

Marijana Petrović^{1*}, Nataša Bojković¹, Mladen Stamenković², Ivan Anić³¹ University of Belgrade, Faculty of Transport and Traffic Engineering, Serbia² University of Belgrade, Faculty of Economics, Serbia³ ITS-Information Technology School, Belgrade, Serbia

A Sensitivity Analysis of ELECTRE Based Stepwise Benchmarking for Policy: the Case of EU Digital Agenda

DOI: 10.7595/management.fon.2018.0003

Abstract: Research question: This paper introduces a new parameter for evaluating the sensitivity of development paths in ELECTRE (*ELimination Et Choix Traduisant la Réalité*) based stepwise benchmarking model. **Motivation:** The aim of the research was to devise a 'path sensitivity coefficient' which measures the sensitivity of a development path depending on parameters set by the decision-maker (DM). The need for such a measure comes from the fact that the hierarchical preorder of benchmarking units is affected by the DM's subjective decisions on the threshold for declaring outranking relations among them. Consequently, this may disrupt the sequence of intermediate benchmarks - transitional targets to be followed on the route to the ultimate policy goals. The proposed approach is an extension of a stepwise benchmarking procedure devised by Petrovic et al. (2014) which evaluates development paths based on their gradualism. **Idea:** In this paper, we propose to characterize development paths by considering both gradualism and sensitivity to outranking thresholds. The standpoint of our research is that in order to make a reliable final choice the DM should be aware that the most gradual path is not necessarily the most stable one. **Tools:** The proposed 'path sensitivity coefficient' combines four levels of uncertainty coming from two threshold values set by the DM. The coefficient has a relative nature, it points to the "distance" from the theoretically ideal solution - a path fully indifferent to threshold values. An approach to merge coefficients associated with gradualism and stability is also introduced. **Data:** To illustrate the proposed approach we applied it in the field of digital transformation of the EU countries. We exploited indicators and data from Digital Agenda Scoreboard regarding specific policy goals to be achieved until 2020. **Findings:** The findings imply that the decision on the most gradual development path should be reconsidered in regards to the influence of model parameters. The final choice depends on whether the DM prefers the smoothness of the path or the overall stability in terms of underlying a partial preorder. **Contribution:** This work contributes to the existing ELECTRE-based stepwise benchmarking by reducing the uncertainty coming from subjective parameterization and allows the DM to be more confident in the final path selection.

Keywords: stepwise benchmarking, ELECTRE, multiple criteria decision aiding, sensitivity analysis, digital agenda targets

JEL Classification: C44, L96, L 98, O57

1. Introduction

Traditionally, benchmarking is seen as a systematic and continuous process of comparing performances of similar organizations or processes, in order to learn from the best performers and thereby improve one's own performance. As a tool for continuous improvement, it gained great popularity during the nineties of the last century, after the success of the Xerox Corporation (Bogan and English, 1994). As explained in detail by

* Corresponding author: Marijana Petrović, E-mail: marijanatpetrovic@gmail.com

Bhutta and Hug (1999) by their benchmarking the company's performance against the practice of its Japanese affiliates, the Xerox Company managed to lower its product unit manufacturing cost by 70 percent (by cutting its suppliers base, reducing the number of defective components, reassigning a number of inspectors to other jobs, cutting the component lead-time, reducing costs of purchased parts etc.) In time, benchmarking has expanded its possible fields of application. A major breakthrough was its migration from private to public sector and the increasing application in the field of public policy (Bessant and Rush, 1998; Lundvall and Tomlinson, 2001; Bruno et al., 2006; Friedewald et al. 2005). The reason underpinning the application of benchmarking in the policy domain is to inform, explain and justify actions of government institutions and organizations in the eyes of the public (Bessant and Rush, 1998; Fagerber, 2001; Papaioannou et. al, 2006; Dominique et al., 2013). Another, but no less important contribution to policy, is that benchmarking enables the transfer of knowledge through finding good practices and their transfer to the decision-making system. Learning from each other is a principal aim of benchmarking that goes with a common conceptual framework for policy modelling known as *policy transfer* and related concepts (terms) *cross-national learning* and *lesson-drawing* (Dolowitz and Marsh, 1996, 2000; Dolowitz, 2003; James & Lodge, 2003; Evans, 2004). In the core of all of them is a process of comparing one to the other and improving by learning from each other. That is why benchmarking is the best way to operationalize the policy transfer. Perhaps the most valuable confirmation of this is the well-known concept of governance in the European Union (EU) – the so called “Open method of coordination-OMC.” The OMC concept was introduced by the European Council of Lisbon in March 2000, with the aim to stimulate member states to learn from sharing their experience of national policies in areas of common interest (Radaelli, 2003). Along with OMC, benchmarking was promoted as a policy tool, as a way to underpin policy transfer based on cross-national learning. This has triggered a series of benchmarking studies in a policy context and opened a debate among scholars about the true contribution of the OMC and benchmarking to policy making (Porte et al., 2001; Radaelli, 2003; Arrowsmith, 2004, Dominique et al., 2013 etc.). The first one is the decision on benchmarking partner-units (mostly countries) that will be engaged in the process. As argued by several authors (Rose, 1991, Bauer, 2009, Petrovic et al., 2012), in order to exploit international benchmarking for policy learning and transfer the countries included should share common policy goals and have international organizational/institutional support (such as the EU countries addressed in this study). The other issue is setting a proper indicator set that reflects policy aspects that are to be benchmarked. For this reason, policy agendas are often followed by extensive work on indicators reflecting progress towards established policy goals (as Digital Agenda Scoreboard exploited in the empirical example of this study). One of the concerns that are in the very essence of benchmarking is the question of a benchmark – the one to look up to i.e., the one whose performance should be emulated. As discussed by Alstete (2008), many benchmarking studies focus solely on ranking leaving out the detection of a benchmark as a next legitimate step of a benchmarking process. A stream of research deals with this issue within a stepwise benchmarking concept (see Petrovic et al., 2017 for a review). Obtaining the development or benchmarking paths for each less successful benchmarking unit makes the essence of stepwise benchmarking. This path is represented by a set of intermediate benchmarks (performance targets) which mark the improvement steps along the route to the ultimate target (policy agenda goals). One idea to operationalize stepwise benchmarking is to use a multi-level outranking approach based on ELECTRE (ELimination Et Choix Traduisant la REalité) as described and applied in public policy domain in Petrovic et al. (2012, 2014, 2017) and Stamenkovic et al. (2016). In the centre of this approach is the overcoming of certain negative aspects of common analytical tools like DEA (Data Envelopment Analysis) or CIs (Composite indices) approach (for comprehensive discussion see Petrovic et al., 2013b). The main concern regarding the value of the model is the question of parameters defined by the decision-maker (DM) i.e., how sensitive the solution is to the DMs' preferences. The focus of this paper is to introduce a new measure – ‘path sensitivity coefficient’ that will help eliminate the uncertainty coming from subjective parameterization and make the decision-maker more confident in the results.

