UNIVERSITÀ DI PISA

DIPARTIMENTO DI INGEGNERIA DELL'ENERGIA DEI SISTEMI DEL TERRITORIO E DELLE COSTRUZIONI

RELAZIONE PER IL CONSEGUIMENTO DELLA LAUREA MAGISTRALE IN INGEGNERIA GESTIONALE

Municipal Waste Performance of NUTS-2 regions in the EU: A Data Envelopment Analysis model and convergence analysis

SINTESI

RELATORI

Prof. Ing. Gionata Carmignani

Dipartimento DESTEC

Dr. Paolo Sospiro
Director EUAbout

IL CANDIDATO

Alberto Fedele
a.fedele4@studenti.unipi.it

Sessione di Laurea Magistrale del 28/04/2021

Municipal Waste Performance of NUTS-2 regions in the EU: A Data Envelopment Analysis model and convergence analysis

Alberto Fedele

Sommario

Questo lavoro di ricerca è volto ad assegnare una valutazione sulla performance di efficienza regionale riguardo la gestione di Municipal Solid Waste con il fine di esaminare gli esiti delle direttive Europee sui rifiuti e monitorarne la loro efficacia. L'area di studio considerata è EU-27 a livello delle regioni NUTS-2. L'analisi del contesto legislativo ha permesso di individuare le condizioni sotto cui le varie regioni operano. Come dataset in totale sono state selezionate 167 NUTS-2 regions di 20 paesi. Per le regioni di 7 nazioni è stata necessaria una manipolazione dei dati, attraverso multiple imputation bootstrap-based EMB, per stimare i dati mancanti. I periodi presi in esame sono 2008-2013. La misurazione dell'efficienza è svolta tramite 4 modelli del metodo non parametrico Benefit-of-Doubt. La valutazione di efficienza è rappresentata dal Composite Indicator che in modo aggregato considera i principali waste treatments. Infine l'analisi sulla convergenza ha permesso di verificare la direzione intrapresa dalle regioni e il loro divario.

Abstract

This research assesses an efficiency evaluation, regarding the Municipal Solid Waste management, of the EU-27 area at NUTS-2 region level. The aim is to examine the impact of the EU waste directives and monitor their effectiveness. The legal context analysis allowed to understand the conditions that influence the regions. In total 167 NUTS-2 regions of 20 countries are considered as a dataset, for 7 countries was necessary a data manipulation, through multiple imputation bootstrap-based EMB, in order to estimate their missing data. The time span examined is the period 2008-2013. The efficiency assessment is done with 4 different models of Benefit-of-Doubt, a non-parametric approach. The measurement is presented by the Composite Indicators which consider the multi criteria of the main waste treatments. The further convergence analysis verified the regional direction taken and allowed to understand the differences.

1. Introduction

Since the beginning of the XX century waste starting to be a significant global issue. A more efficient use of assets is needed to transform waste into new resource. Notwithstanding EU efforts, a significant amount of potential secondary raw materials is still not exploited by the economy.

2. Legal context

The determination of waste in a legislative framework shall set a shared system of regulatory control on which policy makers and institutions must adapt. In the European Union (EU) the legal reference to the waste is the Waste Framework Directive (WFD) which include waste definition. Despite several revisions of the WFD the term waste remained defined in the same way as "… any substance or object which the holder discards or intends or is required to discard…". Not by every Member State (MS) was accepted the definition, applying instead their own national definition of waste leading to a divergence.

2.1. Waste classification

Identify the characteristics of a specific waste is crucial to apply its proper management. The waste is classified on its main three areas: Municipal Solid Waste (MSW), Industrial and Commercial (C&I), and Construction and Demolition (C&D). This regroup is done to avoid, as much as possible, the interconnection of waste data between the different categories.

2.2. Concept of MSW

Despite the MSW does not represent the main percentage of the total amount of waste produced (10% in 2018) is one of the most difficult categories of waste to manage, due not just to the sorting and treatment complexity but also to the high potential harm. Also the definition of MSW has lack of clarity, therefore, each Member State has freedom to interpret MSW, and to decide which waste types and waste sources to include and which are not in the definition of MSW. Mainly is waste produced by households but include also other sources that are similar to it, for nature and composition.

