medzinárodný koncept chudoby. Porovnávame aditívny model s aritmetickým priemerom indikátorov pre niekoľko kombinácii indikátorov od dvoch po všetkých deväť. Taktiež počítame Human Development Index s použitím očakávanej dĺžky života, priemernej mesačnej mzdy a úrovne vzdelania. Výsledky HDI porovnávame s DEA aplikovanou na tie isté parametre. Na základe našich výsledkov sa zaoberáme výhodami a nevýhodami DEA v porovnaní s alternatívnymi metódami.

Kľúčové slová: Data Envelopment Analysis, identifikácia chudoby, aditívny model, Human Development Index

Contents

List of Figures

List of Tables

List of Abbreviations

DEA Data Envelopment Analysis

DMU Decision Making Unit

PPP Purchasing Power Parity

EI Education Index

EYSI Expected Years of Schooling Index

GDP Gross Domestic Product

GNI Gross National Income

HDI Human Development Index

II Income Index

LEI Life Expectancy Index

MYSI Main Years of Schooling Index

PCA Principal Component Analysis

UNDP United Nations Development Program

Introduction

Identification of poverty is a basic step towards its further study as well as policies that deal with it. A difficulty lies in the multidimensionality of poverty. It can hardly be reduced to a single numerical indicator. The multidimensionality of poverty is a well established fact recognized even by large international organizations like United Nations Development Program (UNDP) or the World Bank. A typical approach to deal with it is to construct a weighted index of the indicators. This usually means a decision about the weights in the composite indicator. It is difficult to justify a given decision on weights because we combine values in different units. For example, how should one combine such diverse indicators as income, life expectancy and education attainment in one meaningful number? The weights attached to different dimensions of poverty represent choices which are more or less arbitrary. Moreover, the importance of different dimensions may vary across the population and any decision on weights may be biased towards some of the units.

Second arbitrary choice which comes into the process is the choice of how many poorest districts to consider. Where should we draw a line between the poorest and moderately poor districts? For example, the World bank uses a poverty line of an equivalent of $1 Purchasing Power Parity (PPP) a day as a line under which live people in extreme poverty. However, this line has been criticized because it is considered difficult to live even on $5 PPP a day.

DEA is an alternative approach developed to deal with multidimensionality. It can be viewed as a way to choose individual weights most suitable for particular units. Alternatively, it singles out units having Pareto property. In this thesis we call such districts extremal (not to be confused with extreme in the common sense). We implement DEA in R and apply it to the data of the districts of Slovakia. This is compared to the same number of poorest districts according to their rank in a weighted average of poverty indicators. We also construct an index of the districts according to the Human Development Index(HDI) methodology and compare it with the DEA approach to the same indicators.

The thesis is organized as follows: In Chapter 1 we introduce the topic of multidimensional poverty and discuss the problem of identifying the poorest districts. We include one of the popular multidimensional indicators of poverty the HDI. Then we introduce some desirable poverty indicator properties, based on desirable properties of poverty measures in Havinga et al. (2005). We note some specifics of the available data for the study of poverty and review previous papers dealing with multidimensional poverty in Slovakia.

In Chapter 2 we observe the DEA additive model.

In Chapter 3 we discuss specific features of application of DEA to poverty.

In Chapter 4 we prepare the indicators from Hajko et al. (2011). We normalize their values to the scale from 0 to 1. We examine their corellation analysis and comment on the implications for use of DEA. We also do their PCA.

In Chapter 5 we present the results of the application. For each selected combination of indicators we compare the set of poorest districts obtained from the weighted index method and DEA. We compare both approaches on the map of Slovakia. We also compare results of HDI with DEA. In the Conclusion, we summarize the lessons learned from the application of DEA on the Slovak data.

1 Multidimensional poverty

In 1835 Tocqueville wrote Memoir on Pauperism (cited in Kakwani and Silber (2008)) in which he showed the difference between the poor and the indigents. In some countries the poor were "ill-fed and ill-clothed, living in the midst of a half-uncultivated countryside and in miserable dwellings" while the indigent didn't look impoverished but were living off of social benefits. This was an early account of the multidimensional nature of poverty.

In this chapter we introduce the problem of identification of the multidimensionally poorest districts, then we present some of the approaches to its solution found in the literature on poverty measurement. We introduce HDI which we will use in application later in Chapter 3. We introduce desirable properties of a poverty indicator adapted from desirable properties of poverty measure found in Havinga et al. (2005).

1.1 Identification of the multidimensionally poorest districts

Suppose that poverty can be characterized by $m \in \mathbb{N}$ indicators, where each indicator y_j can take values from 0 to 1 and a higher value of each indicator means that the district is poorer in that indicator. Let us choose a country with $n \in \mathbb{N}$ districts. We can then define a poverty matrix $Y = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1j} & \cdots & y_{1m} \\ \vdots & & & & & \\ y_{i1} & y_{i2} & \cdots & y_{ij} & \cdots & y_{im} \\ \vdots & & & & & \\ y_{n1} & x_{n2} & \cdots & y_{nj} & \cdots & y_{nm} \end{bmatrix}$ as a $n \times m$-matrix of these m indicators of poverty measured for n districts.

Our goal is to single out a set P of multidimensionally poorest districts among them using information of this poverty matrix Y. Note that most papers on multidimensional poverty solve a related but different objective of creating a complete ordering of units. There could be several different approaches to our problem.

1.2 Approach 1: Weighted index

One would be to construct an index $I_i = \sum_{j=1}^{m} w_j y_{ij}$, so that $\sum_{j=1}^{m} w_j = 1$. [1]
We choose a positive integer $r \in \mathbb{N}$. P is then defined as the set of those i with the r highest values of I_i.
Alternatively, we could choose a threshold $z \in \mathbb{R}$ and define $P = \{i : I_i > z\}$.
In the past, the weighted average has been used widely in creation of composite indices. The work Yang (2014)lists more than a 100 composite development indices. One of its weaknesses is that any given choice of weights is a matter of choice. Also, one set of weights is imposed on all the districts even though the relative importance of indicators may vary across the country. Another problem is that we have to choose r or z. In practice, the analyses are sometimes repeated with different choices for r or z to test the robustness of the hypoteses.

1.3 Approach 2: Rawlsian Maximin

Another approach would be $P = \{i : \exists j : y_{ij} = \max_j Y_{ij}\}$ which is called Rawlsian Maximin approach. This is based on a premise that the poorest dimension is the most important to identify. All the dimensions are so important for wellbeing that being the worst in any of them indicates relative poverty. There seem to be no published application of the Rawlsian Maximin approach in the multidimensional context. The number of poorest districts would be dependent on the number of extremes in all the individual indicators.

1.4 Income-based measures

First measures of poverty were based on the idea that high enough income means that the person can provide for his other needs, such as health, nutrition or education. However, analyses such as Laderchi et al. (2003) argue that income alone (whether as Gross Domestic Product (GDP) or Gross National Income (GNI)) does not suffice to characterize the whole reality of poverty and that the poverty is a multidimensional concept.

[1] Analogically, the index could be a weighted geometrical average instead of arithmetic average.

1.5 Human Development Index(HDI)

According to UNDP (1990–2014), HDI is calculated from indicators belonging to the three dimensions of long and healthy life, knowledge and a decent standard of living. HDI is a composite index published by UNDP each year for hundreds of countries. It has been used as a measure of poverty and also for identification purposes. While it was criticized shortly after the publication of the first annual report by McGillivray (1991) and numerous others it remains as an important way to complement GDP as a poverty indicator. The theory behind HDI is capability theory of Sen (1973). We pay special attention to HDI because of its wide use. Also, the data we use in the application of DEA includes indicators of the same basic capabilities as HDI and so the two approaches can be compared.

HDI is a capability index, based on the idea that long and healthy life, knowledge and a decent standard of living are three irreducibly important dimensions of human life. It used to be calculated as a weighted arithmetic average but has been changed to a geometric average in 2010.

1.5.1 Calculation of the HDI

Long and healthy life is represented by life expectancy at birth. Knowledge is reflected in the education index which is a combination of mean years of schooling and expected years of schooling. A decent standard of living is reflected in the indicator GNI per capita in PPP \$. The individual indicators are normalized as follows:

$$\text{Life Expectancy Index (LEI)} = \frac{LE - 20}{85 - 20} \tag{1}$$

$$\text{Main Years of Schooling Index (MYSI)} = \frac{MYS}{15} \tag{2}$$

$$\text{Expected Years of Schooling Index (EYSI)} = \frac{EYS}{18} \tag{3}$$

$$\text{Education Index (EI)} = \frac{\text{MYSI} + \text{EYSI}}{2} \tag{4}$$

The Income Index (II) is normalized in a different way because of the assumption about decreasing returns on income.

$$\text{II} = \frac{ln(\text{GNI}_{pc}) - ln(100)}{ln(75000) - ln(100)} \tag{5}$$

After normalization, the three dimensions are combined into a composite index.

$$\mathrm{HDI} = \sqrt[3]{\mathrm{LEI} \cdot \mathrm{EI} \cdot \mathrm{II}} \tag{6}$$

Note that the extreme values used for normalization purposes have been criticized by Alkire and Foster (2010) because their concrete values are not justified scientifically.

1.6 Desirable properties of extremal poverty indicators

Over the years there have been many attempts to improve poverty indicators and there have been several approaches to develop axioms or desirable properties in order to explicitly state the important properties of the various composite indicators. This list is adapted mostly from Havinga et al. (2005), p.55–57. Havinga et al. (2005) lists desirable properties of poverty measures which we adapted for indicators of extremal poverty. Here we consider some general properties, in Chapter 3 we look at which of them DEA satisfies.

1. Scale invariance - if population in each district doubles and everything else is unchanged, P remains unchanged.

2. Focus property - changes among people above the poverty line do not affect the set P.

3. Monotonicity - all else constant, rise in a poverty indicator of a poor person doesn't cause the district where this person lives to exit the set P.

4. Transfer Principle - all else constant, rising a poverty indicator of a poor person and lowering a poverty indicator of a less poor person by the same amount does not take the district from the set P

5. Ability to compare in time and space. This depends on availability of suitable data more than on theoretical properties of the indicator but it is important for practical purposes.

6. Recognition of different levels of human needs. This feature allows the researchers to distinguish indicators by their different contribution to the concept of poverty.

2 DEA - Data Envelopment Analysis

DEA is a technique used in operational analysis since the CCR model of Charnes et al. (1978). Its main goal is to single out the most efficient “Decision Making Units" (below, DMUs) relatively to a family of units of the same type as well as to measure the efficiency of the remaining ones. DEA is used when the performance of DMUs cannot be measured by a single numerical indicator in itself. A Decision Making Unit (DMU) is a unit able to make decisions about the process of turning inputs into outputs. It is pareto efficient if it can only produce more of an output by either producing less of another output or by consuming more inputs.

If we assume that the most efficient DMUs are producing optimally, we can construct a production function or as it is often called in DEA the efficiency frontier. This is the set of pareto efficient DMUs.

If the number of inputs and outputs together is two or three, it also possible to find the efficient frontier visually. However, DEA uses linear programming approach in order to find the DMUs which form the efficient frontier.