The paper reads as follows. The next section is about benchmarking in the public policy context and stepwise approach. The methodological background is given in Section 3. Section 4 problematizes the sensitivity of the model output and introduces sensitivity parameters. This is followed by an empirical example – benchmarking EU countries towards key Digital Agenda goals. The paper ends with concluding remarks and future research directions.

2. Benchmarking in Public Policy Context and the Need for Stepwise Approach

The main reason that underpins using benchmarking in public policy context is to operationalize the policy transfer based on cross-national learning. The very essence of benchmarking enables a systematic comparison of performance with the aim to find the best practice exemplar and to learn from it. As elaborated by Papaioannou et al.(2006), several classifications of benchmarking recognize policy as the field of

benchmarking application. For example, Huggins (2010) points to *policy benchmarking* as a separate type along with performance and process benchmarking.

Once benchmarking was promoted as a policy tool at the EU level, a number of benchmarking studies, projects and initiatives have been conducted resulting in benchmarking being now more depicted by its practical use than by the management theory. This made establishing the relationship between benchmarking and policy an even more complex task. Relying on the experience in sustainable transport policy domain Gudmundsson et al. (2005) offered an approach to classify the ways in which benchmarking can be implemented in the policy domain. They singled out five basic types (illustrations in Figure 1). The explanations and application examples from the telecommunications/ICT policy is offered by Petrovic et al. (2013a) and summarized in text boxes in Figure 1 (Acronyms in Figure 1 are for major EU benchmarking projects: SIBIS - Statistical Indicators Benchmarking Information Society; BISER - Benchmarking the Information Society in European Regions; UNDERSTAND - 'European Regions UNDER way towards STANDard indicators for benchmarking information society'; TRANSFORM - Benchmarking and Fostering TRANSFORMative Use of ICT in the EU Regions; WSIS-World Summit on Information Society).

Although widely accepted as a public policy tool, the application of benchmarking itself is associated with many issues and challenges. By far the most elaborated concern is its prevailing quantitative nature (for comprehensive discussions see Petrovic et al, 2013a). Policy modelling based on quantitative data leads to simplification of complex policy phenomena. Namely, policy creators often focus on data and numbers instead of on the reasons lying behind the differences (Elmuti and Kathawala, 1997; Arrowsmith et al., 2004; Codagnone et al., 2015). The consequence is that benchmarking loses its essence and becomes a tool for comparing (a kind of a "check list") rather than a tool for learning. Another and associated issue is the discussion on the term benchmark itself: does it represent an etalon for comparing (target performance to be achieved Spendoliny, 1992; McNair and Leibreid, 1994; Zairi, 1994) or is a benchmark a best practice exemplar – the one to look up to.

In the policy context, benchmark is often related to the targets defined in policy agendas/action plans (as key performance targets defined in the EU Digital Agenda). As the one to look up to, the term benchmark refers to the standard of excellence, best practice exemplar, one whose performance is desired and a strategy that should be emulated. The issue of learning from the best is also extensively discussed in terms of policy. Namely, as seen by some authors (see for example Bauer, 2010; Maheshwari et al., 2013; Petrovic et al., 2013a, 2014, 2017), learning from the best may lead to setting unrealistic targets and unachievable policy goals. That is why decision-makers opt for the concept of learning from the similar, tracking corresponding benchmarks and/or some intermediate targets on the route to the ultimate goal. This refers to the process of stepwise benchmarking, i.e., gradual improvements towards performance targets (see Estrada et al., 2009 Park et al., 2012; Petrovic et al., 2014, 2017; Fang, 2015).

The suitability of stepwise benchmarking approach for public policy created the need to define methods to operationalize it. In terms of analytical approach the concepts of corresponding benchmarks, benchmarking paths and stepwise improvements are usually addressed using the linear programming method, Data Envelopment Analysis – DEA, a method first proposed by Charnes, Cooper and Rhodes (1978) (Hong et al., 1999; Seiford and Zhu, 1999; Park et al., al, 2012; Lim et al., 2011; Estrada et al., 2009; Ghahraman and Prior, 2016). However, a number of limitations induced the need to define new tools based on multi-criteria decision aiding - MCDA. These limitations are associated with restricted scope of application imposed by DEA application requirements – the need for input/output orientation of the data, selecting the best path on each level instead of overall path evaluation, infeasibility when working with theoretical maximum (like targets set by policy agendas), etc. (for further discussion see Petrovic et al., 2013b, 2014, 2017).

Among MCDA methods, the ELECTRE multi-level outranking (ELECTRE MLO) appears to be a promising tool for stepwise benchmarking operationalization, although it is not the only MCDA method used in the context of benchmarking (see for example Laise, 2004 who used ELECTRE I or Jovanovic and Delibasic, 2014 who exploited AHP).