3. Legal waste framework

Through the years the EU put much more attention on the reduction of waste and on a safe and proper management. In legislative manner the EU amends higher level waste policy that each MS drafts in their own national and regional directives to create homogeneity on waste strategies and waste treatment. The Environment Action Programme (EAP) was the first directive that analysed the waste management as a matter to be controlled at European level.

The sixth EAP, first to be drafted through a MSs' joint decision, laid down precise goals to lead a shared sustainable development and to protect the environment reporting as main aim the reduction of waste and led to the draft of the 2008/98/EC (rWFD).

3.1. European Waste legislation

The rWFD is the first Directive which a global approach on the regulation regarding waste management that avoided the overlapping of different acts. General objectives are laid down and introduced the "waste hierarchy". This last one is a guiding principle, which the governments have to include in their waste management strategy, that define a priority order on waste management (prevention, reuse, recycle, recovery, disposal).

3.2. European Directives on waste management

Other important European Directives and Regulations are integrating specific aspects of waste management. The landfill Directive 1999/31/EC introduced stringent technical requirements for waste and landfills. The Directive 2000/76/EC on incineration waste has the aim to prevent or limit negative effects on the environment and risks to human health, from the incineration. The Directive 94/62/EC on Packaging and Packaging Waste sets out the rules on managing packaging and packaging waste to harmonise the measures.

4. Toward the Circular Economy

The CE appears to be an answer to the actual global challenges. A key point is to keep the resources in the economy for as long as possible, the transformation from waste to resources should bring solutions for waste management problems. The CE is based on a hierarchical approach known as 9R, a guideline similar to the "waste hierarchy" model, which represents the level of CE implementation leading to a life cycle thinking. The Green Deal is considered the main long term EU strategy to tackle the environmental challenges.

5. Data and sample

The area of interest for the study is the EU-27 area at NUTS-2 regions level. The main source of data is Eurostat even if presented missing data. EU data on MSW are available till the year 2013 which constitutes also the last period considered. The 2008 was selected as starting period, set as a reference for the amendment of the rWFD. The Eurostat dataset structure is taken as a reference to create the dataset used in the analysis. The Waste statistics in the EU are collected on the basis of a voluntary questionnaire with data furnished directly by each MS. The MSs have freedom to define their methods to collect the data under the concepts and formats defined centrally between Eurostat and the MSs. In the final dataset are considered 167 NUTS-2 regions of 20 countries, seven countries had to be excluded due to insufficient information to construct their reliable prediction. The data are classified by waste generation and by the five main treatment operations: landfill, incineration, energy recovery, recycling and composting.

5.1. Methodology

To predict the missing data is selected the Multiple Imputation approach to obtain a robust estimation with less statistical uncertainty. The specific method applied is Amelia which performs multiple imputation generating imputed data set. The algorithm is bootstrap based EMB(Expectation-Maximization with Bootstrapping), represented in Fig 1, that ensures robustness to impute many variables including cross sectional, time series data. It is an extension of the EM (Expectation-Maximization) algorithm, which alternates between an expectation (E-step) and maximization (M-step) steps until convergence. It is operating under the assumptions that each variable is considered with Multi-variate Normal Distribution (MVN) and the missing data are considered as missing at random (MAR). The multiple imputation obtained are combined using the univariate mean of the predicted variables.

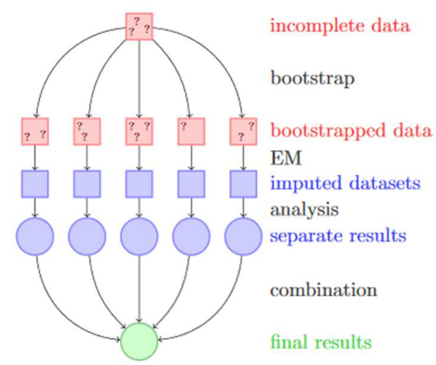

Fig. 1 Representation of EMB algorithm multiple imputation

5.2. Data elaboration

In the time span selected, the Eurostat dataset has a rate of missing data equal to 35.7% that forced to manipulate some regions and exclude the countries with more than 45% of total missing observations. A common setting, for each country imputed, is defined to obtain the most coherent imputations' results which consists in setting cross-section patterns over time in all variables, the data are considered as a time series, it is considered cross-sectional relation among regions, each variable is considered dependent on the previous period value, priors are added to increase the numerical stability, upper and lower bound are set to restrict the imputation range for each variable, number of imputations applied are equal to m=5. To validate the imputations it was used diagnostics tools: compare the densities distributions between observed data and imputed data, check the imputation model on the observed data through overimputation, check the convergence on the models imputation via overdispertion method. Common assumptions were taken: the waste generated correspond to the sum of all the waste treatments and that the national values correspond to the sum of the regional values for the specific period.