2.1 Additive Envelopment Model With Variable Returns to Scale

The additive model has first been introduced by Charnes in 1985.

Let us assume we have $n \in \mathbb{N}$ DMUs, each with $p \in \mathbb{N}$ inputs $x_1, \ldots, x_p$ and $m \in \mathbb{N}$ outputs $y_1, \ldots, y_p$. The real values of these are collected in matrices $X_{n,m}$ and $Y_{n,p}$. By $o \in 1, \ldots, n$ we denote the index of the DMU which we examine. The efficiency for the DMUo can then be expressed using the following model:

$$z_o = \max_{\lambda, s^-, s^+} \sum_{k=1}^{p} s_k^- + \sum_{j=1}^{m} s_j^+ \tag{7}$$

$$\text{subject to } \sum_{i=1}^{n} x_{ik}\lambda_i + s_k^- = x_{ok},\ k = 1, \dots, p, \tag{8}$$

$$\sum_{i=1}^{n} y_{ij}\lambda_i - s_j^+ = y_{oj},\ j = 1, \dots, m, \tag{9}$$

$$\sum_{i=1}^{n} \lambda_i = 1, \tag{10}$$

$$\lambda_i \geq 0, i = 1, \dots, n, s_k^- \geq 0, k = 1, \dots, p, s_j^+ \geq 0, j = 1, \dots, m. \tag{11}$$

Objective function of an envelopment model takes values from the interval $< 0, \infty)$. Efficient DMU in the additive model is one for which the objective function value is 0. Efficient frontier is the border of the convex envelopment formed by the efficient DMUs. It is the set of points which give the optimal solution 0.

On the figure 1 we can see the illustration of the additive model for the case of two outputs with constant input. Points B, C and D are efficient and they form the basis for the efficient frontier, point A is not efficient, it is clear that value of its objective function will be positive.

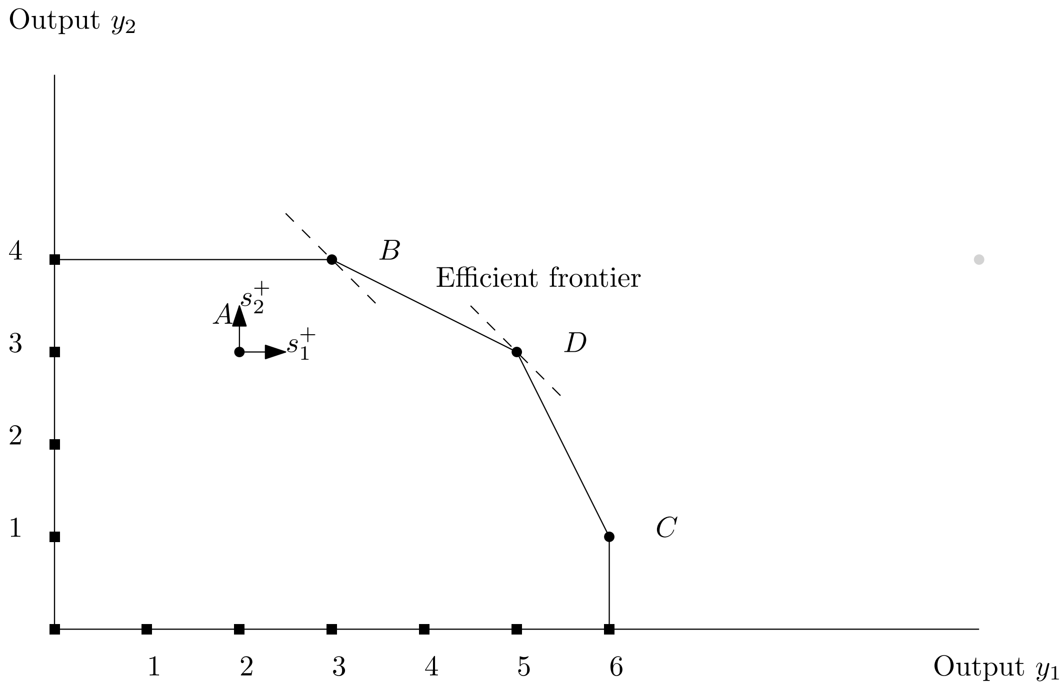

Figure 1: Illustration of the additive model and its efficient frontier

3 DEA Efficiency as a Poverty Indicator

Although most of the DEA application concern desirable outcomes of processes, there are several notable applications also on the undesirable extremes as in Gomes and Lins (2007), Liu and Chen (2009), Brunovský et al. (2009) and Mahlberg and Obersteiner (2001).

Brunovský et al. (2009) applies DEA in context of meteorology to identify extreme weather events. The units in this context are K-day periods of measurements.

Mahlberg and Obersteiner (2001) applied DEA on indicators used in HDI and compared the results with the current weighted index approach to HDI. Although HDI has been used in other works as an indicator of poverty, this application focused on finding the wealthiest countries and not on finding the poorest.

The main specifics of applying DEA to identify the poorest units are as follows:

- DMUs can be districts, individuals, households, cities or countries. For this thesis we consider districts.
- The interpretation of the efficient frontier is finding worst practice rather that best practice. Rather than talking about efficiency we will talk about extremal districts.
- The indicators of poverty are outputs. There are no inputs to be compared.

It is important to note that we use the DEA approach to identify the extremally poor DMUs among the given selection of DMUs. We do not attempt to measure poverty in a cardinal sense or to create a ranking of poor DMUs as in other approaches like HDI.

3.1 Best and Worst Practice Frontiers

In most applications of DEA the focus is on finding the best practice frontier. In the case of identifying poverty it is more meaningful to search for the worst practice frontier. The mathematical model finding worst practice can be the same to the best practice model but the interpretation is different. While computing the best practice frontier we assume that this frontier represents the productivity available in the best possible conditions, the worst practice frontier represents productivity in the worst conditions.

A figure 2 from Liu and Chen (2009) suggests that if we only look at the positive picture of the data presented by the best case frontier, we can ignore that some of the DMUs in the best practice frontier may be also a part of the worst practice frontier which can mean they are at risk. In chapter 5 we test that this does not happen with the districts of Slovakia.

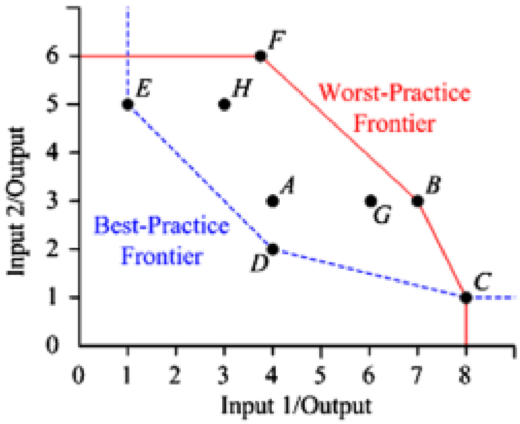

Figure 2: Illustration of the worst practice frontier

DEA models are capable to model DMUs with multiple inputs and outputs. However, the indicators of poverty such as mean household income or life expectancy of men are all modelled as outputs in this paper. We normalize tham and subtract them from 1 to make them into indicators of poverty as in section 1.

Worst practice frontier is then interpreted as the poverty frontier - frontier of the combinations of poverty indicators which are found to be the most poor according to our selection of indicators.

Identifying poverty is a specific problem because poverty is considered as an outcome for individual, district or country. If we compare individuals from the same region or comparable regions or countries, it is sensible to consider all the indicators of poverty as outputs and to consider any inputs as equal. In this case, the additive model can be simplified by setting $x_{ij} = 1$ for all $i, j \in M, N$. From that follows that the $\forall k \in 1, \ldots, ps_k^- = 0$. According to Halická (2013), if we have only one constant input as in our case, we obtain the same extremal districts with or without the $\sum_{i=1}^{n} \lambda_i = 1$ condition (variable vs. constant returns to scale).

$$z_o = \max_{\lambda, s^+} \sum_{i=1}^{m} s_i^+ \tag{12}$$

$$\text{subject to } \sum_{i=1}^{n} y_{ij}\lambda_i - s_j^+ = y_{oj} \; \forall j \in 1, \dots, m \tag{13}$$

$$\sum_{i=1}^{n} \lambda_i = 1 \tag{14}$$

$$\forall i \in 1, \dots, n, \forall j \in 1, \dots, m, \lambda_i \geq 0, s_j^+ \geq 0. \tag{15}$$

Denote $(\hat{\lambda}, \hat{s^+}) = \text{argmax} z_o$. We call j a reference DMU for DMU$_o$ if $\hat{\lambda}_j > 0$. For an efficient DMU i we denote a set of all DMUs for which it is a reference DMU as a reference group of DMU i We will call these sets reference groups in this paper. s^+ informs us in which indicators is the DMU not the poorest.
$P = i, z_i = 0$.
The additive model with constant returns to scale is equivalent to CCR model as regards the identification of extremal districts. Because we have only 1 constant input $x_i 1 = 1, \; i = 1, \dots, n$, the conditions $\sum_{k=1}^{p} v_k x_{ko} = 1$ and $\sum_{j=1}^{m} u_j y_{ij} - \sum_{k=1}^{p} v_k x_{ik} \leq 0, \; i = 1, \dots, n$ reduce to $\sum_{j=1}^{m} u_j y_{ij} \leq 1, \; i = 1, \dots, n$.

$$CCR_o = \max_{u_r, v_i} \sum_{j=1}^{m} u_j y_{oj} \tag{16}$$

$$\text{subject to } \sum_{j=1}^{m} u_j y_{ij} \leq 1, \; i = 1, \dots, n, \tag{17}$$

$$u_j \geq 0, \; j = 1, \dots, m. \tag{18}$$

This allows us to interpret the model intuitively as maximizing the efficiency score through possible positive weights so that no district will have efficiency score larger than 1. In CCR model, P is defined as $P = o, CCR_o = 1$.
Interesting property of this DEA approach is that P always includes P computed by Rawlsian Minimax approach. However, P obtained by DEA approach sometimes finds more districts. Using the CCR formulation we can see that if $y_{o\hat{j}} = \max y_{i\hat{j}}, \; i = 1, \cdots, n$, if we select $u_j = 0, j \neq \hat{j}, \; u_{\hat{j}} = \frac{1}{y_{oj}}$ in CCR model, the objective value of CCR_o will be 1 and the district o will be singled out as extremally poor.

3.2 Desirable properties of extremal poverty indicators and DEA

In this section we discuss to which extent Havinga's requirements are satisfied. We adapt them from Havinga et al. (2005) to apply to the question solved in this thesis. We consider these properties more generally in Chapter 1, here we look at which of them DEA approach satisfies.