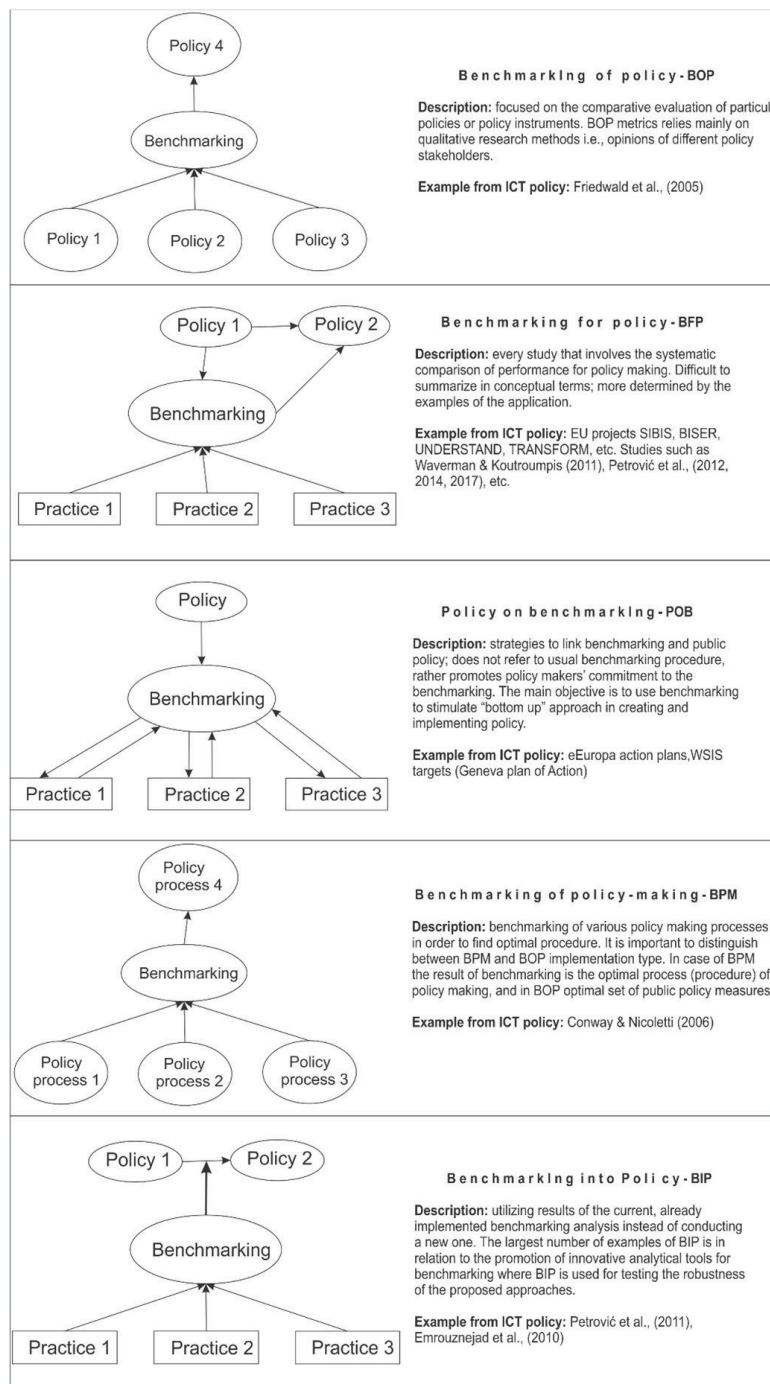


Figure 1: Five different possible roles of benchmarking in relation to public policy with examples from the field of telecommunications/ICT (according to Gudmundsson et al., 2005 and Petrovic et al., 2013a)

3. ELECTRE based Model for Stepwise Benchmarking in Public Policy Domain

According to the progress regarding the desired policy targets, the ELECTRE MLO produces hierarchical levels of performance. It also provides the pairwise relations between benchmarking units, which reveals better performing ones, termed as "intermediate benchmarks". The sequence of intermediate benchmarks constitutes the development path. The specific procedure of the model allows for the selection of the optimal path among the multiple choices.

As well as other methods of the ELECTRE family, the ELECTRE based multi-level outranking method – ELECTRE MLO is based on pairwise comparisons of units under observation (alternatives) for each criterion and the computation of concordance and discordance indices. The concordance index C_{ij} denotes the degree of dominance of one unit over another. To construct hierarchical levels of performance, the modified concordance index is defined in the following form (Anic and Larichev, 1996; Bojkovic et al., 2010).

$$C_{ij} = \frac{\sum_{k \in K_{ij}^+} \omega_k}{\sum_{k \in K_{ij}^+ \cup K_{ij}^-} \omega_k}, \tag{1}$$

In equation (1), ω_k is the weight of criteria k , while the sets K_{ij}^+ and K_{ij}^- are those where the performance score of the alternative A_i is higher (K_{ij}^+) and lower (K_{ij}^-) than the score of the alternative A_j .

The discordance index d_{ij} prevents the declaration of dominance if an alternative is significantly bad against a certain criterion. It is expressed as

$$d_{ij} = \max_{k \in K_{ij}^-} \left((a_{jk} - a_{ik}) / |I_k| \right). \tag{2}$$

where a_{jk} and a_{ik} are the scores for alternatives A_j and A_i with respect to criterion k , and $|I_k|$ is the scaled-score range of criterion k .

The assertion “alternative A_i outperforms alternative A_j ”, (denoted as $A_i \succ A_j$) is valid if the values of the concordance and discordance indices C_{ij} and d_{ij} are above and below predefined threshold values p and q , respectively.

The alternative A_i appears at a certain performance level L_n when the following equation holds:

$$\max \left\{ t \mid (\exists A_j \in L_t) A_i \prec A_j \right\} = n - 1. \tag{3}$$

This means that A_i is at the level of performance L_n if there is at least one alternative located at the level L_{n-1} that outperforms A_i . The hierarchical structure of the model (visualized in the form of a ‘relation tree’) allows the decision maker to have a number of development paths at disposal. These paths are labelled as evolutionary development paths and consist of a sequence of alternatives at the higher levels, called “intermediate benchmarks”. A procedure for paths’ evaluation according to their gradualism or smoothness is proposed in Petrovic et al. (2014) and further explored in Petrovic et al. (2017). It is based on coefficient named *relative path variation* ρ_i^π which measures the distance of each possible path from the most gradual one - where each step presumes equal performance gain (see Petrovic et al., 2014, 2017 for detailed explanation and calculations).

4. Path Sensitivity Analysis of the Stepwise Benchmarking Model

Parameterization of MCDA-based models is an ever-present issue, with many research directions. An emerging stream of research is based on indirect preference elicitation procedures where parameters of the model are defined without asking the decision maker to explicitly assign such values. This approach requires less cognitive effort and is more user-friendly given that the decision maker, although familiar with the problem, usually knows little about the mathematical formalization of the model (Corrente et al., 2013). Such approach yielded several methods, with robust ordinal regression being the most popular one (Corrente et al., 2013; Kadzinski et al., 2016; Ciomek et al., 2018). This method sorts or ranks alternatives

in a *learning by example* sense (Greco et al., 2010). It is within stepwise benchmarking modelling that indirect elicitation of parameters was introduced by Petrovic et al. (2017).

