MULTIPLE CRITERIA DECISION MAKING
Vol. 8 2013

Mateusz Zawisza[*]
Bogumił Kamiński[*]
Wit Jakuczun[*]
Agnieszka Gładysz[**]

COMPOSITE EVALUATION OF BROADBAND INTERNET ACCESS IN POLAND

Abstract

The level of access to Internet is constantly evaluated and promoted by electronic communications regulators around the world. The issue is especially important in countries, such as Poland, where where Internet access is highly heterogenous among local markets. The objective of this paper is to identify socio-economic factors that influence the level of Internet access in local communities (gminas) in Poland.

The definition of Internet access involves multiple criteria and encompasses in particular its availability, adoption, speed, quality of service and price. In the paper we propose a two-phase approach to a comparison of Internet access in various gminas. First we use Data Envelopment Analysis (DEA) to evaluate Internet broadband access depending on their demographic characteristics based on data from 2010 and 2011 collected by Poland's Office of Electronic Communications (UKE). In the second stage we explain the obtained DEA effectiveness indices using supervised learning techniques with the socio-economic status of the community as explanatory variables. We show that in the period under study rural communities experienced greater Internet access improvement than urban communities, therefore catching up with large cities and reducing technological gap. Moreover, we identify drivers of broadband Internet advancement, including: community type, community education and age structure, computerization level in schools and Herfindahl-Hirschman

[*]Warsaw School of Economics, Institute of Econometrics,al . Niepodległości 162, 02-554 Warsaw, e-mails: *mz35109@sgh.waw.pl* (M. Zawisza), *bkamins@sgh.waw.pl* (B. Kamiński), *wit.jakuczun@gmail.com* (W. Jakuczun)
[**]Urząd Komunikacji Elektronicznej, Kasprzaka 18/20, 01-211 Warszawa, e-mail: *A.Gladysz@uke.gov.pl*

competition index. We show that an effective regulation may foster the advancement of fixed-location broadband Internet access.

1 Introduction

The deployment of broadband technology brings to the society versatile economic and social benefits by its positive influence on (1) Gross Domestic Product (see Czernich et al., 2011; Holt and Jamison, 2009; Koutroumpis, 2009), (2) employment and job creation (see Katz, 2009; Katz et al., 2010), (3) research & development sector (see OECD, 2008), (4) reduction of business costs through e.g. cloud computing (see Zhang et al., 2010), (5) retail, services, manufacturing & industrial sectors (see Fornefeld et al., 2008), (6) education sector through e.g. building human capital and (7) health care sector. Broadband Internet is regarded as a general purpose technology (see Bresnahan and Trajtenberg, 1995), which, similarly to the invention of electricity, engines or railways, fundamentally changed the way an economic activity is now organized. General purpose technology is characterized by its pervasive use in a wide range of sectors bringing about vast productivity gains and enabling new opportunities across an entire economy (see Kelly and Rossotto, 2012). Therefore, fostering the development of broadband technology is regarded as an important policy making objective of many governments around the world. In particular, National Regulatory Authorities (NRAs) are dedicated to support the development of broadband infrastructure and regulation.

Due to the regional diversity of market competition, service accessibility and market penetration, NRAs face the problem of addressing their regulatory objectives and policies taking into account local market heterogeneity. The variety of factors describing the level of market development (service availability, adoption, speed, quality of service and price) makes it difficult to determine which local markets perform better than others. Therefore, we apply Data Envelopment Analysis (DEA) to enable the Polish NRA (UKE – Urząd Komunikacji Elektronicznej) to assign an efficiency score to every region and calculate the distance between them. This can be done without the need of defining the weights of variables a priori which is the disadvantage of many other methods dedicated to the assessment of broadband advancement (see Bayasyan et al., 2011; Commission of EC, 2008; Grubesic, 2010; ITU, 2009; TechNet, 2003).