- **Scale invariance** Because we normalize partial indicators prior to the use of DEA, the method used is scale invariant.
- **Focus property** Whether an indicator using DEA has poverty focus in terms of individuals or not depends on the indicators of poverty which we choose as outputs. The use of average values like average income do not have a poverty focus because they relate to the whole population, the use of other dimensions like the rate of social benefit claims in a population or the rate of unemployed can satisfy this axiom. If we interpret the poverty focus to mean focusing on the poor districts in the country than using DEA gives us a poverty focus because we find the poorest districts.
- **Monotonicity** Rise in one of the indicators of a poor person (for example income, education attained or health status) ceteris paribus cannot make the district exit the set P.
- **Transfer principle** Transfer from the poorer person to the richer person cannot take the district out of P.
- **Ability to compare in time and space** This depends on availability of appropriate data. Use of standard DEA models assumes that all the indicator values are known and measured the same way. This can be a problem because there is often a lack of common methodology in collecting poverty data across countries or across time periods. The study Hajko et al. (2011) was only conducted in 2009. For some of the data which are collected annually can be compared the set

of poorest districts according to DEA. However, every DEA analysis is relative to the data in a given year.

- **Recognition of different levels of human needs** A basic assumption of DEA is that all the dimensions are equally important. Use of assurance regions or of additive model with weights can attach different importance to different dimensions in DEA analysis. However, this would require further justification. Also, logarithm can be applied on the income related dimensions to capture the assumption of decreasing returns to scale.

DEA approach provides scale invariance, monotonicity and transfer principle. Focus property and ability to compare in time and space can be assured by selection of suitable indicators. Recognition of different levels of human needs could be achieved but we do not attempt it in this thesis.

4 Application of DEA: Data

This section is dedicated to the dataset used in our application of the DEA in poverty identification. This section concerns the criteria used in choosing the data, description of the selected data and notes on corellation and PCA results.

4.1 Poverty data

In order to identify poor districts with DEA it is necessary to have a dataset which is aggregated at the level of districts. Follows a general framework for data classification and some examples of datasets available to researchers which proved useful when searching for regional data.

4.1.1 NUTS classification of regions

NUTS is abbreviation for Nomenclature of territorial units for statistics. It is the EU wide classification.

- NUTS 1: There are 98 of them in EU. One NUTS 1 element includes the whole Slovak Republic (SR).
- NUTS 2: There are 273 of them in EU. SR includes 4 oblasts.
- NUTS 3: There are 1324 of them in EU. SR includes 8 regions.
- LAU1 (Local Administrative Unit, previously NUTS 4): 8397 LAU1 units in EU, 79 districts in SR.
- LAU2: 120432 LAU2 units in EU, 2928 municipalities in SR.

Different level of NUTS/LAU would be required for different DEA analysis. For analysing of the whole EU, NUTS 1 would be the best option, for analysing SR, best option seems to be LAU1. Different approach could be analysing NUTS 3 level of Central European countries. Important fact to note is that there is much more public data available at higher levels of NUTS. More information containing also other countries of the EU can be found in Charron et al. (2014).

4.1.2 SILC

SILC EU is a EU wide dataset evaluating social conditions, deprivations and poverty at levels NUTS 1, 2 and 3. Therefore it is not suitable for analysis of SR alone because 8 is a low number but it could be useful for DEA analysis of central Europe or EU. This dataset has been used by Bartošová and Želinský (2013).

4.1.3 Regional data

RegDat database úrad Slovenskej republiky (2009) includes various variables which are connected to poverty, for example unemployment statistics, however, it has only data at levels NUTS 1-3, therefore it is not suitable for the purposes of my DEA analysis. Possible use could include using DEA with a time series analogically to the Brunovský et al. (2009) to compare poverty indicators in various years and to use years as DMUs. This would be an analysis of the poorest years as opposed to poorest districts or regions. Competitive Regions 21 Hajko et al. (2011) was the regional dataset used in this paper. Section 5 includes more information about the selected indicators.

4.2 Multidimensional Poverty in Slovakia

Multidimensional poverty in Slovakia has been studied before, notably by Michálek (2004), Džambazovič (2007), Rochovská and Horňák (2008), Gerbery et al. (2009), Michalek and Zarnekow (2012), Ivanová (2013), Bartošová and Želinský (2013) and Bauer et al. (2014).

Michálek (2004) made a weighted index for districts of Slovakia. He used indicators: the lack of employment or education, large and/or single-parent family, an overcrowded apartment, the lack of a bathroom and of a car. For further analysis, it could be interesting to apply DEA to his data. Two of his indicators (lack of employment and of education) is used also in our data. One reason why we didn't decide to use same data as Michalek was the fact that his analysis is ten years old and we prefer the most recent data available.

Džambazovič (2007) describes historical and sociological connotations of poverty in Slovakia. He observes different modes of poverty emergent in the transformation towards the market economy and concentrates on the changes in the sociological discourse on

poverty. He talks also about spatial aspects of poverty and marginal regions. According to him, the regional differences stem from the 19^{th} and early 20^{th} century because of different pace of industralization. During socialism the industralization was planned in the central parts which lead to stagnation of periferal regions.
Rochovská and Horňák (2008) concentrate on subjective perception of poverty in the marginal regions of Slovakia. The authors chose two periferal regions to conduct a survey about perception of poverty and quality of life. More than a third of respondents who wrote that they considered their household poor wrote that they were very satisfied or reasonably satisfied with the material situation of the household. This is an interesting result because it reveals that even the subjective notion of poverty is multidimensional.
Gerbery et al. (2009) mentions 32 national indicators of poverty for SR. Most of these indicators are made on the basis of EU SILC (data for 8 regions) which does not provide enough units for the purpose of DEA. However, one of the indicators called "rate of marginal districts out of all districts" identifies so called marginal districts on the basis of 8 indicators: index of aging, life expectancy of men and women, net migration, rate of registered unemployment, average monthly wage of employees, rate of the people in material deprivation and number of cases of tuberculosis on 100 thousand people. A district is called marginal if at least 4 of its indicators have marginal values. That means that the district has worse values than is the value of the last quartile. In some cases the last quartile is the bottom quartile and in some the top quartile, depending on the nature of the indicator. In this thesis, we consider 4 out of the indicators used in this national indicator: life expectancy of men and women, rate of registered unemployment and average monthly wage of employees.
Michalek and Zarnekow (2012) provide 340 indicators for rural districts of Slovakia. However, focus of our thesis is not on actual application but on the DEA methodology which to our knowledge has not been used before in the context of poverty. Using so many indicators with DEA would be problematic because these indicators would have to be grouped and aggregated for the use with DEA. It is not a good practice in DEA to have more indicators than units studied.
Bartošová and Želinský (2013) review the history of poverty measurement in Czechoslo-

vakia. Prior to split, the research of poverty was not encouraged because the existence of poverty did not agree with the ideological claim of the existence of equality among the citizens. They also discuss the differences in poverty in the Slovakia and Czech Republic 15 years after the split.

4.3 Criteria

The choice of our data was based on several criteria.

1. Choose data with dimensions suitable for the use of DEA. I was seeking 5-10 outputs and 50-150 units. (Compare with Sarkis (2007), p. 1–2) This led to choosing data at the level of districts.

2. Good quality of data information. I didn't want data with many missing values. For this reason I decided not to use data for Multidimensional Poverty Index published by Oxford Poverty and Human Development Initiative

3. Results should be comparable with another approach to identify poor units.

4.4 Selected Data

Data we used have been collected in a study of Hajko, Klatik and Tunega Competitive Regions 21 Hajko et al. (2011). The study was not focused on poverty but on regional disparities affecting business in Slovakia. Data was obtained from publicly available statistics and a survey of entrepreneurs from all 79 districts of Slovakia. The authors list 106 parameters for 79 districts of Slovakia. We have chosen those listed parameters which are internationally accepted to relate to poverty. These parameters included:

- Development potential of the district (DP) was a score according to results of a survey question with 6 possible answers. How do you perceive the development potential of your district? The answers ranged from 1 – will be the slowest developing district in the Slovak Republic to 6 – will be the fastest developing district in the Slovak Republic.

- Social benefit claims (SBC) in €.

- Inflow of foreign direct investments (IFI) - no unit. Inflow of foreign direct investments into the district by the year 2008 per capita.
- Life expectancy of men and women (LEM, LEW) in years.
- Registered unemployment rate (RUR) in %. The registered unemployment rate is calculated according to the methodology of the Ministry of Labour, Social Affairs and Family as a proportion of available jobseekers to the total economically active population of the district.
- Share of long-term jobseekers (SLJ) -it is an index. This index is a function of the length of the period of unemployment of all jobseekers and is calculated from the number of people in 11 different groups according to the length of the registered period.
- Average monthly wage (AMW) in slovak crowns. It is the average monthly wage in industry (natural persons). The data are from the last year before the change of currency to €.
- Level of education (LOE) - a survey question with 6 levels. How do you perceive the level of education of people in your district? 1 – as the lowest among all districts in Slovakia 6 – as the highest among all districts in Slovakia

4.5 Normalization

The data had different ranges. For example, the share of long term jobseekers was in percent value, the inflow of direct investments was in Slovak Crowns. According to Sarkis (2007), common practice in DEA methodology is dividing each Input and Output by its mean. However, I have decided to use a different procedure instead to make all outputs in scale of 0 to 1.

For the outputs the increase of which reflects increasing poverty the formula I used was $\frac{Y_i - Y_{min}}{Y_{max} - Y_{min}}$. For outputs like Life expectancy which rise as poverty decreases I have used analogous formula: $1 - \frac{Y_i - Y_{min}}{Y_{max} - Y_{min}}$. To sum up, for the use of DEA I have normalized my data to a scale from 0 to 1 where 1 indicates the most poor of the districts in a given variable and 0 the least poor.

For normalized indicator values of all districts, please refere to B.

4.6 Correlation Analysis

A corellation plot (Figure 3) showed that the variables could be divided into highly correlated groups: Corellation could be a problem if we attempted to rank the districts