Regardless of the way in which the parameters are obtained, many MCDA based studies sought to investigate the sensitivity and robustness analysis of the proposed solution(s) (e.g., &, 1997; Tervonen et al., 2009; Roy, 2010). Kadzinski et al. (2016) designed the methodology to evaluate the recommendation of the completed robust ordinal regression. They explained several potential directions to approach the robustness issue, such as defining the range for the possible rank of alternatives (Beynon and Barton, 2008) or by indicating various intensity measures of the dominance relation within the method (Ahn and Park, 2008). The importance of this line of research is not solely related to MCDA methods and it can be seen as a state of the art question in academia recently (for example ARWU ranking sensitivity analysis has been of interest in a number of studies such as Dobrota & Dobrota, 2016; Maricic et al., 2017; Zornic et al., 2014).

The aim of this paper is to offer an approach in a similar fashion. We propose a methodological extension of the stepwise benchmarking model in order to present to the decision maker the level of robustness of the model output. Our work is about threshold values p and q that are subjectively set by the DM and reflect his tolerance in declaring dominance relations. As such, they are subject to change, whereby it should be known to what extent these changes affect the model output, namely development paths.

In this paper, we introduce a measure that will reflect how *far* the selected path from ideally insensitive one is. We label this measure as *path sensitivity coefficient* (σ_π). It combines four levels of uncertainty coming from concordance (p) and discordance (q) threshold values.

An alternative can be *pq-insensitive*, meaning that changes of both parameters p and q will not significantly affect the position of an alternative in the relation tree. On the other hand, the alternative can be *p-sensitive* if the change of parameter p creates a significant alteration of alternatives' position in the relation tree, while parameter q does not have the same influence. We can describe the *q-sensitivity* in a similar way. Finally, the alternative will be considered *pq-sensitive* if both threshold changes impose significant changes in the alternative's position in the relation tree. With S_p we will denote all alternatives on a given path that are only *p-sensitive*, and analogously we define S_q as well as S_{pq} , the set of alternatives that are both *p-sensitive* and *q-sensitive*.

For further description, let the path π contain alternatives $A_i = A_{i_1}, A_{i_2}, \dots, A_{i_m}$ where A_{i_m} is an alternative from the top level (core alternative) and let $A_i^\pi = \{A_{i_1}, \dots, A_{i_m}\}$ be a set of alternatives from the path π and not in the core. In order to evaluate the sensitivity in the ELECTRE MLO, we depict a normalized coefficient σ_π representing the measure of path sensitivity. We define σ_π in the following manner:

$$\sigma_i^\pi = \frac{0.5|A_i^\pi \cap S_p| + 0.5|A_i^\pi \cap S_q| + |A_i^\pi \cap S_{pq}|}{m-1} \tag{4}$$

path sensitivity coefficient (σ_π) can take values from 0 to 1. Path sensitivity coefficient is also a relative measure meaning that the closer σ_π is to value 0, it is less sensitive to the DM's preferences, i.e., the path is more robust and thus more preferable. By this equation, we measure the level of sensitivity in the model, whereas we give twice as much significance to the pq-sensitivity compared to p and q-sensitivity. The parameter for the measure of path sensitivity will take value 0 when there are no sensitive elements alongside the path while the value 1 will represent the path in which all alternatives in the path are pq-sensitive.

Now, the decision maker is provided with two coefficients that can assist the decision maker in making the final decision of a stepwise benchmarking path. He can consider both parameters independently or he can

weigh the importance of one relative to the other. In that case, the coefficients may be merged into a new one, φ_i^π in following way:

$$\varphi_i^\pi = \alpha \rho_i^\pi + (1 - \alpha) \sigma_i^\pi . \quad (5)$$

If the *smoothness* of the path is preferred over the *stability*, α will be higher than 0.5 and vice versa.

Two coefficients offered address two important features of a development path, its 'gradualism', i.e., the balanced improvement steps along the whole path and 'stability', i.e., insensitivity to the parameters of the model. As demonstrated later in the paper (within the empirical example) the most gradual path is not necessarily the most stable one. Given the fact, the output is affected by the way the model is calibrated, it is up to model designers to make the decision maker aware of this influence. Namely, the standpoint of our approach is that the decision maker should be provided with both information in order to make a reliable choice.

5. Empirical Example – the Case of EU Digital Agenda Policy Goals

In this section, we illustrate how sensitivity coefficient improves the decision on optimal development path within a benchmarking exercise of EU countries according to Digital agenda targets. Digital Agenda for Europe (DAE, 2010) is the first of seven flagships initiatives in , the EU's strategy to bring smart sustainable and inclusive growth. The DAE was launched in May 2010 and includes 101 actions, grouped around seven priority areas. Besides, the Digital Agenda contains which are to be achieved. These are:

1. the entire EU to be covered by broadband;
2. the entire EU to be covered by broadband above 30 Mbps;
3. 50 % of the EU to subscribe to broadband above 100 Mbps;
4. 50 % of the population to buy online;
5. 20 % of the population to buy online cross-border;
6. 33 % of SMEs to make online sales/purchases;
7. the difference between roaming and national tariffs to approach zero;
8. to increase regular internet usage from 60 % to 75 % by 2015, and from 41 % to 60 % among disadvantaged people;
9. to halve the proportion of the population that has never used the internet from 30 % to 15 %;
10. 50 % of citizens to use eGovernment by 2015, with more than half returning completed forms;
11. all key cross-border public services, to be agreed by the Member States in 2011, to be available online;
12. to double public investment in ICT R&D to € 11bn;
13. to reduce energy use of lighting by 20%,

All goals are quantitative by nature and summarize the so-called digital transformation of the EU countries. Progress against these targets is measured annually through the.² This is a specific data base for public policy creators with different tools that can be used but without any signposts on intermediate targets that can support moving step-by-step to the target values. That is why there is a need for a model that will enable to consider all targets within one model and identify possible transitional goals. The relation tree as an outcome of specific and sophisticated ranking scheme and suggestion on development paths for each underperforming country can serve this purpose. The relation tree reflecting the progress of 27 EU countries towards the DAE key specific goals in 2012 is presented in Figure 2 (for details on benchmarking units, indicators, scores, and thresholds, see Petrovic et al., 2014). Note that 11 indicators reflecting 9 key performance targets are used to obtain the relation tree in Figure 2.