According to the Polish NRA, the level of competition, as well as the level of demand on the broadband access market, are distributed very unequally throughout the country (see Gaj, 2012). Various market specific characteristics call for different approaches towards market regulation and the allocation of funds. Therefore, UKE concentrates on geographically differentiated regulation scheme and makes it one of its main priorities in its three-year strategic plan. It states that the "identification of areas with unsatisfied demand for fixed line services and broadband internet access is crucial for consumer-oriented policy consisting in improving service accessibility in rural areas and widening the choice of service providers in urbanized areas" (see Gaj, 2012). Applying multi-criteria DEA models enables such an identification and narrows the number of regions used for further, more detailed analysis. Other DEA applications, important from the NRA's point of view, may include setting precise and feasible aims concerning the development of local markets through comparing communities with similar input values. Moreover, operators and third party investors may find the DEA results obtained for local markets valuable and interesting as they reflect the effectiveness of regulatory actions in individual regions and are comparable in time. Therefore, DEA models have many potential applications because of their versatility and simplicity in interpreting the outcomes.

In the literature there are many applications of DEA method to (1) telecommunications markets (see Badasyan et al., 2011; Fernandez-Menendez et al., 2009; Giokas and Pentzaropoulos, 2008; Grubesic, 2010; Kang, 2009; Lam and Lam, 2005; Lam and Shiu, 2008; Sastry, 2009; Tsai et al., 2006; Uri, 2003; Zhu, 2004) and (2) the evaluation of regional economic competitiveness in the UK by creating the omposite competitiveness index (see Huggins, 2003), (3) the evaluation of countries with respect to Gross Domestic Product (see Growiec, 2012) and (4) the performance measurement of banks and public companies (see Brockett et al., 1997; Ho and Zhu, 2004; Zhu, 2000). However, none of these papers tries to explain the DEA score by control variables. Therefore, we propose a second step consisting in modeling DEA scores by control variables with the help of supervised learning techniques. A similar approach is taken with respect to the assessment of local governments by Alfonso and Fernandes (2008). The authors use a Tobit regression to explain DEA indices.

In the paper we propose a two-step procedure for the performance evaluation of regions according to specified criteria and explaining differences in performance among regions by a supervised learning technique. In the first step, we apply Data Envelopment Analysis to assess relative technological advancement of fixed location broadband Internet deployment in individual communities. In the second step, we apply supervised learning techniques (regression trees and random forests) to explain and predict differences in DEA scores depending on socio-economic characteristics of local markets. We show the possibility of effective welfare-improving regulation by influencing some of the explanatory variables that are of partial control of regulator and might foster the deployment of broadband Internet. The novelty of the paper consists in (1) applying the proposed two-step procedure to the data concerning telecommunications local markets in the period of 2010-2011 acquired by Polish telecommunications regulator (UKE) and (2) proposing a new analytic approach based on explaining DEA scores by control variables with the help of supervised learning techniques.

The article is organized as follows. In Section 2 we present sources and structure of data. In Section 3 we describe the proposed two-step procedure consisting of Data Envelopment Analysis part (Section 3.1) and supervised learning approach (Section 3.2). Results of this two-step procedure are presented in Section 4. All figures and computations are performed in the statistical programming environment GNU R (see R Core Team, 2012).

2 Broadband Access Data Characteristic

The employed data are obtained from the following dataset sources:

1. Local Markets Dataset of Poland's Office of Electronic Communications (UKE) from 2010 and 2011.
2. Local Data Bank of Poland’s Central Statistical Office (GUS) from 2010 and 2011.
3. GUS data covering average income in communities from 2010 and 2011.
4. National Census of Population and Housing 2002 provided by GUS.
5. Financial budgets of local governments provided by the Agency of Public Information (BIP) of Poland's Ministry of Finance from 2010.

Local Markets Dataset of UKE allows to obtain data concerning the situation on the market of fixed-location broadband Internet from 2010 and 2011. In particular, the dataset covers the number of subscribers (adoption) and the number of households with ability to subscribe to broadband (availability). The dataset contains also information on fixed-location broadband technology, bundled products and the infrastructure ownership. The UKE data allow to draw conclusions as to the broadband availability, adoption and competition. Based on those data we calculate additionally the Herfindahl-Hirschman Index[1] (HHI) which is a measure of the competition level on the local market. Each data observation describes one local market (gmina). There are 2 462 gminas in the dataset out of 2 479, grouped by three types as follows: 304 urban communities, 599 urban-rural communities and 1 559 rural communities.