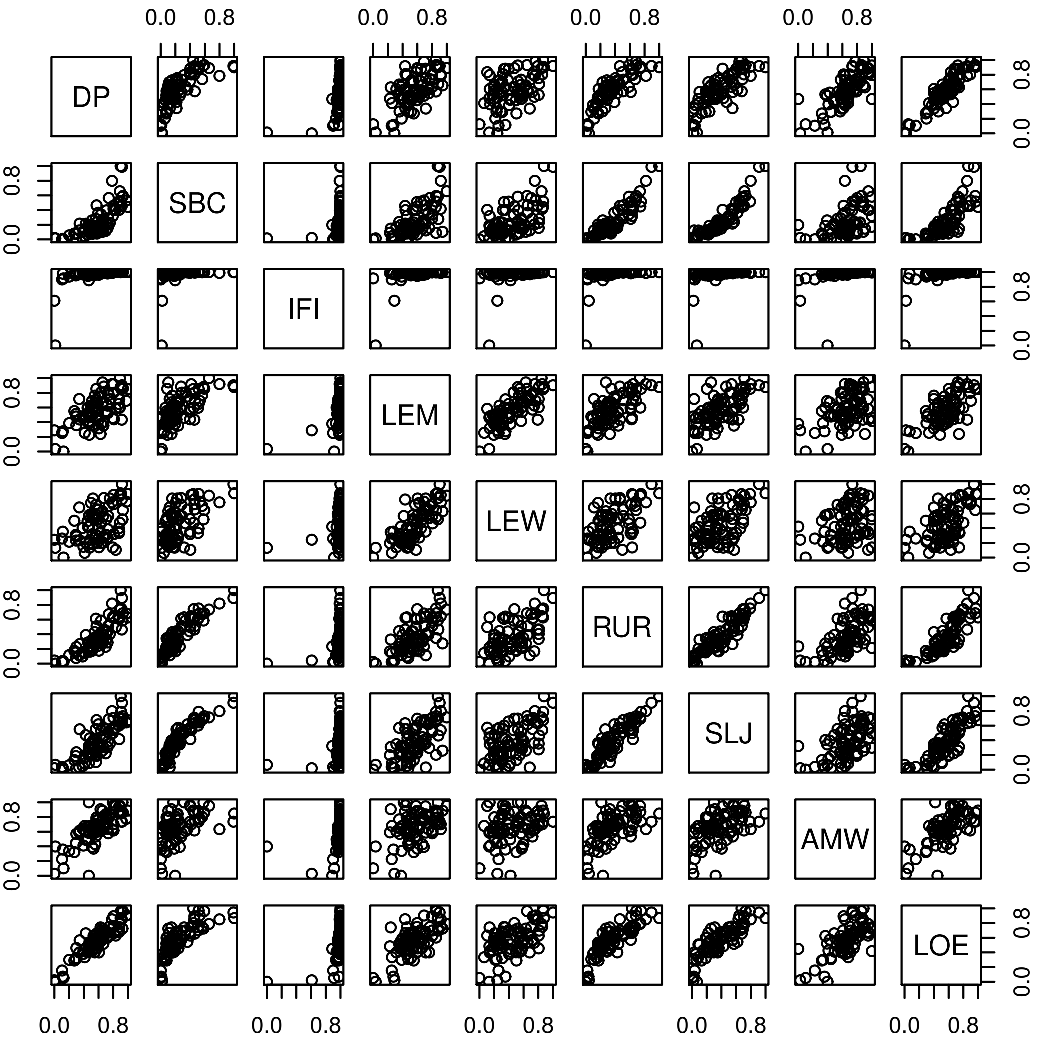

Figure 3: Multiplot of chosen variables

by poverty but because our objective is the identification of the poorest districts this is not the case. Sarkis has ambiguous stance towards removing corellating inputs and outputs. He shows on an example that even when two outputs are perfectly corellated, the model with both and just one of them shows minor difference in the efficiency scores. Using less outputs in an analysis can reduce "time in data acquisition, storage, calculation"Sarkis (2007). The concrete efficiency scores of inefficient DMUs are not important in our application of DEA. The results important for this paper is the set P of "poorest districts". Therefore we don't need to be concerned about differences in efficiency scores. What could cause concern is that the number of parameters influences the number of "poor" districts. We will return to this in more detail in the next section.

Table 1: Corellations of 9 indicators

	DP	SBC	IFI	LEM	LEW	RUR	SLJ	AMW	LOE
DP	1.00	0.76	0.47	0.65	0.50	0.86	0.81	0.77	**0.92**
SBC	0.76	1.00	0.23	0.66	0.62	**0.91**	**0.92**	0.49	0.77
IFI	0.47	0.23	1.00	0.39	0.23	0.32	0.29	0.38	0.48
LEM	0.65	0.66	0.39	1.00	0.78	0.69	0.64	0.44	0.70
LEW	0.50	0.62	0.23	0.78	1.00	0.58	0.58	0.29	0.58
RUR	0.86	**0.91**	0.32	0.69	0.58	1.00	**0.91**	0.60	0.85
SLJ	0.81	**0.92**	0.29	0.64	0.58	**0.91**	1.00	0.54	0.83
AMW	0.77	0.49	0.38	0.44	0.29	0.60	0.54	1.00	0.69
LOE	**0.92**	0.77	0.48	0.70	0.58	0.85	0.83	0.69	1.00

4.7 PCA

The component analysis of normalized indicators suggests that there is a common factor which accounts for the most of the variability, namely, the regional poverty (or wealth because the number signs are negative). Another important factor is foreign investment.

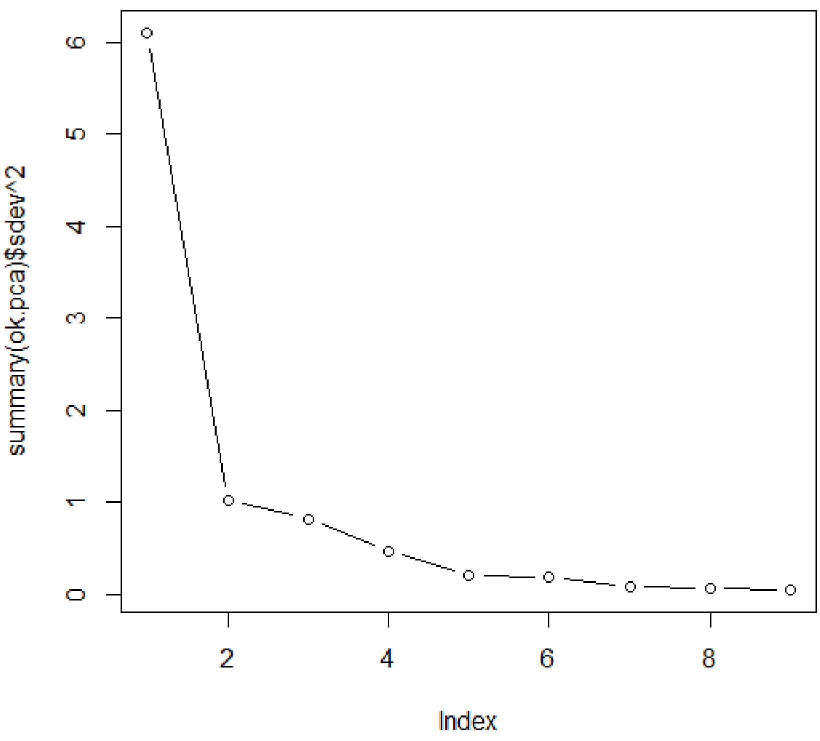

Figure 4: Elbow diagram of PCA

Table 2: Principal components

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9
DP	-0,375	0,208	-0,172	0,072	-0,123	0,375	-0,333	-0,596	0,398
SBC	-0,361	-0,265	-0,194	-0,299	0,146	-0,392	-0,084	-0,440	-0,545
IFI	-0,190	0,656	0,545	-0,422	0,172	-0,166	-0,017	0,013	0,004
LEM	-0,327	-0,187	0,455	0,246	-0,707	-0,247	0,164	-0,052	0,015
LEW	-0,283	-0,422	0,504	0,305	0,588	0,160	-0,123	0,027	0,085
RUR	-0,379	-0,108	-0,195	-0,205	-0,125	-0,129	-0,588	0,604	0,147
SLJ	-0,370	-0,161	-0,225	-0,309	0,140	-0,069	0,651	0,092	0,484
AMW	-0,286	0,434	-0,288	0,661	0,207	-0,374	0,108	0,113	-0,065
LOE	-0,379	0,131	-0,056	-0,005	-0,100	0,656	0,242	0,248	-0,525
SDev	2,470	1,011	0,907	0,687	0,449	0,427	0,292	0,251	0,232

5 Application of DEA: Results

In this chapter, we apply DEA to 2, 9, 8, 7 and 3 poverty indicators. The organization of this chapter is designed in order to be able to compare different approaches to the problem of identification of the poorest districts considering the same indicators. We compare results of DEA with the results of a weighted index and with the Rawlsian Maximin approach. In the case of 3 indicators, we calculate the HDI and compare the results with it. We present the results spatially on a map of Slovakia.

5.1 Illustrative example of 2 indicators: Development potential and social benefit claims

First we analyse the 2 dimensional case. It is the simplest one to use with DEA. It has a good feature that it can be easily illustrated and checked on the figure 5 so that people not very familiar with DEA can get an intuition of what it does.

In Table 3 the values of objective function are listed by the alphabetical order of the districts they correspond to. The row numbers refer to the tens of alphabetic rank and the column numbers to the units. The optimal value **0** is bolded because it corresponds to the extremally poor districts. In Table 4 we can see the parameter values for the

Table 3: Values of the objective function (2 parameters)

	0	1	2	3	4	5	6	7	8	9
0		-1,10	-1,31	-0,96	-0,80	-1,88	-1,89	-1,80	-1,78	-1,79
10	-0,89	-1,20	-1,15	-0,93	-1,10	-1,28	-1,45	**0,00**	-1,41	-0,98
20	-1,32	-0,48	-0,93	-0,65	-1,13	-1,24	-1,27	-1,21	-0,69	-1,21
30	-0,95	-0,53	-1,23	-0,60	-1,56	-1,19	-0,52	-0,87	-1,25	-1,25
40	-1,30	-1,53	-1,08	-1,13	-1,51	-1,34	-0,06	-1,35	-1,24	-1,00
50	-1,17	-1,39	**0,00**	**0,00**	-0,33	-1,19	-0,57	-1,60	-1,23	-1,27
60	-0,62	-0,36	-0,80	-0,97	-0,67	-0,70	-1,14	-1,19	-0,36	-1,45
70	-1,65	-1,04	-1,25	-0,31	-0,52	-1,03	-1,24	-1,06	-1,27	-1,54

extremally poor districts. We can see that compared with the Rawlsian maximin approach (simply considering extremes in each individual dimension), Revuca is added to the set of the extremally poor districts P.

Table 4: Parameter values for the extremally poor districts (DP and SBC)

	Development Potential	Social Benefit Claims
Gelnica	**1,00**	0,44
Revúca	0,92	0,99
Rimavská Sobota	0,90	**1,00**

5.1.1 Reference Groups

- Gelnica: Gelnica, Poltár, Sobrance, Veľký Krtíš
- Revúca: all districts except other districts in the poverty frontier
- Rimavská Sobota: none except itself

The list of reference groups allows us to see which extremally poor district is a reference district for all the districts. The program maximizes the sum of slacks between a district i and a point x located on the poverty frontier. The point x is always a linear combination of the extremally poor districts. If a district in the poverty frontier j contributes to this combination, i is included in the reference group of j. In the 2 dimensional case on Figure 5 we can see what does it mean geometrically. For each district we find the furthermost feasible point in its nonnegative quadrant. For most districts this will be Revúca, only some districts which do not include it in their nonegative quadrant will select Gelnica. In order to check the possibility that some of the districts could be extremally poor but also extremally wealthy according to DEA, we computed also the wealthiest districts.