* Corresponding author: Marijana Petrović, E-mail: marijanatpetrovic@gmail.com

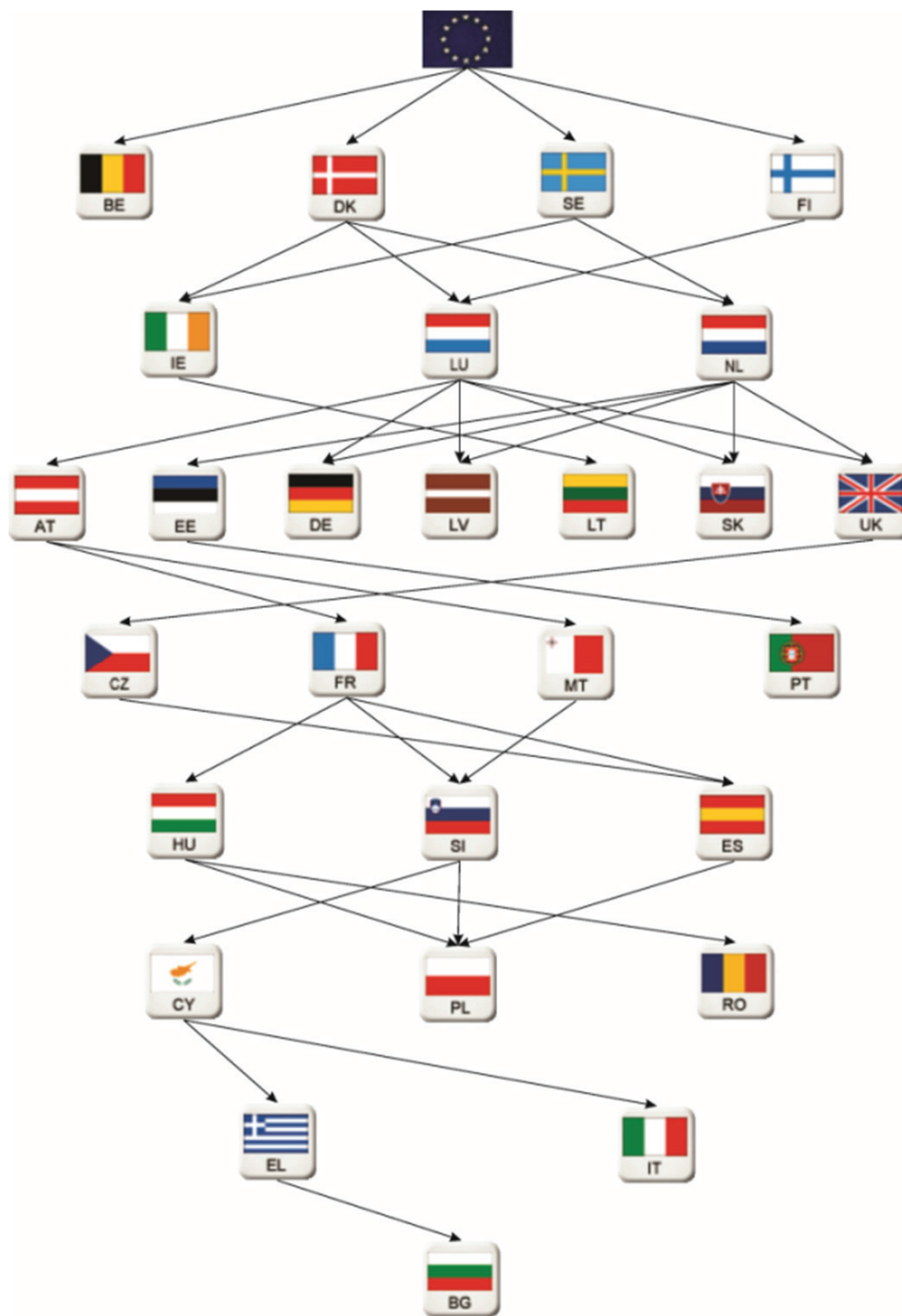


Figure 2: Relation tree (Adapted from Petrovic et al., 2014)

There are nine levels of performance (Figure 2), whereby the first level represents the desired targets to be achieved by the EU as a whole as well as by individual countries. The countries from the neighboring levels are connected with arrows which represent the outranking relations and indicate corresponding benchmarks. Countries that are positioned in an adjacent, higher level are potential policy benchmarks for those on the lower level if they are connected with the outranking/dominance relations. On the other hand, the lack of relation indicates a strong contrast between adjacent level countries and they are therefore not considered as suitable sources for cross-national (policy) learning.

Possible development paths are depicted for Poland, a country on the seventh level of performance, and presented in Table 1. We have opted for Poland to be an illustrative example due to the largest possible number of development paths in the example.

Table 1: Possible development paths for Poland and associated measures

	Development path	ρ_i^π	σ_π
1	POL-SPA-CZE-UK-NED-SWE	0.48	0.40
2	POL-SPA-FRA-AUT-LUX-DEN	0.57	0.20
3	POL-HUN-FRA-AUT-LUX-FIN	0.58	0.10
4	POL-SPA-CZE-UK-NED-DEN	0.58	0.40
5	POL-SPA-FRA-AUT-LUX-FIN	0.60	0.20
6	POL-HUN-FRA-AUT-LUX-DEN	0.61	0.10
7	POL-SPA-CZE-UK-LUX-DEN	0.62	0.40
8	POL-SPA-CZE-UK-LUX-FIN	0.62	0.40
9	POL-SLO-FRA-AUT-LUX-DEN	0.62	0.10
10	POL-SLO-MLT-AUT-LUX-DEN	0.62	0.20
11	POL-SLO-MLT-AUT-LUX-FIN	0.66	0.20
12	POL-SLO-FRA-AUT-LUX-FIN	0.66	0.10

The DM has gained 12 different development paths obtained from the ELECTRE MLO stepwise benchmarking procedure (Table 1). A large number of possibilities invokes possible uncertainty for the decision maker and sensitivity analysis is fully observed in this case. According to the value of relative path variation that represents the level of gradualism of the improvement, the set of intermediate policy benchmarks to be followed consists of four countries – Spain, the Czech Republic, the United Kingdom, the Netherlands, and Sweden. At the same time, this path gained the highest value of the sensitivity coefficient

σ_π in Poland's path selection, which calls for the DM's further appraisal. The final decision will depend on the DM's view of the importance of the path gradualism over stability (expressed by parameter α in equation (5)). If he decides that gradualism is of three or more times higher importance than stability ($\alpha \geq 0.75$), then the above mentioned path will be his final choice. If he, however, decides to give equal importance to both path characteristics (to set $\alpha = 0.5$), then parameter φ_i^π defined from the equation (5) will yield that the best path for Poland is path number 3 consisting of Hungary, France, Austria, Luxembourg and Finland.