The data concerning socio-economic characteristics of communities are derived from the following sources: (1) Local Data Bank of GUS, (2) National Census of Population and Housing 2002 of GUS and (3) GUS data on demand on average income in communities. Local Data Bank provides statistics of population size in communities, as well as age structure of population. There is no available data on current population education structure on the level of communities. However, we find out that the education structure in communities

[1] Herfindahl-Hirschman Index is the sum of squared market shares, i.e. where is the market share of company operating on the market with competitors. The maximum value of HHI is and indicates the monopoly market. The lower HHI (the minimum bound is 0), the more competitive market and higher number of competitors with no significant market power. The usage of Herhindal-Hirschman Index in the evaluation of local markets’ competition performance is presented in Zawisza and Kamiński (2012).

from 2002 is a good proxy for the education structure in 2011. Therefore, we make use of the National Census of Population and Housing 2002 to acquire information about the education structure in 2002. Moreover, GUS Local Data Bank is the source for data concerning computerization level in schools. It enables us to calculate the number of computers with broadband Internet access per pupil in primary and secondary schools in each community. We claim that high level of access to Internet in schools might be an important demand factor influencing the household decisions whether or not to subscribe to broadband. Another potential driver for broadband demand might be the wealth of community citizens, which is measured in our paper by the average income in community. This statistics is calculated on demand by GUS.

The data concerning local government expenditures on telecommunication services and infrastructures is obtained from the BIP of Poland's Ministry of Finance. The data consist of quarterly budgets of communities. In our analysis we calculate the sum of expenditures in 2010.

3 Internet access analysis procedure

The proposed procedure consists of two steps. In the first step, we assign a performance index to each local market with regard to its fixed-location broadband advancement depending on its demographic characteristic. We obtain these indices by using Data Envelopment Analysis (DEA) technique (see Charnes et al., 1978; Cooper et al, 2006; Guzik, 2009; Zhu, 2009). In the second step, DEA indices are explained by socio-economic variables. Understanding the mechanism of the influence of socio-economic variables on broadband advancement might be of great use for a regulator to tailor its policies according to recommendations provided by the model. The modeling part is done with the use of supervised learning techniques, in particular regression trees and random forests (see Hastie et al., 2009; Kamiński and Zawisza, 2012; Koronacki and Ćwik, 2008; Walesiak and Gatnar, 2009).

3.1 Data Envelopment Analysis step

The aim of the Data Envelopment Analysis (DEA) step is to provide a single index describing the level of fixed-location broadband technological advancement for each community. The DEA efficiency index takes into account three criteria[2]:

1. Availability of fixed location broadband refers to the ability of a household to subscribe to broadband via at least one fixed-location broadband technology regardless of whether the household actually subscribes to it or not. It is measured by the number of households with at least one broadband provider.

[2] These criteria are the outputs of DEA method, i.e. variables whose high values are desirable.

2. Adoption of fixed location broadband refers to the actual use of Internet and is measured by the number of households that are subscribed to fixed location broadband services.
3. Competition of fixed location broadband market refers to the competition pressure put by the providers on each other, which has a direct impact on terms and conditions of services offered, e.g. price level, transfer speed, download limits. It is measured by the number of households with at least two broadband service providers.

Since the above three criteria are expressed in absolute terms, it is important to normalize them. Otherwise, larger communities would have higher DEA index values. Therefore, we include the following variables:

- community population size,
- the number of households in community

as DEA inputs[3]. These inputs are objective demographic characteristics of the local market and so they can be treated as exogenous and beyond the control of telecommunications regulator or any government in short or medium term. The aim of the first step of the procedure is to return DEA scores that capture only the objective and technological aspect of fixed location broadband Internet deployment. We do not include any other variables as inputs, since they would unnecessarily justify a poor deployment of broadband Internet in some communities. The explanation of DEA scores is provided in the second step of the procedure proposed.

As the result of DEA method, we assign a single effectiveness index to each local market. The index is normalized between 0% and 100%. The DEA index measures the relative technological performance of fixed location broadband deployment in a given community. The higher DEA index, the higher the evaluation of broadband market performance. The community with DEA efficiency of 100% denotes that it is relatively efficient, e.g. there is no other community in the dataset examined, which is better with respect to all three criteria given the same inputs. On the other hand, communities with DEA efficiency lower than 100% are regarded as not efficient, which means that there is at least one theoretical community that performs better. For instance, DEA efficiency of 25% means that there is a better performing community, which meets all three output criteria having inputs one-fourth as low. For each ineffective community the DEA method returns a benchmarking set, i.e. the set of communities which should be regarded as communities to be followed for a worse performing community.