5.2.1. Regional Data manipulated

By the Eurostat Denmark is totally missed, had to be integrated with data furnished by the Danish National Statistical Institute, which constitute of MSW treatment at NUTS-3 level referred to the period 2013-2019. Before the imputation was applied a reclassified following the Eurostat structure. On the dataset created was applied the EMB-algorithm.

For France the data provided by Eurostat are just the periods 2008, 2010 and 2011. In order to obtain a reliable imputation through EMB algorithm it was done the manipulation twice, the first considering as a time span the years 2008-2011 and the second considering the entire timeframe using as input the result of the previous imputation.

The missing data on Germany, concerning the waste treatments of landfill, incineration and energy recovery, were completed using data provided by the German National Statistical Institute which reported the statistics collected from 2006 till 2017 in matter of MSW but based on a different collection method. On the data reclassified under Eurostat criteria was calculated the rates for the missing waste treatments, then used on the value of Eurostat waste generated to estimate the missing data. The residuals were fulfilled using the variable average calculated on the German dataset weighted on their standard deviation.

For Italy, in order to run the EMB algorithm the Italian data had to be integrated with data from ISPRA, assuming that the amount of recycling per capita at NUTS-2 level multiplied for the population can represent the regional MSW recycling treatment. The additional data were added just for the periods 2011-2013 because the ISPRA included in one value material and biological recycling and just for these years Eurostat reported the composting amount.

For the data of Poland on NUTS-2 regions, the dataset provided by Eurostat is missing just the variable referred to energy recovery for the periods between 2008 and 2011 and the EMB algorithm was run directly.

Concerning Romania, due to unsatisfying results obtained in the diagnostic analysis of the first computation, through an heuristic approach, a different imputation configuration was set. Was increased the priors to 5%, consider a polynomial relation of first grade on the time and shrunk the bounds' confidence level.

The Spanish data had to be integrated with data from the National Statistical Institute. Before to run the imputation the data were reclassified under Eurostat criteria. After the computation in few regions remained a residual value. The gap was filled through a redistribution based on the waste treatment mean rate calculated from the Spanish reports weighted on the CV.

6. Methodology

6.1. Data Envelopment Analysis (DEA)

The model applied to evaluate the efficiency is the DEA, used when there is no information on the exact functional relation between the variables. The DEA is an econometric linear programming (LP) approach that allows to identify the efficiency of each Decision Making Units (DMU) by measuring endogenously the relative efficiencies between themselves and to find peer units which can represent the good practices or benchmarking. It's a non-parametric approach that evaluates the efficiency requiring only minor assumptions. The simple restriction is that all the DMUs are considered belonging to a specific production frontier on which they lie on or below its efficient level. The DEA can handle multiple inputs and multiple outputs and aims to inputs minimisation to entail the decrease of inefficiency (if input oriented) or outputs maximisation to improve the efficiency (if output oriented); the choice of the specific model orientation is based on the factors' control. The final result of the DEA is a ranking of the DMUs based on the Composite Indicator (CI) measured which assesses a performance score to each DMU.

6.2. Selection of inputs and outputs

The CI assessed is built on five sub-indicators representing the waste treatments, in fact aggregates simple indicators which contain the numerical information on the relevant aspects to obtain a unique measurement. The simple indicators weightings determination is done through a relative perspective between the observed data which allows to evaluate the specific DMU in the best possible light in relation to the others. It is considered under control, by the policy makers, just the waste treatments and not the value of waste generation thus considered as an uncontrollable variable. The raw values of the five waste treatments are normalize on the waste generation value of the same region to deal with the rates and obtain comparable values. The five waste treatments rate are considered the outputs in the model and thus is used an output oriented approach.

6.3. Benefit-of-Doubt (BoD)

The model selected for the analysis is BoD technique that is based on the DEA methodology. It allows to capture the complexity and the multidimensionality of the MSW managements taking in consideration in the efficiency evaluation an involvement of multi criteria. The BoD-model is formally tantamount to the multiplier formulation of the DEA-model with all performance indicators treated as outputs and a 'dummy input' equal to one for all observations. In the specific case of the study the number of regions considered is equal to 187 and the simple indicator representing the main waste treatments are 5.