Conclusion

This paper addresses the problem of development path's stability within the existing stepwise benchmarking model. It introduces sensitivity coefficient as the measure of sensitivity analysis of the concordance and discordance threshold values. By mitigating fluctuations imposed by threshold values that are subject to change and errors, the decision-maker can make the final choice on intermediate targets towards ultimate policy goal. As illustrated in the empirical example of the Digital Agenda specific policy targets, the sensitivity coefficient can be considered in conjunction with relative path variation, thus allowing the policy creators to better understand the status of the solution.

The weak point of the ELECTRE based stepwise benchmarking model is its full dependency on a partial preorder created by underlying outranking method - ELECTRE MLO. In this paper, we have addressed this issue and introduced sensitivity parameter aimed at resolving the problem.

The ELECTRE MLO provides a partial preorder as its final output based on synthesizing preference relations among alternatives observed through a concordance and discordance index. The decision maker is requested to define the cutting level p and veto threshold q to define the outranking relation. However, since the decision maker is not aware of the sensitivity of the output coming from these two parameters, in this study we have introduced sensitivity coefficient which helps the DM to evaluate the stability of each solution (proposed development path). Namely, the decision maker now has two path features addressed – the relative path variation and the path sensitivity coefficient devoted now both to gradualism and stability. The DM can combine these two parameters into the final decision. We proposed an aggregate measure in the paper as a function of the DM's preferences. The DM can decide, according to problem characteristics, whether he/she prefers smoothness of the path or the overall stability in terms of underlying a partial preorder. The important feature of the sensitivity coefficient is in its relative nature. It means that it does not only rank development paths but also points to the "distance" from the theoretically ideal solution - a path fully indifferent to threshold values. It should be noted that the full scope of sensitivity analysis would include another condition that affect the results of the stepwise benchmarking model.

One issue that may change the hierarchical levels of performance is the addition or exclusion of the benchmarking units/alternatives in the model. This is to be the subject of future research.

Through this procedure, we have excluded the possibility of a sensible solution in terms of path stability. However, the question of rank reversal and possible change of partial preorder after an alteration of within alternatives has not been resolved within stepwise benchmarking methodology and it will be investigated in our future research.

Regarding some future empirical applications, the 'path sensitivity coefficient' should be included in stepwise benchmarking procedure based on the new Digital Agenda Scoreboard indicators (those that constitute the new EU Digital Economy and Society index - DESI). Some investigations have been made regarding one dimension - connectivity (Petrovic et al., 2017) but without sensitivity analysis.

Future research should also aim at compiling a data pre-processing procedure, i.e., scoring based on the DM's performance appraisal introduced in Petrovic et al. (2017) with the sensitivity coefficient presented in this paper. This would improve the quality of the model output but also impose a specific research challenge – how to avoid the information overload, i.e., exposing the decision maker to too many coefficients of path quality.

Acknowledgements

This work was supported by the Ministry for Education and Science of the Republic of Serbia within the projects "Critical infrastructure management for sustainable development in postal, communication and railway sector of the Republic of Serbia", Project number: TR36022, 2011-2017 and "The risks of financial institutions and markets in Serbia - microeconomic and macroeconomic approach", Project number: 179005.

REFERENCES

- [1] Ahn, B. S., & Park, K. S. (2008). Comparing methods for multiattribute decision making with ordinal weights. *Computers & Operations Research*, 35(5), 1660-1670. DOI:10.1016/j.cor.2006.09.026
- [2] Alstete, J. W. (2008). Measurement benchmarks or "real" benchmarking? An examination of current perspectives. *Benchmarking: An International Journal*, 15(2), 178-186. DOI:10.1108/14635770810864884
- [3] Anic, I., & Larichev, O. (1996). The ELECTRE method and the problem of acyclic relation between alternatives. *Automation and Remote Control*, 8, 108-118.
- [4] Arrowsmith, J., Sisson, K., & Marginson, P. (2004). What can 'benchmarking' offer the open method of co-ordination?. *Journal of European Public Policy*, 11(2), 311-328. DOI: 10.1080/1350176042000194458
- [5] Bauer, J. M. (2010). Learning from each other: promises and pitfalls of benchmarking in communications policy. *info*, 12(6), 8-20. DOI: 10.1108/14636691011086017
- [6] Bessant, J., Rush, H. (1998) *Approaches to benchmarking: the case of 'framework conditions' and ICT-Os*, Working paper prepared for the Institute of Prospective Technological Studies, European Commission.
- [7] Beynon, M. J., & Barton, H. (2008). A PROMETHEE based uncertainty analysis of UK police force performance rank improvement. *International Journal of Society Systems Science*, 1(2), 176-193. DOI:10.1504/IJSS.2008.021918
- [8] Bhutta, K. S., & Huq, F. (1999). Benchmarking—best practices: an integrated approach. *Benchmarking: An International Journal*, 6(3), 254-268. DOI:10.1108/14635779910289261
- [9] Bogan, E., English, M.J. (1994) *Benchmarking for Best Practices: Winning Through Innovative Adaptation*, New York: McGraw-Hill, Inc.
- [10] Bojkovic, N., Anic, I., & Pejic-Tarle, S. (2010). One solution for cross-country transport-sustainability evaluation using a modified ELECTRE method. *Ecological Economics*, 69(5), 1176-1186. DOI:10.1016/j.ecolecon.2010.01.006
- [11] Bruno, I., Jacquot, S., & Mandin, L. (2006). Europeanization through its instrumentation: benchmarking, mainstreaming and the open method of co-ordination... toolbox or Pandora's box?. *Journal of European public policy*, 13(4), 519-536. DOI:10.1080/13501760600693895
- [12] Ciomek, K., Ferretti, V., & Kadzinski, M. (2018). Predictive analytics and disused railways requalification: Insights from a Post Factum Analysis perspective. *Decision Support Systems*, 105, 34-51.
- [13] Charnes, A., Cooper W. W., & Rhodes, E. L. (1978). Measuring the Efficiency of Decision Making Units. *European Journal of Operational Research*, 2(6), 429-444. DOI:10.1016/0377-2217(78)90138-8
- [14] Codagnone, C., Misuraca, G., Savoldelli, A., & Lupiañez-Villanueva, F. (2015). Institutional isomorphism, policy networks, and the analytical depreciation of measurement indicators: The case of the EU e-government benchmarking. *Telecommunications Policy*, 39(3), 305-319. DOI:10.1016/j.telpol.2015.01.005