To assess the performance improvement between 2010 and 2011, we apply DEA to the dataset consisting of communities measured in both years. Hence, the dataset consists of 4 924 observations, i.e. twice the number of communities

[3] These variables are the inputs of DEA method, i.e. variables whose low values are desirable.

in the dataset examined. We apply the input-oriented DEA method with increasing returns to scale. The calculation is performed in GNU R statistical programming environment (see R Core Team, 2012) with the help of the "Benchmarking" package (see Bogetoft and Otto, 2011).

3.2 Supervised learning step

In the second step, we explain DEA indices obtained in the first step. To model DEA indices, we use explanatory variables that are both beyond and of partial control by the national regulatory authority. Therefore, understanding the mechanism of their influence on broadband advancement might be of great use for a regulator to tailor its policies according to recommendations provided by the model. Although some variables are beyond the control of the regulator, it is important to include them in the second step of the procedure, since they may influence variables that are under the regulator's control.

Candidates for variables used in the modeling are the following:

- Computerization level in local schools (SchoolComp),
- Local government expenditures on telecom infrastructure and service (GovExponTelecomInf and GovEspOnTelecomSer),
- Average income in a community (AvgIncome),
- Herfindahl-Hirschman index in a community (HHI),
- Age structure of community population (WorkAge, PreWorkAge),
- Education structure of community population (Prim- & HighEduc).

To model DEA indices with the help of explanatory variables, we can use any of the supervised learning techniques, e.g. linear regression, artificial neural nets or generalized additive model (see Hastie et al., 2009; Kamiński and Zawisza, 2012; Koronacki and Ćwik, 2008; Walesiak and Gatnar, 2009). However, we take the advantage of the regression tree technique, because of its simplicity of interpretation and ease of visualization. In particular, we apply a conditional inference framework in the induction of regression trees (see Hothorn et al., 2006). Additionally, we present the ranking of the importance of variables with the help of random forest technique.

The calculation is performed in GNU R statistical programming environment (see R Core Team, 2012) with the help of the "party" package (see Horthorn et al., 2006) and the "randomForest" package (see Liaw and Wiener, 2002).

4 Results of Internet Access Analysis

4.1 Data Envelopment Analysis – I stage

In the first step, the input-oriented DEA method with increasing returns to scale is employed on the dataset of 4 924 observations (twice the number of communities for 2010 and 2011). A DEA model includes three output criteria: (1) the number of households with at least one broadband provider (availability), (2) the number of households that subscribe to fixed location broadband services (adoption) and (3) the number of households with at least two broadband service providers (competition). DEA inputs are: (1) community population size and (2) the number of households in a community. The result of the DEA method are efficiency indices assigned to each local market in a specified year. The distribution of DEA indices separately for 2010 and 2011 is shown using box-plots and density functions in Figure 1.

Moreover, we consider two approaches: (1) unweighted community observations and (2) community observations weighted by their population size. We show that using the technique of weighting observations by their population sizes significantly influences: (1) calculated summary statistics, (2) their interpretation and (3) implications concerning the performance evaluation of the fixed location broadband market in Poland. The approach of not weighting communities by their population sizes treats communities as units of interest, regardless of their population sizes, whereas the approach of weighting puts emphasis on the citizen, not the community. Summary statistics calculated with the use of weighted observations is a more reliable picture of Poland's fixed-location broadband market performance, since it is taken from the viewpoint of the citizen and not a single community. The unweighted technique assumes that each community is of the same importance and has the same population size. As a result, the unweighted approach puts higher relative importance to rural communities of low population size in comparison with urban communities.