6.3.1. Basic BoD-model

The first model applied is the basic version of BoD, setting less constrain possible. Analytically, it is expressed by the LP:

$$CI_k = max_{w_{k,q}} \sum_{i=1}^{q} w_{k,i}\, y_{k,i} \qquad (1)$$

s.t.

$$\sum_{i=1}^{q} w_{k,i}\;\; y_{k,i} \leq 1 \qquad j = 1; \dots; k; \dots.; N$$

$$w_{k,1} \geq 0 \qquad \mathrm{i} = 1; \;\dots\dots; \mathrm{q}$$

This model is applied to each NUTS-2 region k, finding the optimal CI_k evaluation. The $w_{k,i}$ represent the most favourable weight for the region k regarding the waste treatment i; $y_{k,i}$ is the simple indicator of the observed performance relative to the waste treatment i of the region k. The only constraints set are the upper bound of the CI_k set up to 1 and the non-negativity of the weights assigned to each observed performance.

6.3.2. Constrained BoD-model

A constrained BoD-model is applied to face the problem of compensation between the simple indicators in the LP (1). In LP (2) is imposed an importance weight of at least 5% to each simple indicator and up to 80% to don't overrate the best results, in addiction the sum of the importance weights is set to 1.

$$CI_k = max_{w_{k,q}} \sum_{i=1}^{q} w_{k,i}\, y_{k,i} \qquad (2)$$

s.t.

$$\sum_{i=1}^{q} w_{k,i}\;\; y_{k,i} \leq 1 \qquad j = 1; \dots; k; \dots.; N$$

$$\sum_{i=1}^{q} w_{k,i} \;\; = 1 \qquad j = 1; \dots; k; \dots.; N$$

$$w_{k,i} \geq 0.05 \qquad \mathrm{i} = 1; \;\dots\dots; \mathrm{q}$$

$$w_{k,i} \leq 0.8 \qquad \mathrm{i} = 1; \;\dots\dots; \mathrm{q}$$

6.3.3. Bad BoD model

Before each output was considered desirable, for that reason the landfill treatment was considered inverse. In the bad BoD-model is introduced the presence of undesirable indicators. Moreover is added a set of restrictions which impose a rank ordering. To represent the undesirable factors are introduced in the model two more variables, the performance reported by the simple indicator $y_{k,s}^{-}$ and the relative weight assigned $v_{k,s}$ to make a distinguish with the desirable performance that is represented by the variable $y_{k,i}^{+}$.

$$CI_k = max_{w_{k,i};v_{k,s}} \sum_{i=1}^{r} w_{k,i} y_{k,i}^{+} - \sum_{s=1}^{l} v_{k,s} y_{k,s}^{-} \qquad (3)$$

s.t.

$$\sum_{i=1}^{r} w_{k,i}\ y_{k,i}^{+} - \sum_{s=1}^{l} v_{k,s} y_{k,s}^{-} \leq 1 \qquad j = 1; \dots; k; \dots.; N$$

$$\sum_{i=1}^{r} w_{k,i} + \sum_{s=1}^{l} v_{k,s} = 1 \qquad j = 1; \dots; k; \dots.; N$$

$$w_{k,i} \geq 0.05 \qquad \mathrm{i} = 1; \dots\dots; \mathrm{r}$$

$$v_{k,s} \geq 0.05 \qquad \mathrm{s} = 1; \dots\dots; \mathrm{l}$$

In the specific case the undesirable factor s is referred just to the landfill, so l is equal to 1

The ranking order set in based on the five stage level of the waste hierarchy.

$$v_{k,1} \geq w_{k,4};\ w_{k,4} \geq w_{k,3};\ w_{k,3} \geq w_{k,2};\ w_{k,2} \geq w_{k,1}$$

6.3.4. Directional Distance BoD-model

This version is built on the ideas of the directional distance function models, taking in account the existence of a preference structure including a directional penalty. The directional BoD-model assesses the performance evaluation on the output distance to the frontier in g-units that represents the exact direction in which improvements may be sought.