- [15] Conway, P., & Nicoletti, G. (2006). Product market regulation in the non-manufacturing sectors of OECD countries. DOI:10.1787/18151973
- [16] Corrente, S., Greco, S., Kadzinski, M., & Słowiński, R. (2013). Robust ordinal regression in preference learning and ranking. *Machine Learning*, 93(2-3), 381-422. DOI:10.1007/s10994-013-5365-4
- [17] DAE (2010)-Digital Agenda for Europe, European Commission, Brussels, 26.8.2010, COM(2010) 245 final/2, 2010, Retrieved from: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2010:0245:FIN:EN:PDF> (February, 2016).
- [18] Dobrota, M., & Dobrota, M. (2016). ARWU ranking uncertainty and sensitivity: What if the award factor was Excluded? *Journal of the Association for Information Science and Technology*, 67(2), 480-482. DOI: 10.1002/asi.23527
- [19] Dolowitz, D. Marsh, D. (2000) Learning from Abroad: the Role of Policy Transfer in Contemporary Policy-Making', *Governance*, 13 (1), p. 5–24. DOI: 10.1111/0952-1895.00121
- [20] Dolowitz, D. (2003). A policy maker's guide to policy transfer. *The Political Quarterly*, 74(1), 101–108. DOI: 10.1111/1467-923X.t01-1-00517
- [21] Dolowitz, D., & Marsh, D. (1996). Who learns what from whom: a review of the policy transfer literature. *Political studies*, 44(2), 343-357. DOI: 10.1111/j.1467-9248.1996.tb00334.x
- [22] Dominique, K. C., Malik, A. A., & Remoquillo-Jenni, V. (2013). International benchmarking: Politics and policy. *Science and Public Policy*, 40(4), 504-513. DOI: 10.1093/scipol/scs128
- [23] Elmuti, D., & Kathawala, Y. (1997). An overview of benchmarking process: a tool for continuous improvement and competitive advantage. *Benchmarking for Quality Management & Technology*, 4(4), 229-243. DOI:10.1108/14635779710195087
- [24] Emrouznejad, A., Cabanda, E., & Gholami, R. (2010). An alternative measure of the ICT-Opportunity Index. *Information & management*, 47(4), 246-254. DOI:10.1016/j.im.2010.04.002
- [25] Estrada, S. A., Song, H. S., Kim, Y. A., Namn, S. H., & Kang, S. C. (2009). A method of stepwise benchmarking for inefficient DMUs based on the proximity-based target selection. *Expert Systems with Applications*, 36(9), 11595-11604. DOI:10.1016/j.eswa.2009.03.035
- [26] Evans, M. (Ed.) (2004) *Policy Transfer in a Global Perspective*, Ashgate, Aldershot
- [27] Fagerberg, J. (2001). Benchmarking: A new and useful tool for policy learning?. *Working Papers on Innovation Studies*, 20010621 Retrieved from: http://folk.uio.no/janf/downloadable_papers/fagerberg_benchmarking.pdf
- [28] Fang, L. (2015). Centralized resource allocation based on efficiency analysis for step-by-step improvement paths. *Omega*, 51, 24-28. DOI:10.1016/j.omega.2014.09.003
- [29] Friedewald, M., Kimpeler, S., & Hawkins, R. (2005). Benchmarking national and regional policies in support of the competitiveness of the ICT sector in the EU. : Fraunhofer IRB Verlag, 2005. Retrieved from: http://ec.europa.eu/enterprise/sectors/ict/files/ict-pol-finrep_en.pdf
- [30] Ghahraman, A., & Prior, D. (2016). A learning ladder toward efficiency: Proposing network-based stepwise benchmark selection. *Omega*, 63, 83-93. DOI:10.1016/j.omega.2015.10.004
- [31] Greco, S., Słowiński, R., Figueira, J., & Mousseau, V. (2010). Robust ordinal regression. Trends in multiple criteria decision analysis, 142, 241-283. DOI 10.1007/978-1-4419-5904-1
- [32] Gudmundsson, H. Wyatt, A., Gordon, L. (2005). Benchmarking and Sustainable Transport Policy: Learning from the BEST Network, *Transport Reviews*, 25(6), p. 669–690. DOI:10.1080/01441640500414824
- [33] Hong, H. K., Ha, S. H., Shin, C. K., Park, S. C., & Kim, S. H. (1999). Evaluating the efficiency of system integration projects using data envelopment analysis (DEA) and machine learning. *Expert Systems with Applications*, 16(3), 283-296. DOI:10.1016/S0957-4174(98)00077-3
- [34] Huggins, R. (2010). Regional competitive intelligence: benchmarking and policy-making. *Regional Studies*, 44(5), 639-658. DOI:10.1080/00343400802331312
- [35] James, O., & Lodge, M. (2003). The limitations of 'policy transfer' and 'lesson drawing' for public policy research. *Political studies review*, 1(2), 179-193. DOI: 10.1111/1478-9299.t01-1-00003
- [36] Jovanovic, B., & Delibasic, B. (2014). Application of integrated QFD and fuzzy AHP approach in selection of suppliers. *Management: Journal of Sustainable Business and Management Solutions in Emerging Economies*, 19(72), 25-35. DOI: 10.7595/management.fon.2014.0018
- [37] Laise, D. (2004). Benchmarking and learning organizations: ranking methods to identify "best in class". *Benchmarking: An International Journal*, 11(6), 621-630. DOI:10.1108/14635770410566528
- [38] Kadzinski, M., Ciomek, K., Rychly, P., & Słowiński, R. (2016). Post factum analysis for robust multiple criteria ranking and sorting. *Journal of Global Optimization*, 65(3), 531-562. DOI:10.1007/s10898-015-0359-3
- [39] Lim, S., Bae, H., & Lee, L. H. (2011). A study on the selection of benchmarking paths in DEA. *Expert Systems with Applications*, 38(6), 7665-7673. DOI:10.1016/j.eswa.2010.12.148
- [40] Lundvall, B-A., Tomlinson, M. (2001). Learning-by-comparing: reflections on the use and abuse of international benchmarking. In Sweeney G. (Ed.), *Innovation, Economic Progress and Quality of Life*, Cheltenham: Edward Elgar. p.203-231.