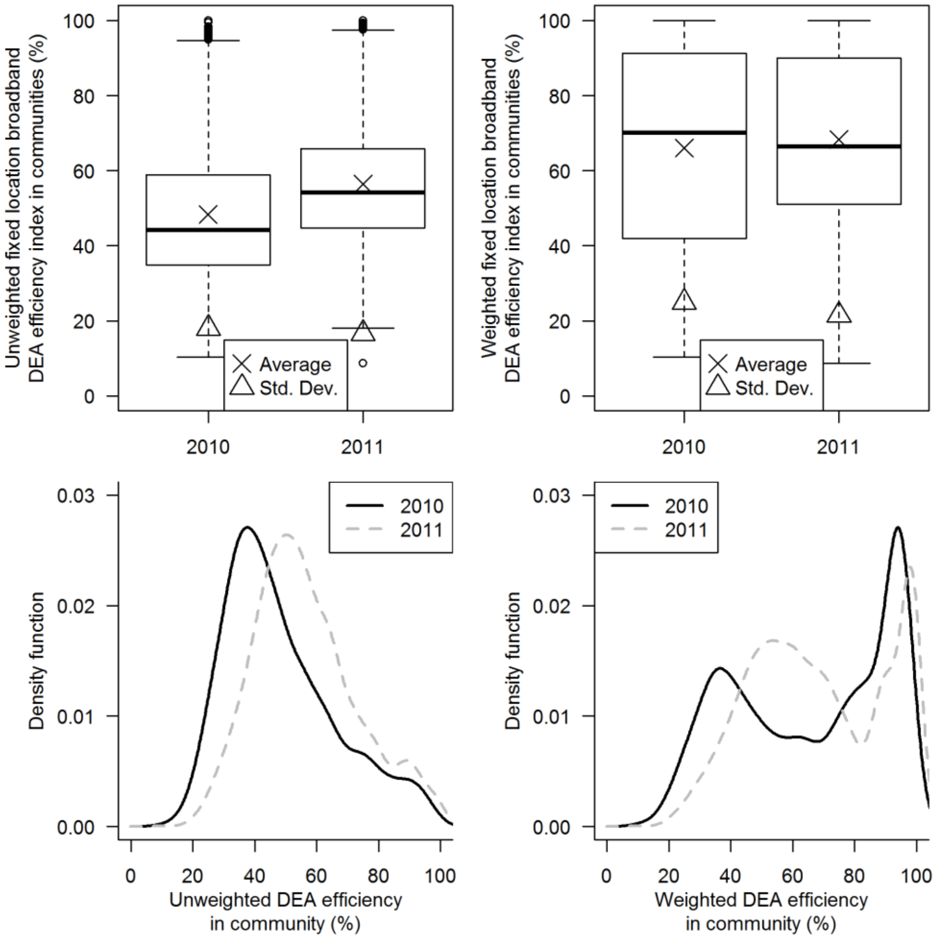

Figure 1. Unweighted (left) and weighted (right) box-plots (top) and density functions (bottom) of Data Envelopment Analysis (DEA) efficiency indices in communities for 2010 and 2011

The upper panels of Figure 1 consist of box-plots of DEA efficiency indices. The upper left panel presents the box-plot of unweighted observations, thus it deals with communities as units of interest. Based on this box-plot we can draw three important conclusions. First, the average of DEA efficiency indices over communities in both years is quite low: 48.3% in 2010 and 56.5% in 2011. This is due to the fact that most communities in our datasets are rural, which perform much worse than urban ones. However, in unweighted approach rural and urban communities have the same weights, left panels tell mainly the story of rural areas. Second, the upper left panel of Figure 1 indicates low dispersion of DEA efficiency indices with interquartile range of 23.9 pp. in 2010 and 21.1 pp. in 2011, i.e. half of communities differed in their DEA efficiency indices by no more than 23.9 pp. in 2010 and 21.1 pp. in 2011. This is again due to the fact that most observations in the dataset are similar to each other, as they are poor performing rural communities. Third, in terms of dynamics, we observe a significant improvement of DEA efficiency indices over the years 2010 and 2011. The median of DEA indices increases by 10.0 pp. from 44.2% to 54.2%. This large improvement is again due to an impressive progress in small, rural communities, which constitute over 60% of all communities.