$$CI_k = max_{w_{k,i};v_{k,s}} \frac{1}{1+(\sum_{i=1}^{r} w_{k,i} y_{k,i}^{+} - \sum_{s=1}^{l} v_{k,s} y_{k,s}^{-})} \qquad (4)$$

s.t.

$$\sum_{i=1}^{r} w_{k,i}\ y_{k,i} - \sum_{s=1}^{l} v_{k,s} y_{k,s}^{-} \leq 1 \qquad j = 1; \dots; k; \dots.; N$$

$$\sum_{i=1}^{r} w_{k,i}\ g_i^{+} + \sum_{s=1}^{l} v_{k,s} g_s^{-} = 1 \qquad j = 1; \dots; k; \dots.; N$$

$$\frac{w_{k,i} g_i^+}{\sum_{i=1}^{r} w_{k,i}\, g_i^+ + \sum_{s=1}^{l} v_{k,s} g_s^-} \geq 0.05 \qquad \mathrm{i} = 1;\ \ldots\ldots;\mathrm{r}$$

$$\frac{v_{k,s} g_s^-}{\sum_{i=1}^{r} w_{k,i}\, g_i^+ + \sum_{s=1}^{l} v_{k,s} g_s^-} \geq 0.05 \qquad \mathrm{s} = 1;\ \ldots\ldots;\mathrm{l}$$

$$w_{k,i} \geq 0 \qquad \mathrm{i} = 1;\ \ldots\ldots;\mathrm{r}$$

$$v_{k,s} \geq 0 \qquad \mathrm{s} = 1;\ \ldots\ldots;\mathrm{l}$$

In order to obtain more reliable results it is applied a robust order-m framework for the LP (3) and (4) to mitigate the presence of outliers and have a more robust model after the introduction of further assumptions. This is done by executing a Monte Carlo simulation procedure in which B iterations (b = 1,…,B) are performed in which on sub-sample of m regions drawn is applied the LP. The final CI score is the average value of the B CI calculated.

6.4. Advantages and disadvantages of the DEA in our application

The results obtained by the BoD-model should be harder to contest by the single actors because the setting doesn't define any preference for specific subjects due to its non-parametric way and using the optimal weights in the regional evaluation.

Including some assumptions in the model is a trade-off between guide towards a better reflection of the actual situation and leave the flexibility in the self-determination. In matter on the weight determination, each region is evaluated in the most favorable way, the DMUs have different weight importance level creating problems in the comparisons. Exogenous factors are not considered in the evaluation but in the real situation could influence in the measurement of performance. Moreover the approach evaluates the different regions between themselves that could be possible include some improper confrontation.

6.5. Convergence methods

Convergence methods are applied to understand the regional disparities and evaluate the dispersion trend. It is decided to apply the approaches of σ-convergence and β-convergence, these models are selected because have a respective mutual relationship and can extent their outputs. The β-convergence allows to understand the reduction in disparities across European regions, an absolute β-convergence model is used the ordinary least squares (OLS) estimation. The model is considered as linear but not non-linear in terms of variables, in order to represent that the treatments with higher values should be more difficult to increase. It is decided to apply an estimation through the log-log relationship. The final OLS absolute model applied to evaluate the presence of the β-convergence is

$$\frac{1}{T}\sum_{t=1}^{T}\ln\left(\frac{Y_{k,t+1}}{Y_{k,t}}\right) = \ln\beta_0 + \beta_1 \ln X_{k,t_0} + \ln U_{k,t_0}$$

Which in $Y_{k,t}$ represents the growth rate of the waste treatment under consideration for the region k in the period t; β_0 and β_1 are the parameters to be estimated; $X_{k,t}$ represents the initial level in the period t of the waste treatment for the region k; $U_{k,t}$ represents the random residuals.

β_1 represents an estimation of elasticity and part of the condition of convergence $\beta_1 \leq 0$.

The second analysis applied is the σ-convergence approach to determine the standard dispersion deviation. In fact, even if results $\beta_1 \leq 0$ from the β-convergence, it's not guaranteed a decrease in dispersion. There is convergence when differentiation between regions decrease over time, here was measured the dispersion by the coefficient of variation. The model should be represented by

$$cv = \alpha_0 + \alpha_1 t + U_t$$

α_0 *and* α_1 are the parameters to be estimated, to exist the σ convergence should result $\alpha_1 < 0$.