- [41] Maheshwari, D., & Janssen, M. (2013). Measurement and benchmarking foundations: Providing support to organizations in their development and growth using dashboards. *Government Information Quarterly*, 30, S83-S93. DOI:10.1016/j.giq.2012.11.002
- [42] Maricic, M., Zornic, N., Pilcevic, I., & Dacic-Pilcevic, A. (2017). ARWU vs. Alternative ARWU Ranking: What are the Consequences for Lower Ranked Universities?. *Management: Journal of Sustainable Business and Management Solutions in Emerging Economies*, 22(1), 1-14. DOI:10.7595/management.fon.2017.0002
- [43] May, P. J. (1992). Policy learning and failure. *Journal of public policy*, 12(04), 331-354.
- [44] McNair, C.J. and Leibfried, K.H.J. (1992), *Benchmarking: A Tool for Continuous Improvement*, HarperCollins Publishers, New York, NY.
- [45] Papaioannou, T., Rush, H., Bessant, J. (2006) Benchmarking as a policy-making tool: from the private to the public sector, *Science and Public Policy*, 33(2), p.91–102. <https://doi.org/10.31DOI:52/147154306781779091>
- [46] Park, J., Bae, H., & Lim, S. (2012). A DEA-based method of stepwise benchmark target selection with preference, direction and similarity criteria. *International Journal of Innovative Computing, Information and Control*, 8(8), 5821-5834.
- [47] Petrovic, M., Bojkovic, N., Anic, I., & Petrovic, D. (2012). Benchmarking the digital divide using a multi-level outranking framework: Evidence from EBRD countries of operation. *Government Information Quarterly*, 29(4), 597-607. DOI:10.1016/j.giq.2012.05.008
- [48] Petrovic, M., Bojkovic, N., Anic, I., Stamenkovic, M., & Tarle, S. P. (2014). An ELECTRE-based decision aid tool for stepwise benchmarking: An application over EU Digital Agenda targets. *Decision Support Systems*, 59, 230-241. DOI:10.1016/j.dss.2013.12.002
- [49] Petrovic, M., Bojkovic, N., Stamenkovic, M., & Anic, I. (2017). Supporting performance appraisal in ELECTRE based stepwise benchmarking model. *Omega*. In Press, DOI:10.1016/j.omega.2017.07.002
- [50] Petrovic, M., Tarle, S. P., Gospić, N. (2013a). *Benchmarking and telecommunications policy*, Faculty of Transport and Traffic Engineering, Belgrade, Serbia (In Serbian).
- [51] Petrovic, M., Tarle, S. P., Anic, I. (2013b). Pitfalls and promises of analytical tools for benchmarking, Proc. of The 7th International Working Conference
- [52] Porte, C. D. L., Pochet, P., & Room, B. G. (2001). Social benchmarking, policy making and new governance in the EU. *Journal of European Social Policy*, 11(4), 291-307.
- [53] Radaelli, C. M. (2003). *The Open Method of Coordination: A new governance architecture for the European Union?*. Swedish Institute for European Policy Studies.
- [54] Rose, R. (1991). What is lesson-drawing? *Journal of Public Policy*, 11(1), 3–30. DOI: 10.1017/S0143814X00004918
- [55] Roy, B. (2010). Robustness in operational research and decision aiding: A multi-faceted issue. *European Journal of Operational Research*, 200(3), 629-638. DOI:10.1016/j.ejor.2008.12.036
- [56] Spendolini, M. J. (1992). *The benchmarking book*. New York: American Management Association
- [57] Stamenkovic, M., Anic, I., Petrovic, M., & Bojkovic, N (2016). An ELECTRE approach for evaluating secondary education profiles: evidence from PISA survey in Serbia. *Annals of Operations Research*, 245:337-358. DOI:10.1007/s10479-015-1823-7
- [58] Tervonen, T., Figueira, J. R., Lahdelma, R., Dias, J. A., & Salminen, P. (2009). A stochastic method for robustness analysis in sorting problems. *European Journal of Operational Research*, 192(1), 236-242. DOI:10.1016/j.ejor.2007.09.008
- [59] Triantaphyllou, E., & Sanchez, A. (1997). A sensitivity analysis approach for some deterministic multi criteria decision making methods. *Decision Sciences*, 28(1), 151-194. DOI: 10.1111/j.1540-5915.1997.tb01306.x
- [60] Waverman, L., & Koutroumpis, P. (2011). Benchmarking telecoms regulation—the telecommunications regulatory governance index (TRGI). *Telecommunications Policy*, 35(5), 450-468. DOI: 10.1016/j.telpol.2011.03.006
- [61] Zairi, M. (1994). Benchmarking: the best tool for measuring competitiveness. *Benchmarking for Quality Management & Technology*, 1(1), 11-24. DOI: 10.1108/14635779410056859
- [62] Zornic, N., Markovic, A., & Jeremic, V. (2014). How the top 500 ARWU can provide a misleading rank. *Journal of the Association for Information Science and Technology*, 65(6), 1303-1304. DOI: 10.1002/asi.23207

Received: 2017-10-12

Accepted: 2018-01-29



About the Author

Marijana Petrović

University of Belgrade, Faculty of Transport and Traffic Engineering, Serbia
marijanatpetrovic@gmail.com

Dr Marijana PETROVIĆ is an Associate Professor at the Faculty of Transport and Traffic Engineering, University of Belgrade, Serbia. Her research areas are engineering management and policy in transport and communications, with special interest in policy modelling and quality management with the application of mathematical models and software solutions.



Nataša Bojković

University of Belgrade, Faculty of Transport and Traffic Engineering, Serbia
nbojkovic@yahoo.com

Dr Nataša BOJKOVIĆ is an Associate Professor at the Faculty of Transport and Traffic Engineering, University of Belgrade, Serbia. Her field of interest includes transport sustainability, urban mobility, policy modelling, management methods and decision support systems in transport and communications.



Mladen Stamenković

University of Belgrade, Faculty of Economics, Serbia
mladen@ekof.bg.ac.rs

Dr Mladen STAMENKOVIĆ is an Assistant Professor of Economic and mathematical methods and models (Operational Research) at the Faculty of Economics, University of Belgrade, Serbia. His research area is multiple criteria decision aiding with applications in economic policy.



Ivan Anić

ITS-Information Technology School, Belgrade, Serbia
ivan.anic@its.edu.rs

Dr Ivan ANIĆ works at the ITS-Information Technology School, Belgrade, Serbia. His area of research is multi-criteria decision making with applications. He especially deals with outranking methods and supporting software solutions. He has published a number of scientific papers in this area.