On the other hand, the upper right panel of Figure 1 tells the story from the viewpoint of a citizen, not a community. As a result, this picture reflects the actual situation more reliably. Based on this panel we can draw the following three conclusions. First, the average of DEA efficiency indices over citizens in both years is quite high 66.0% in 2010 and 68.3% in 2011, especially in comparison to the analogous statistics from the upper left panel. These differences are due to the fact that most citizens live in large cities with well performing broadband market. Second, the upper right panel of Figure 1 indicates a high dispersion of DEA efficiency indices with interquartile range of 49.3 pp. in 2010 and 39.0 pp. in 2011, i.e. half of population differed in their DEA efficiency indices by no more than 49.3 pp. in 2010 and 39.00 pp. in 2011. This indicates large inequalities among Polish citizens from various regions as regards Internet access. Third, in terms of dynamics, we see two interesting facts. The dynamics of DEA efficiency inequality is improving quite remarkably, since the interquartile range decreases by 10.3 pp. from 49.3 pp. in 2010 to 39.0 pp. in 2011. It is reached mainly by the significant upward shift of first quartile by 9.1 pp. from 42.0% to 51.8% and also a minor decrease of third quartile by 1.3 pp. from 91.3% to 90.0%. At the same time the dynamics of level of DEA efficiency stagnates, as the average increases slightly by 2.3 pp. from 66.0% to 68.3% and the median decreases slightly by 3.6 pp. from 70.1% to 66.5%.

The changes observed are due to the fact that poor performing rural communities are catching up, whereas the best performing, mainly urban communities do not progress anymore.

The bottom panels of Figure 1 provide a more comprehensive picture of DEA efficiency indices, as they present density functions. As before, we distinguish the case of unweighted observations (bottom left panel) and weighted observations (bottom right panel). Each panel looks different; they provide two complimentary pictures of the fixed-location broadband market performance in Poland.

The bottom left panel of Figure 1 confirms conclusions drawn from the upper left box-plot panel. On the level of communities we observe: (1) low values of DEA efficiency indices, as the mode[4] of distribution is just ca. 40% in 2010 and 50% in 2011, (2) a remarkable improvement of DEA efficiency over the period

[4] The most probable value of random variable.

2010-2011, as the distribution shifts to the right, (3) small dispersion of DEA efficiency, as most of probability mass is located within an interval of ca. 30%-60%.

The bottom right panel of Figure 1 confirms conclusions drawn from the upper right box-plot panel. On the level of citizens we observe: (1) high values of DEA efficiency indices, as the mode of distribution is just nearly 100% in both years, (2) large dispersion of DEA efficiency, as there is a lot of probability mass assigned to low values of ca. 30% and 100% and (3) the remarkable improvement of inequalities in the DEA efficiency distribution over the period 2010-2011, as the left tail and a significant part of the left distribution mass moved to the right, while the right distribution mass moved slightly toward the center.

The dynamics of DEA efficiency change in the period 2010-2011 is depicted in Figure 2. DEA efficiency indices of both 2010 (the horizontal axis) and 2011 (the vertical axis) are presented on a scatterplot. Each single point represents a community. We distinguish three type of communities: (1) urban communities marked as black circles, (2) urban-rural communities marked as dark grey crosses and (3) rural communities marked as light dark triangles. Moreover, we draw a 45-degree line, which indicates communities with the same DEA efficiency index values in both years. Points lying above this line are communities that experienced an efficiency improvement in the period examined. Analogically, points below this line are communities which regressed into a worse efficiency DEA index. Additionally, for all data points we estimate a spline function presented here with a black thick line to illustrate the average change in DEA efficiency index among communities.

As we can see in Figure 2, rural communities constitute the majority of the dataset and its points are primarily located in lower values of DEA efficiency index (60% and less). However, we also observe a significant improvement among rural communities, since most of them lie above the 45-degree line. The estimated spline function also lies above this line, which indicates that average communities with low and medium DEA index experienced an efficiency improvement of nearly 10 pp.