5.1.2 Extremally Wealthy Districts With 2 Variables

The resulting wealthiest districts with 2 variables were Bratislava I, Bratislava II, Bratislava III and Bratislava IV. We can see that none of the districts were both wealthiest and poorest. We obtained similar results also with considering more indicators. Because Bratislava is the capital and the biggest city of Slovakia, it is not very

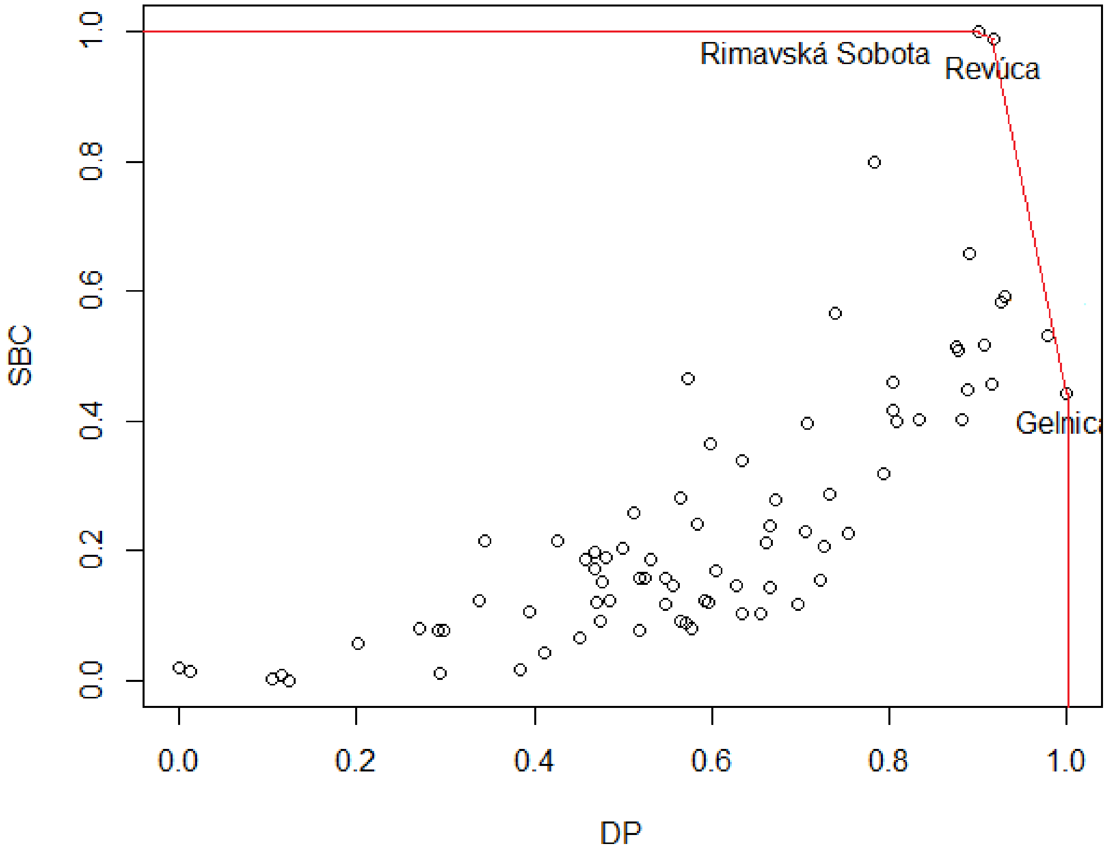

Figure 5: Graphically represented poverty frontier in 2D case obtained using R's plot function

surprising that all the wealthiest districts are there. In some of the selections of indicators also Košice II was singled out as one of the extremaly wealthy districts. Košice is the second largest city of Slovakia.

5.2 9 Dimensional Case

Values of the objective function of the complete analysis of 9 variables can be found in table 5. Note that two districts, Sobrance and Kosice II, were omitted from this analysis because the information of their average monthly wage is not publicly available. Table 6 shows parameter values for the extremally poor districts. Each extremal district in this case has at least one parameter equal to its largest possible value. This means that in this case we would have gotten the same result using Rawlsian Maxmin principle. Even though some of the districts, for example Gelnica, are extremal in more than

Table 5: Values of the objective function (9 parameters)

	0	1	2	3	4	5	6	7	8	9
0		-3,82	-4,55	-2,33	**0,00**	-7,71	-7,10	-6,41	-7,15	-6,15
10	-2,75	-3,53	-2,03	-2,65	-4,01	-4,01	-4,41	**0,00**	-5,12	-3,27
20	-4,48	-0,47	-2,65	-1,72	-4,65	-5,01	-3,61	-1,56	-3,69	-2,82
30	**0,00**	-4,37	-1,68	-5,07	-4,33	**0,00**	-2,38	-4,36	-3,37	-4,64
40	-4,87	-3,47	-3,29	-4,95	-4,96	**0,00**	-4,15	-4,15	-3,54	-4,35
50	-4,32	**0,00**	**0,00**	-0,83	-3,82	-1,20	-5,53	-3,54	-4,33	-1,01
60	-2,82	-2,62	**0,00**	**0,00**	-3,88	-3,94	**0,00**	-5,37	-5,76	-3,04
70	-4,13	-0,37	-1,41	-3,77	-4,39	-2,97	-4,04	-5,12		

1 dimension, number of districts in P is not lowered. This is because IFI has six districts with the maximum value. This result confirms findings of precious studies like Michálek (2004) that the poorest regions of Slovakia are located on east, south and close to the border. Table 10 shows the ranks which the DEA extremally poor

Table 6: Parameter values for the extremally poor districts (9 parameters)

	DP	SBC	IFI	LEM	LEW	RUR	SLJ	AMW	LoE
Bardejov	0,79	0,32	0,9996	0,46	0,14	0,63	0,50	**1,00**	0,74
Gelnica	**1,00**	0,44	**1,0000**	0,85	0,86	0,63	0,67	0,89	**1,00**
Levoca	0,92	0,46	**1,0000**	0,43	0,79	0,46	0,63	0,88	0,85
Medzilaborce	0,87	0,52	**1,0000**	0,65	0,40	0,62	0,66	0,86	0,70
Poltár	0,98	0,53	**1,0000**	0,69	0,76	0,68	0,64	0,78	0,90
Revúca	0,92	0,99	0,9989	0,90	**1,00**	0,89	0,91	0,74	0,94
Rim. Sobota	0,90	**1,00**	0,9978	0,88	0,88	**1,00**	**1,00**	0,85	0,87
Stropkov	0,83	0,40	**1,0000**	0,58	0,11	0,50	0,57	0,95	0,63
Svidník	0,81	0,40	**1,0000**	0,44	0,32	0,62	0,56	0,96	0,66
Trebišov	0,89	0,66	0,9992	**1,00**	0,84	0,74	0,71	0,95	0,79

districts obtained with the arithmetic mean method. Average rank of the extremally poor districts was $11,2$, while average of all ranks is 39 and average of lowest 10 ranks is $5,5$. All extremally poor districts were in the poorest third of districts according to the average. 10 **poorest districts according to the mean of** 9 **indicators**:

Table 7: Mean ranks of the DEA extremally poor districts (9 parameters)

	R_a
Bardejov	23
Gelnica	5
Levoča	13
Medzilaborce	14
Poltár	8
Revúca	2
Rim. Sobota	1
Stropkov	24
Svidník	19
Trebišov	3

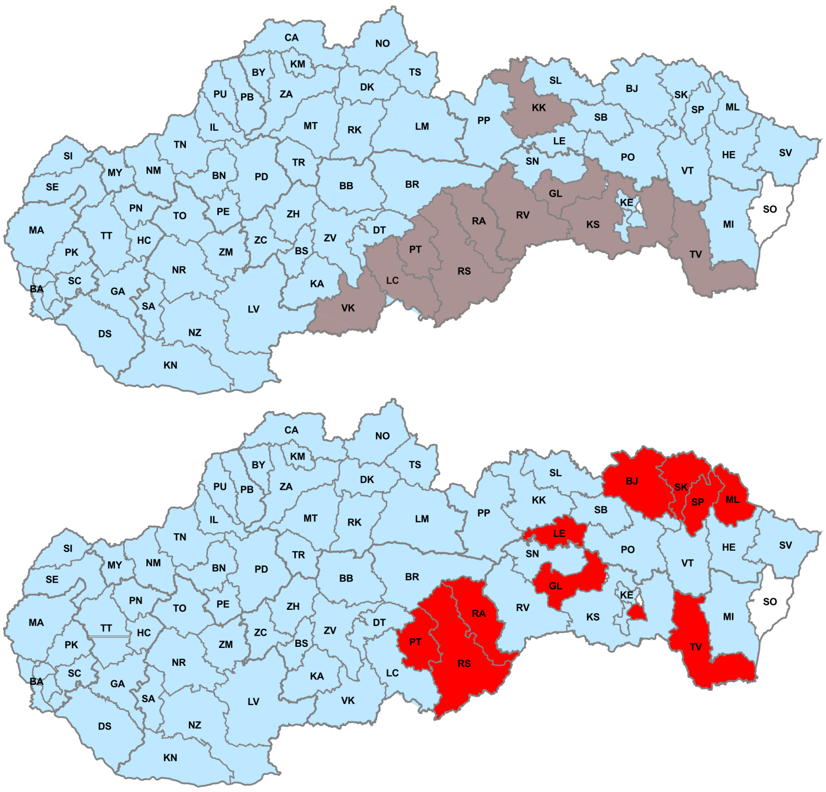

Figure 6: Comparison of P according to average and DEA of 9 indicators

Rimavská Sobota, Revúca, Trebišov, Kežmarok, Gelnica, Rožňava, Veľký Krtíš, Poltár, Lučenec, Košice - okolie.

5.3 8 Dimensions: Without Inflow of Foreign Direct Investments

I have excluded foreign investments dimension because it is not so clear that it really indicates poverty and this dimension is not widely used for the purpose of measuring poverty as opposed to the rest of the measures.
Table 8 lists values of the objective function.

Table 8: Values of the objective function (8 parameters)

	0	1	2	3	4	5	6	7	8	9
0		-3,82	-4,53	-2,36	**0,00**	-6,71	-6,71	-6,31	-7,07	-6,13
10	-2,75	-3,53	-2,49	-2,69	-4,00	-4,00	-4,38	**0,00**	-5,09	-3,27
20	-4,47	-0,47	-2,65	-1,72	-4,61	-4,90	**0,00**	-1,56	-3,65	-2,82
30	-1,58	-4,36	-1,68	-5,05	-4,33	-2,02	-2,38	-4,36	-3,36	-4,64
40	-4,86	-3,46	-3,29	-4,95	-4,95	-0,64	-4,14	-4,14	-3,54	-4,35
50	-4,28	**0,00**	**0,00**	-0,82	-3,80	-1,66	-5,51	-3,53	-4,28	-1,61
60	-2,81	-3,27	-2,03	-1,59	-3,84	-3,93	**0,00**	-5,35	-5,70	-3,04
70	-4,13	-0,44	-1,94	-3,76	-4,38	-2,96	-4,02	-5,08		

Table 9 lists the extremally poor districts for 8 indicators and their parameter values. We can see that the poverty frontier includes less districts than the number of

Table 9: Parameter values for the DEA extremally poor districts (8 parameters)

	DP	SBC	LEM	LEW	RUR	SLJ	AMW	LOE
Bardejov	0,79	0,32	0,46	0,14	0,63	0,50	**1,00**	0,74
Gelnica	**1,00**	0,44	0,85	0,86	0,63	0,67	0,89	**1,00**
Revúca	0,92	0,99	0,90	**1,00**	0,89	0,91	0,74	0,94
Rimavská Sobota	0,90	**1,00**	0,88	0,88	**1,00**	**1,00**	0,85	0,87
Trebišov	0,89	0,66	**1,00**	0,84	0,74	0,71	0,95	0,79

dimensions. This is because Gelnica is extremal in 2 dimensions and Rimavska Sobota in 3.