7. Results and general conclusions

As expected different BoD-models brought to different results. In fact the LP (1) and LP (2) are the more flexible but for some regions the evaluation should not reflect properly the actual situation. On the other side the LP (3) aimed to reflect the rWFD guidelines but the first priority scale represented is the third layer of the waste hierarchy because there isn't any measurement for prevention and reuse. Regarding the LP (4) the direction was set by the author based on the MSW EU targets defined by the directives (2008/98/EC set 65% for recycling and 1999/31/EC set a max landfill rate to 10%) but not for every waste treatments is set a defined targets. Despite the different results obtained, with a scatterplot representation of the models' ranking and the relative values of Pearson's correlation, is detected a linear relation between the models. The descriptive statistics analysis of CI values shows a growing trend, which could be a condition to a regional performance convergence towards better performances. Larger room for improving were registered in the early periods. Is not demonstrated yet the best fit and so each LP is considered. To examinate the best performers and identify a benchmark is considered the range of CI assessed in the periods believing that the exact overall CI lie in this range. The CI national deviation decrease from the yearly average in the time span, the decremental of dispersion can be interpreted

as an evidence in favor of convergence. The convergence analysis was applied considering two different settlement, first selecting each waste treatments separated and second aggregate the best waste treatment (recycling and composting) according to the waste hierarchy level of priority. The most consistent convergence is registered in the second case demonstrating the presence of β-convergence and σ-convergence statistically significant.
It is noticed that, even in the same country, the regions are performing differently. The best performers in terms of waste treatments are localized in Northern and Central European regions along all the periods analysed, as represents the geographic distribution of average regional rank (Fig.2). However the worst performers are mostly in the Eastern Europe.

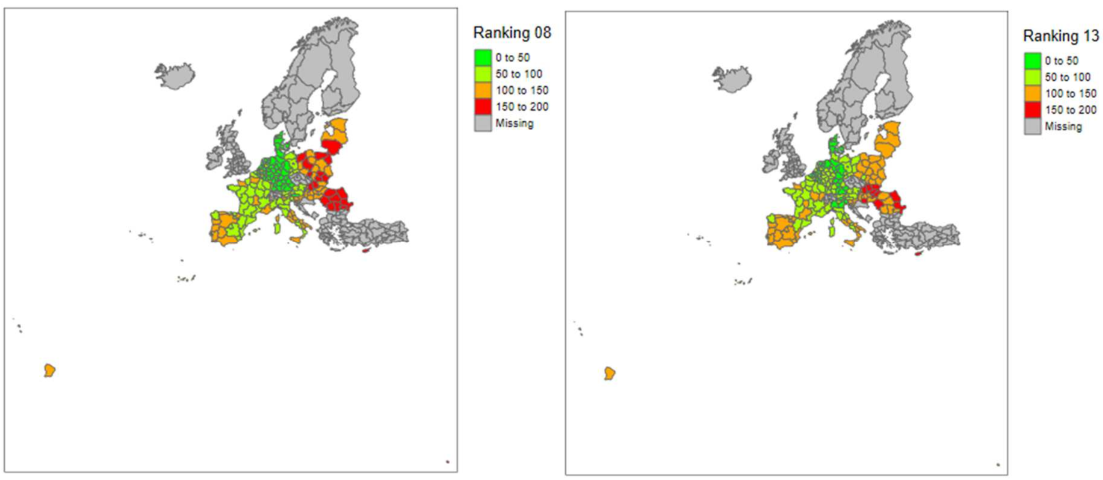

Fig.2 Geographic distribution of NUTS-2 overall regional ranking for the first and last period.

The best improvements are achieved by the regions that moved more away from landfilling, however the worst countries remained those whom still rely mainly on the landfill. Such huge division between regions on the landfill rate is suppose rely mainly on two reasons, the national waste dispositions period of introduction and the year of entrance in the EU. In relation to the timeframe analysed, it is noticed a general convergence towards a better waste treatment concerning MSW between the European NUTS-2 regions. One fact that is assumed influenced the latter convergence is the entry into force of the rWFD. However the convergence level registered is too slow to achieve the EU CE targets. The EU shall develop a common form of regulation to avoid commercial confusion and to have a reliable monitor and evaluation of the MSs. The next steps of the study can be foreseen as the application of the conditional BoD-models. In conclusion the study is considered as a beginning and the results obtained need to be confirmed by additional research.