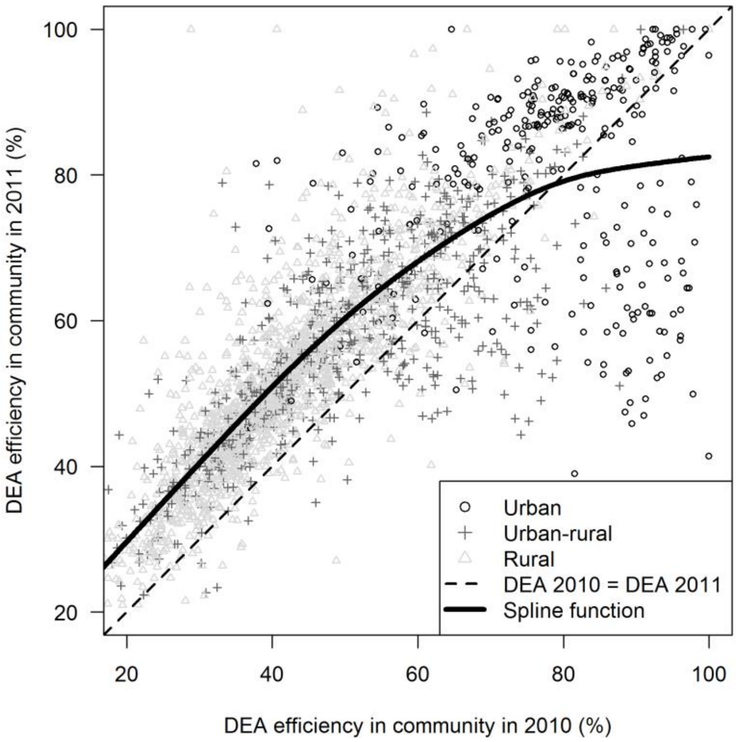

Figure 2. Scatterplot and spline function of relationship between DEA efficiency index values in 2010 and in 2011

On the other hand, as we can see in Figure 2, communities with high DEA index values (80% and more) in 2010 could not have been able to maintain on average a high level of efficiency. Many points with high efficiency in 2010 lie below the 45-degree line, which indicates a regress. These points represent mainly urban and urban-rural communities. This might be due to two reasons. First, some urban communities with saturated broadband market might indeed have experienced some efficiency loss caused by the decrease of the number of subscribers. Second, the phenomena observed might be related to the quality of data. It is possible that data collected for some communities is too optimistic in regard to three criteria in comparison to 2011. The exact source finding would require to analyze each atypical DEA value decrease on its own.

Moreover, communities with DEA efficiency index values of ca. 75%-80% in 2010 experienced on average the same level of efficiency in 2011, as the interval of 75%-80% is the region where the estimated spline function crosses the 45-degree line.

4.2 Supervised learning – II stage

In the second step, we model the DEA efficiency index. In order to do this, we use control variables that are excluded from the first DEA procedure step. We capture the relationship between explanatory variables and the DEA efficiency index by the technique of regression tree. The result of induced regression tree is depicted in Figure 3.

The regression tree provides us with prediction rules concerning the level of DEA efficiency index. A single prediction rule is a path in a graph from the root of the tree to a leaf. For instance, the path following to the leaf number 4 results in the rule that predicts the very high level of DEA index of 85%. The rule has the following form:

IF(CommunityType=urban) AND (HighEduc>11%) → (DEAefficiencyIndex=85%)

This means that an urban community with the share of population with higher education over 11% has on average the DEA efficiency index of 85%. There are 113 communities that fulfill these two conditions. In the case of urban communities with lower than 11% share of higher education in population the average DEA index value is 76%. Hence, the share of population with higher education contributes positively to the level of fixed location broadband advancement. It might be that citizens with a higher educational degree are more aware of Internet advantages and demand for this kind of services. Moreover, higher education population share may not be the direct cause of higher efficiency, but may be correlated with other factors that matter directly.

As we can see to the right of the first node in Figure 3, the remaining cases are concerned with urban-rural and rural communities. All leaves on the right predict lower DEA efficiency index values than those on the left of the first node. Hence, the type of community is crucial for determining the expected level of efficiency. Of course, it is not the administrative decision that matters here, but the fact that community type is strongly correlated with other factors fostering or hampering the development of fixed location broadband infrastructure, e.g. in sparsely populated rural areas the high cost of vast infrastructure may make it unprofitable for entrepreneurs to invest, especially that the demand from less educated population for the broadband is also lower. Hence, regulators should take into account the community type they want to support.