Table 10 shows the ranks which the DEA extremally poor districts obtained with the arithmetic mean method. Average rank of the extremally poor districts is 7, while average of all ranks is 39 and average of lowest 5 ranks is 3. All extremally poor districts were in the poorest third of districts according to the average. Figure 7 allows

Table 10: Mean ranks of the DEA extremally poor districts (8 parameters)

	R_a
Bardejov	23
Gelnica	6
Revúca	2
Rim. Sobota	1
Trebišov	3

us to compare the two approaches to 8 indicators. The 5 poorest districts according to average are Rim. Sobota, Revúca, Trebišov, Kežmarok, Rožňava.

5.4 7 dimensions: Without Inflow of Foreign Direct Investments and Without Average Monthly Wage

The exclusion of average monthly wages allowed me to include Sobrance and Košice III in the calculations. However, neither of them were included as extremely poor. Table 12 lists the extremally poor districts obtained by DEA with 7 indicators. We can compare figures 8 and **??**. Both include districts mostly in the east and south of Slovakia. As we can see, Gelnica and Rimavská Sobota keep being extremal in more than one dimension and so prevent the model from selecting more districts as extremal. Also, all the district included in the poverty frontier are extremal in at least one dimension.

4 **poorest districts according to the mean of indicators**:

Revúca, Rimavská Sobota, Rožňava, Kežmarok.

Table 13 shows the ranks which the DEA extremally poor districts obtained with the

Table 11: Values of the objective function (7 parameters)

	0	1	2	3	4	5	6	7	8	9
0		-3,81	-4,37	-2,29	-2,98	-6,30	-5,93	-5,73	-6,36	-5,68
10	-2,52	-3,43	-2,60	-2,49	-3,87	-3,86	-4,16	0,00	-4,69	-3,23
20	-4,25	-0,40	-2,68	-1,55	-4,41	-4,09	-4,01	-3,43	-1,49	-3,27
30	-2,65	-2,01	-4,15	-1,68	-4,56	-4,13	-2,14	-2,33	-4,20	-3,35
40	-4,51	-4,62	-3,41	-3,82	-4,78	-4,73	-0,57	-3,95	-3,91	-3,50
50	-4,18	-4,14	0,00	0,00	-0,68	-3,37	-1,91	-5,15	-3,40	-3,97
60	-2,47	-1,55	-2,74	-3,31	-2,94	-2,75	-3,43	-3,82	0,00	-5,22
70	-5,23	-3,04	-4,12	-0,81	-2,29	-3,59	-4,16	-2,69	-3,71	-4,71

Table 12: Parameter values for the extremally poor districts (7 parameters)

	DP	SBC	LEM	LEW	RUR	SLJ	LOE
Gelnica	**1,00**	0,44	0,85	0,86	0,63	0,67	**1,00**
Revúca	0,92	0,99	0,90	**1,00**	0,89	0,91	0,94
Rimavská Sobota	0,90	**1,00**	0,88	0,88	**1,00**	**1,00**	0,87
Trebišov	0,89	0,66	**1,00**	0,84	0,74	0,71	0,79

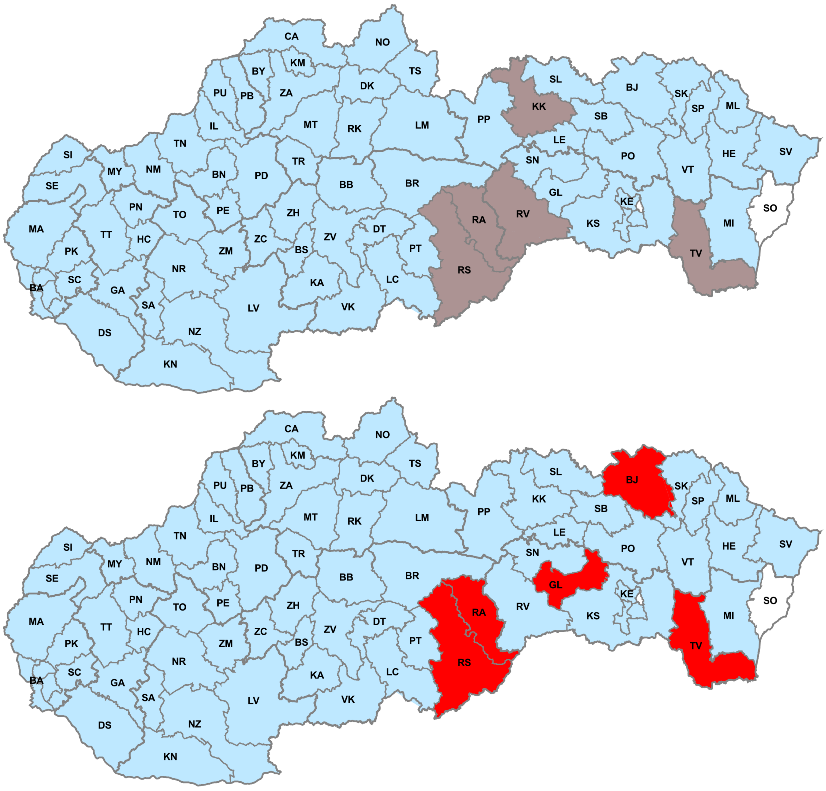

Figure 7: Comparison of P according to average and DEA of 8 indicators

arithmetic mean method. Average rank of the extremally poor districts is $3,5$, while average of all ranks is 39 and average of lowest 4 ranks is $2,5$. All extremally poor districts were in the poorest decile of all districts according to the average.

Table 13: Mean ranks of the DEA extremally poor districts (7 parameters)

	R_a
Gelnica	6
Revúca	1
Rim. Sobota	2
Trebišov	5

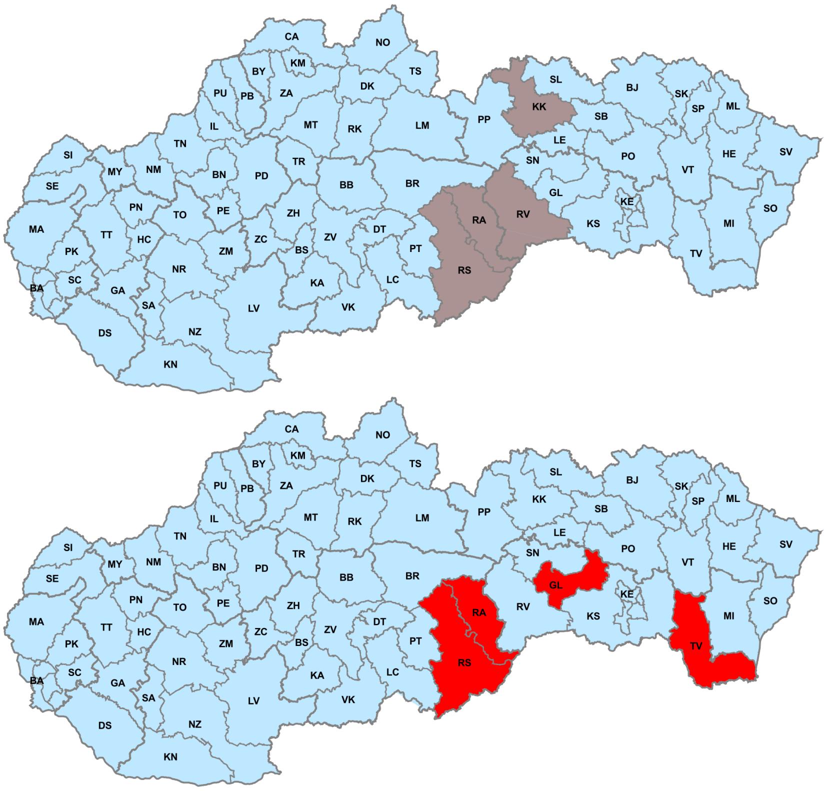

Figure 8: Comparison of P according to average and DEA of 7 indicators

5.5 3 Dimensions: Comparison of HDI and DEA Approach to the Same Indicators

5.5.1 HDI

According to UNDP (1990–2014) HDI is computed based on capability approach. Indicators used regard life expectancy, income and education level of the population. Dataset for this paper included data for Slovak districts capturing these three dimensions. Follows a comparison of results of HDI computed using the available data and results of DEA approach to HDI. The computation of HDI had to be slightly simplified because there was only one educational index available. Also, average monthly wage was used as a proxy for the income indicator used in standard HDI. AMWo here means the original value of AMW before the normalization.

$$LEI = \frac{LEM + LEW}{2} \tag{19}$$

$$AMWI = \frac{ln(AMWo) - ln(12092)}{ln(38863) - ln(12092)} \tag{20}$$

$$HDI = \sqrt[3]{LEI \cdot AMWI \cdot LOE} \tag{21}$$

Two districts, Sobrance and Košice III, had to be removed from the analysis because the data about average income was not publicly available due to only two companies employing people from these districts.

5.5.2 Results of DEA

Table 14: Parameter values for the extremally poor districts (3 parameters)

	LEI	AMWI	LOE
Bardejov	0,59	**1,00**	0,74
Gelnica	1,72	0,89	**1,00**
Revúca	1,90	0,74	0,94
Trebišov	1,84	0,95	0,79

We can see that Bardejov, Gelnica and Trebišov are each extremal in one dimension while Revúca is selected nontrivially by the model.

5.5.3 Results of HDI

Table 15: Parameter values for the poorest districts for HDI

	LEI	AMWI:	LOE
Gelnica	0,85	0,89	**1,00**
Kežmarok	0,85	0,74	0,97
Revúca	0,90	0,74	0,94
Veľký Krtíš	0,86	0,86	0,96

Gelnica ranked first, Revúca third, Bardejov 13^{th} and Trebišov ranked 10^{th} with HDI. We can see that the scatterplot is not so easy to read as the case with 2 indicators. So

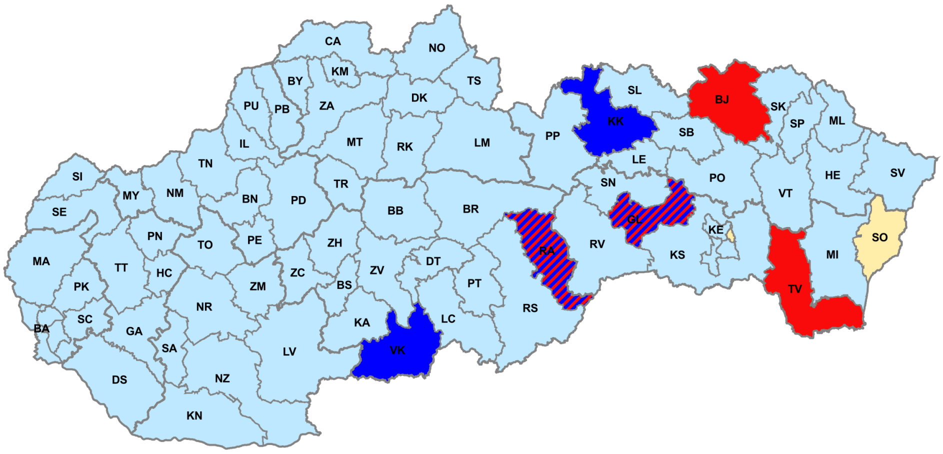

Figure 9: Comparison of DEA approach to HDI and HDI. DEA results are red, HDI are blue and excluded districts are marked yellow.

even though theoretically it should be possible to find the frontier visually, practically it is not an easy task and DEA is helpful even in this case to find the extremally poor districts.

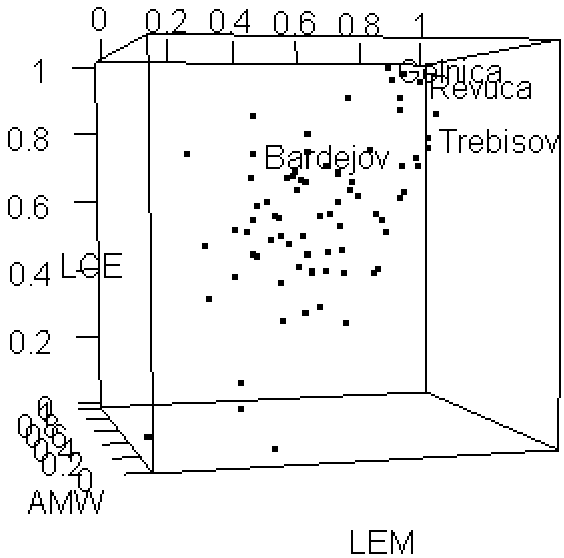

Figure 10: 3D Scatterplot of HDI indicators with DEA results

Conclusion

What does DEA bring to poverty analysis? Since the resulting sets P for DEA and average indicators were different, DEA gives us additional information about the poorest districts. DEA always finds at least the same districts as Rawlsian Minimax approach and sometimes it singles out a few more.
Our results were comparable with other analyses of multidimensional poverty in Slovakia. Similarly to Michálek (2004), extremally poor districts were found mostly in east and south of Slovakia. Gerbery et al. (2009) list on page 56 districts Revúca, Rimavská Sobota, Trebišov and Kežmarok that have been singled out as marginal in all the years 2005–2008. This result is similar to our DEA analysis of 7 parameters where we found extremally poor districts Revúca, Rimavská Sobota, Trebišov and Gelnica.
DEA does not require us to specify either weights of the indicators or the number of the poorest districts, the model determines the number of the poorest districts from the dataset itself.

DEA allows us to identify the extremally poor districts. DEA approach means that the results are necessarily relative to the dataset. Identifying the most poor districts does not tell us about their absolute poverty, only about their lack against the other districts.
The number of chosen poverty indicators influences the number of DEA extremally poor districts. Interesting result was that for the extremally rich districts, the number of indicators didn't influence their number so much. Plausible explanation for this phenomenon is the urban character of rich districts which causes the indicators of poverty to be more corellated with each other. This was also the result of Pacione's analysis of data on Scotland in 1995 as cited by Veselovská (2013).
If there are DMUs which have multiple deprivations so they are extremally poor in more than one dimension, they can cause the set of DMUs in the effective frontier to be reduced.
There could be a concern that most results are simply the districts with extreme values of each individual indicator of poverty as in the Rawlsian maximin approach. Assurance region model could be used to exclude districts with only one extreme value.

However, such an approach would insert arbitrariness in selecting the lower and upper bound into our calculations. The model used in this paper on the Slovak data shows several cases where some of the points cannot be simply inferred by looking at the table.

DEA is a method which is especially sensitive to data. It cannot work with missing values and publicly available data on poverty indicators often includes countries or districts where some values are missing. Furthermore, the dimensionality of data is constrained by the requirements of DEA (For more details, see Sarkis (2007)). However, as we have seen, there is publicly available multidimensional data which can be meaningfully used with DEA and it is plausible that the quality of available data will improve in the future.

There currently exists no publicly available and generally reliable standard software for DEA, so we implemented DEA in R using opensource library limsolve.

As regards desirable poverty identification properties, DEA approach as used here provides scale invariance, monotonicity and transfer principle. We can assure focus property and ability to compare in time and space by selection of suitable indicators. Recognition of different levels of human needs could be achieved but we do not attempt it in this thesis.

Bibliography

Alkire, S. and Foster, J. E. (2010). Designing the inequality-adjusted human development index.

Bartošová, J. and Želinský, T. (2013). The extent of poverty in the Czech and Slovak republics 15 years after the split, *Post-Communist Economies* **25**(1): 119–131.

Bauer, R., Knezova, J. and Steiner, A. (2014). Governing development of regions-case of Slovakia, *ERSA conference papers*, number ersa14p440, European Regional Science Association.

Brunovský, P., Lapin, M., Melicherčík, I., Somorčík, J. and Ševčovič, D. (2009). Risks due to variability of k-day extreme precipitation totals and other k-day extreme events, *J. Hydrol. Hydromech.* **4**(57): 250–263.

Charnes, A., Cooper, W. W. and Rhodes, E. (1978). Measuring the efficiency of decision making units, *European journal of operational research* **2**(6): 429–444.

Charron, N., Dijkstra, L. and Lapuente, V. (2014). Mapping the regional divide in europe: A measure for assessing quality of government in 206 european regions, *Social Indicators Research* pp. 1–32.

Džambazovič, R. (2007). Chudoba na Slovensku, *Diskurz, rozsah a profil chudoby. Univerzita Komenského v Bratislave* .

Gerbery, D., Bodnárová, B. and Fico, M. (2009). Monitoring národných indikátorov v oblasti chudoby a sociálneho vylúčenia.

Gomes, E. and Lins, M. E. (2007). Modelling undesirable outputs with zero sum gains data envelopment analysis models, *Journal of the Operational Research Society* **59**(5): 616–623.

Hajko, J., Klátik, P. and Tunega, M. (2011). Konkurencieschopné regióny 21, *Podnikateľská aliancia Slovenska, Bratislava* .

Halická, M. (2013). *DEA modely*, Faculty of Mathematics, Physics and Informatics, Comenius University in Bratislava, Bratislava.

Havinga, I., Kamanou, G., Morduch, J., David, I. G., Gibson, J., Ward, M., Kakwani, N. and Son, H. H. (2005). *Handbook on poverty statistics: Concepts, Methods and Policy Use*, United Nations.

Ivanová, E. (2013). Priame zahraničné investície a ekonomický rast v regiónoch SR, *Sociálno-ekonomická revue* p. 26.

Kakwani, N. and Silber, J. (2008). Introduction: Multidimensional poverty analysis: Conceptual issues, empirical illustrations and policy implications, *World Development* **36**(6): 987–991.

Laderchi, C. R., Saith, R. and Stewart, F. (2003). Does it matter that we do not agree on the definition of poverty? a comparison of four approaches, *Oxford development studies* **31**(3): 243–274.

Liu, F.-h. F. and Chen, C.-L. (2009). The worst-practice DEA model with slack-based measurement, *Computers & Industrial Engineering* **57**(2): 496–505.

Mahlberg, B. and Obersteiner, M. (2001). Remeasuring the HDI by data envelopement analysis, *Available at SSRN 1999372* .

McGillivray, M. (1991). The human development index: yet another redundant composite development indicator?, *World Development* **19**(10): 1461–1468.

Michálek, A. (2004). Meranie chudoby územnỳch celkov (regiónov) na príklade okresov SR, *Sociológia-Slovak Sociological Review* (1): 7–30.
URL: *http://www.ceeol.com/aspx/issuedetails.aspx?issueid=81b1b07f-341a-436f-9e35-559df99d9dec&articleId=f0926c17-6044-432e-8d7b-d5ea50089214*

Michalek, J. and Zarnekow, N. (2012). Application of the rural development index to analysis of rural regions in Poland and Slovakia, *Social indicators research* **105**(1): 1–37.

Rochovská, A. and Horňák, M. (2008). Chudoba a jej percepcia v marginálnych regiónoch Slovenska, *Geografia Cassoviensis* **2**: 152–156.

Sarkis, J. (2007). Preparing your data for dea, *Modeling data irregularities and structural complexities in Data Envelopment Analysis*, Springer, pp. 305–320.

Sen, A. (1973). Poverty, inequality and unemployment: some conceptual issues in measurement, *Economic and Political Weekly* pp. 1457–1464.

UNDP (1990–2014). *Human Development Report*, UNDP Human Development Report Office, New York.

úrad Slovenskej republiky, Š. (2009). *Štatistická ročenka Slovenskej republiky*, Štatistickỳ úrad Slovenskej republiky.

Veselovská, Z. (2013). Prístupy k priestorovému meraniu chudoby, *Geographia Cassoviensis* **VII**(2): 75–82.

Yang, L. (2014). *An Inventory of Composite Measures of Human Progress*, Human Development Report Office.

Appendix A

A Implementation of DEA in R

Library we used for implementing linear programming was limsolve. There are publicly available specialized DEA solvers but none of them is considered standard.

```
#Correlation analysis and Principal Components Analysis
dist<-read.table("districts.txt",header=TRUE, sep="\t")
di<-dist[1:79,2:10]; rownames(di)<-dis[1:79,1];
plot(di) #Corellation plot
cor(di)     #corellation table
ok.pca<-prcomp(ok,scale=T)
ok.pca
summary(ok.pca)
plot(summary(ok.pca)$sdev^2,type="b")
```

A.1 DEA Additive Model

```
library(limSolve)

#########################################################
#77 districts, limsolve, 9 parameters


d=77;
p=9;
dis<-di[c(-26,-61),c(1,2,3,4,5,6,7,8,9)] #26 is Kosice III, 61 Sobrance.
DEAa=c(rep(0, d),rep(1,p))
a <- DEAa*(-1)
DEAA3=cbind(rbind(t(dis),c(rep(1, d))),rbind(diag(-1,p),c(rep(0,p))))
x<-rep(1,d)
sol<-matrix(1,d,d+p)
i<-1
```

```
while(i <= d)
{
DEAb3<-c(t(dis)[1:p,i],1)
x[i]<-linp(E = DEAA3, F = DEAb3, G = NULL, H = NULL, a,
ispos = TRUE, int.vec = NULL, verbose = TRUE)$solutionNorm #value of objective fun
sol[i,]<-linp(E = DEAA3, F = DEAb3, G = NULL, H = NULL, a,
ispos = TRUE, int.vec = NULL, verbose = TRUE)$X #value or argmax lambda and s
i<-i+1;
}
x
sol

#########################################################
#77 dmus, limsolve, 8 parameters

d=77; #settings
p=8; #settings
dis<-di[c(-26,-61),c(1,2,4,5,6,7,8,9)]       #settings
DEAa=c(rep(0, d),rep(1,p))
a <- DEAa*(-1)
DEAA3=cbind(rbind(t(dis),c(rep(1, d))),rbind(diag(-1,p),c(rep(0,p))))
x<-rep(1,d)
sol<-matrix(1,d,d+p)
i<-1
while(i <= d)
{
DEAb3<-c(t(dis)[1:p,i],1)
x[i]<-linp(E = DEAA3, F = DEAb3, G = NULL, H = NULL, a,
ispos = TRUE, int.vec = NULL, verbose = TRUE)$solutionNorm #value of objective fun
sol[i,]<-linp(E = DEAA3, F = DEAb3, G = NULL, H = NULL, a,
ispos = TRUE, int.vec = NULL, verbose = TRUE)$X #value or argmax lambda and s
```

```
i<-i+1;
}
x
sol

########################################################
#79 districts, limsolve, 7 parameters


d=79; #settings
p=7; #settings
dis<-di[,c(1,2,4,5,6,7,9)] #settings
DEAa=c(rep(0, d),rep(1,p))
a <- DEAa*(-1)
DEAA3=cbind(rbind(t(dis),c(rep(1, d))),rbind(diag(-1,p),c(rep(0,p))))
x<-rep(1,d)
sol<-matrix(1,d,d+p)
i<-1
while(i <= d)
{
DEAb3<-c(t(dis)[1:p,i],1)
x[i]<-linp(E = DEAA3, F = DEAb3, G = NULL, H = NULL, a,
ispos = TRUE, int.vec = NULL, verbose = TRUE)$solutionNorm #value of objective fur
sol[i,]<-linp(E = DEAA3, F = DEAb3, G = NULL, H = NULL, a,
ispos = TRUE, int.vec = NULL, verbose = TRUE)$X #value or argmax lambda and s
i<-i+1;
}
x
sol


########################################################
```

```
#77 districts, limsolve, 3 parameters (HDI)

d=77;
p=3;
dis<-di[c(-26,-61),c(4,8,9)]
DEAa=c(rep(0, d),rep(1,p))
a <- DEAa*(-1)
DEAA3=cbind(rbind(t(dis),c(rep(1, d))),rbind(diag(-1,p),c(rep(0,p))))
x<-rep(1,d)
sol<-matrix(1,d,d+p)
i<-1
while(i <= d)
{
DEAb3<-c(t(dis)[1:p,i],1)
x[i]<-linp(E = DEAA3, F = DEAb3, G = NULL, H = NULL, a,
ispos = TRUE, int.vec = NULL, verbose = TRUE)$solutionNorm #value of objective fur
sol[i,]<-linp(E = DEAA3, F = DEAb3, G = NULL, H = NULL, a,
ispos = TRUE, int.vec = NULL, verbose = TRUE)$X #value or argmax lambda and s
i<-i+1;
}
x
sol
```

A.2 Notes to other tried R methods

- Package boot: sensitivity analysis is possible, however, equality constraints are not implemented.
- Package linprog: it is impossible to access the solution apart from whole model.

Appendix B

B Poverty indicator values for all 79 districts

Table 16: Normalized poverty indicators

Name	DP	SBC	IFI	LEM	LEW	RUR	SLJ	AMW	LOE
Bánovce nad Bebravou	0,70	0,12	0,99	0,52	0,31	0,23	0,22	0,80	0,67
Banská Bystrica	0,52	0,08	0,99	0,36	0,32	0,19	0,21	0,64	0,52
Banská Štiavnica	0,67	0,28	1,00	0,89	0,81	0,46	0,46	0,71	0,70
Bardejov	0,79	0,32	1,00	0,46	0,14	0,63	0,50	1,00	0,74
Bratislava I	0,01	0,01	0,00	0,03	0,13	0,00	0,07	0,40	0,00
Bratislava II	0,00	0,02	0,61	0,29	0,24	0,04	0,02	0,03	0,02
Bratislava III	0,10	0,00	0,90	0,25	0,26	0,02	0,04	0,23	0,15
Bratislava IV	0,12	0,00	0,91	0,00	0,00	0,02	0,00	0,10	0,05
Bratislava V	0,11	0,01	0,98	0,27	0,35	0,04	0,02	0,36	0,07
Brezno	0,73	0,29	1,00	0,80	0,57	0,51	0,51	0,58	0,63
Bytča	0,55	0,16	1,00	0,76	0,54	0,43	0,30	0,71	0,40
Čadca	0,65	0,10	1,00	0,94	0,63	0,28	0,26	0,89	0,72
Detva	0,63	0,34	1,00	0,81	0,47	0,59	0,54	0,61	0,70
Dolný Kubín	0,67	0,14	0,99	0,40	0,20	0,39	0,31	0,68	0,59
Dunajská Streda	0,48	0,15	0,99	0,54	0,58	0,22	0,28	0,67	0,45
Galanta	0,34	0,12	0,97	0,72	0,68	0,11	0,03	0,59	0,40
Gelnica	1,00	0,44	1,00	0,85	0,86	0,63	0,67	0,89	1,00
Hlohovec	0,40	0,11	0,97	0,48	0,15	0,16	0,15	0,41	0,43
Humenné	0,73	0,21	1,00	0,52	0,39	0,44	0,38	0,77	0,67
Ilava	0,47	0,12	0,99	0,51	0,42	0,20	0,09	0,58	0,51
Kežmarok	0,91	0,52	1,00	0,85	0,87	0,75	0,79	0,74	0,97
Komárno	0,75	0,23	1,00	0,71	0,81	0,40	0,36	0,84	0,63
Košice - okolie	0,80	0,46	1,00	0,80	0,73	0,62	0,71	0,62	0,89
Košice I	0,61	0,17	0,96	0,32	0,14	0,18	0,35	0,60	0,40

Table 16: Normalized poverty indicators

Name	DP	SBC	IFI	LEM	LEW	RUR	SLJ	AMW	LOE
Košice II	0,47	0,20	0,89	0,38	0,42	0,23	0,32	0,00	0,45
Košice III	0,47	0,17	0,99	0,42	0,57	0,21	0,30	NA	0,42
Košice IV	0,50	0,20	0,98	0,63	0,75	0,17	0,36	0,64	0,53
Krupina	0,80	0,42	1,00	0,92	0,85	0,64	0,54	0,70	0,75
Kysucké Nové Mesto	0,56	0,15	0,96	0,67	0,60	0,35	0,40	0,43	0,57
Levice	0,60	0,36	1,00	0,75	0,65	0,39	0,62	0,64	0,54
Levoča	0,92	0,46	1,00	0,43	0,79	0,46	0,63	0,88	0,85
Liptovský Mikuláš	0,52	0,16	0,99	0,50	0,18	0,27	0,36	0,59	0,42
Lučenec	0,74	0,56	1,00	0,77	0,70	0,67	0,69	0,80	0,75
Malacky	0,27	0,08	0,97	0,55	0,50	0,17	0,13	0,32	0,29
Martin	0,60	0,12	0,99	0,42	0,25	0,24	0,31	0,61	0,50
Medzilaborce	0,87	0,52	1,00	0,65	0,40	0,62	0,66	0,86	0,70
Michalovce	0,57	0,46	0,99	0,70	0,66	0,51	0,68	0,76	0,66
Myjava	0,56	0,09	0,99	0,46	0,31	0,33	0,10	0,65	0,51
Námestovo	0,57	0,09	1,00	0,85	0,55	0,35	0,20	0,79	0,60
Nitra	0,48	0,12	0,99	0,40	0,24	0,16	0,20	0,68	0,45
Nové Mesto nad Váhom	0,30	0,08	0,98	0,36	0,29	0,23	0,24	0,57	0,46
Nové Zámky	0,58	0,24	0,99	0,66	0,52	0,33	0,36	0,75	0,46
Partizánske	0,63	0,15	1,00	0,24	0,35	0,33	0,31	0,91	0,74
Pezinok	0,38	0,02	1,00	0,46	0,40	0,10	0,05	0,64	0,37
Piešťany	0,47	0,09	0,99	0,23	0,19	0,17	0,20	0,59	0,48
Poltár	0,98	0,53	1,00	0,69	0,76	0,68	0,64	0,78	0,90
Poprad	0,34	0,22	1,00	0,44	0,24	0,27	0,55	0,61	0,56
Považská Bystrica	0,55	0,12	0,99	0,55	0,27	0,35	0,25	0,58	0,56
Prešov	0,67	0,24	1,00	0,35	0,27	0,44	0,57	0,77	0,52
Prievidza	0,63	0,10	0,99	0,38	0,20	0,28	0,24	0,64	0,55
Púchov	0,45	0,07	0,96	0,60	0,50	0,19	0,22	0,67	0,40
Revúca	0,92	0,99	1,00	0,90	1,00	0,89	0,91	0,74	0,94

Table 16: Normalized poverty indicators

Name	DP	SBC	IFI	LEM	LEW	RUR	SLJ	AMW	LOE
Rimavská Sobota	0,90	1,00	1,00	0,88	0,88	1,00	1,00	0,85	0,87
Rožňava	0,78	0,80	0,99	0,92	0,75	0,82	0,80	0,63	0,85
Ružomberok	0,53	0,19	0,98	0,60	0,58	0,28	0,40	0,37	0,62
Sabinov	0,89	0,45	1,00	0,60	0,47	0,74	0,69	0,87	0,80
Šaľa	0,51	0,26	0,96	0,69	0,58	0,23	0,34	0,39	0,52
Senec	0,29	0,01	0,98	0,48	0,23	0,08	0,02	0,44	0,31
Senica	0,52	0,16	0,99	0,60	0,81	0,34	0,28	0,68	0,45
Skalica	0,46	0,19	0,95	0,60	0,42	0,29	0,24	0,49	0,41
Snina	0,88	0,40	1,00	0,71	0,35	0,64	0,42	0,95	0,68
Sobrance	0,93	0,58	1,00	0,88	0,52	0,59	0,70	NA	0,73
Spišská Nová Ves	0,71	0,40	0,99	0,54	0,53	0,44	0,55	0,73	0,66
Stará Ľubovňa	0,71	0,23	1,00	0,55	0,31	0,33	0,44	0,84	0,68
Stropkov	0,83	0,40	1,00	0,58	0,11	0,50	0,57	0,95	0,63
Svidník	0,81	0,40	1,00	0,44	0,32	0,62	0,56	0,96	0,66
Topoľčany	0,59	0,12	0,99	0,46	0,39	0,28	0,34	0,70	0,55
Trebišov	0,89	0,66	1,00	1,00	0,84	0,74	0,71	0,95	0,79
Trenčín	0,41	0,04	0,98	0,25	0,06	0,14	0,09	0,68	0,33
Trnava	0,20	0,06	0,94	0,39	0,22	0,12	0,06	0,34	0,30
Turčianske Teplice	0,66	0,21	1,00	0,67	0,56	0,30	0,44	0,80	0,68
Tvrdošín	0,58	0,08	0,99	0,46	0,20	0,34	0,19	0,80	0,60
Veľký Krtíš	0,93	0,59	1,00	0,86	0,65	0,69	0,73	0,86	0,96
Vranov nad Topľou	0,88	0,51	1,00	0,60	0,33	0,61	0,60	0,88	0,75
Žarnovica	0,56	0,28	0,98	0,57	0,72	0,61	0,55	0,53	0,57
Žiar nad Hronom	0,43	0,21	0,98	0,49	0,21	0,39	0,47	0,49	0,66
Žilina	0,29	0,08	0,96	0,51	0,31	0,15	0,19	0,44	0,32
Zlaté Moravce	0,72	0,15	0,98	0,63	0,27	0,29	0,31	0,64	0,60
Zvolen	0,48	0,19	0,99	0,47	0,14	0,22	0,43	0,59	0,48