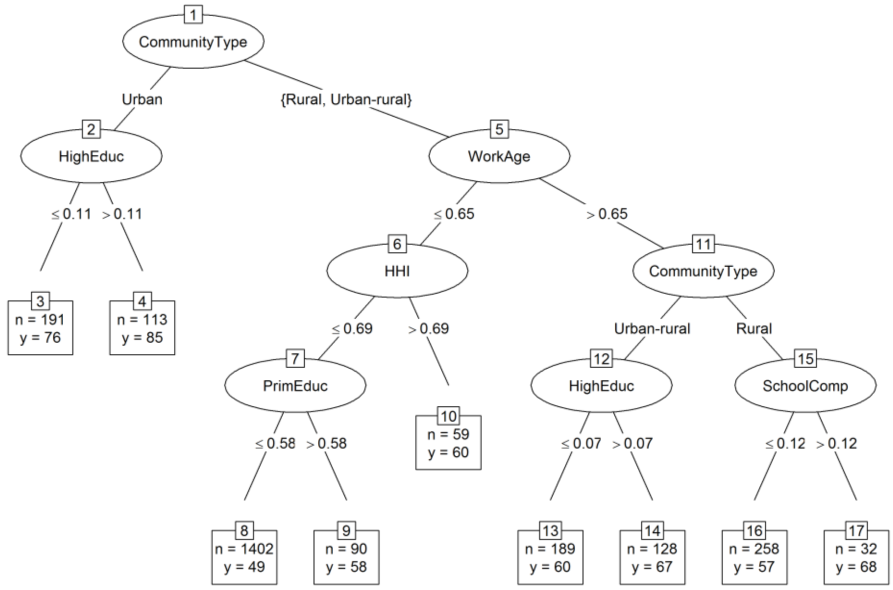

Figure 3. Induced regression tree for the explanation of DEA efficiency index dependent on socio-economic factors and the community type

The next splitting criterion is the share of production age population. The communities with this share larger than 65% have on average higher values of DEA efficiency index, as predictions in leaves on the right side are nearly always higher than those in the two leaves on the left side. For those communities with high production age population share the final splitting criterion depends on the community type. For urban-rural communities, what matters is a higher education population share. If this share is higher than 7%, then an urban-rural community achieves, on average, a DEA efficiency index of 67%, otherwise only 60%. This shows again that the educational structure matters, as it influences the demand for Internet services. On the other hand, for rural communities it is the level of computerization in schools that matters more. If the number of computers with broadband Internet in schools per pupil is higher than 12%, then a rural community can expect on average the DEA efficiency score of 68%, otherwise a score of only 57% is achieved. This shows that in rural areas the impact of demand on the part of children is crucial for household decisions concerning Internet subscription.

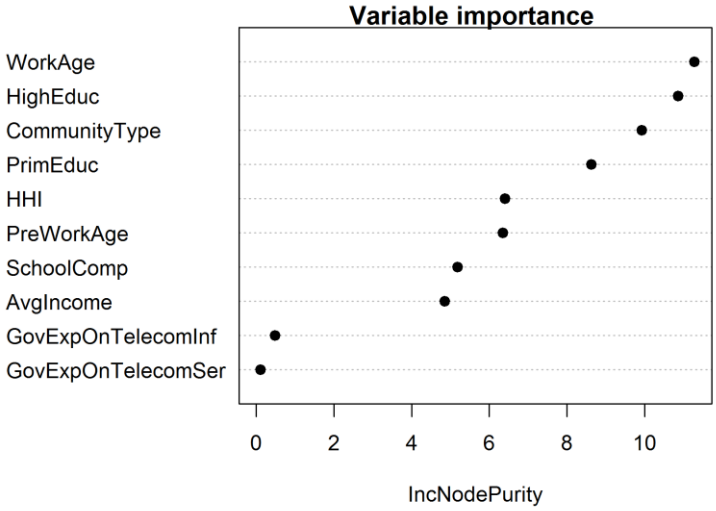

Figure 4. The ranking of socio-economic variable importance in predicting DEA efficiency

As it is shown in Figure 3, the most important determinants of DEA efficiency that occur in a regression tree are: community type, population age structure, population educational structure, school computerization level, and Herfindahl-Hirschman index. The importance of these variables is also confirmed in Figure 4. According to the ranking of variable importance, the most important are variables concerned with population structure with respect to its age and education. The community type is also very important. The Herfindahl-Hirschman Index and average income also play some role. Less significant are local government expenditures on telecommunication infrastructure and services.

The results of the analysis highlight that public policy concentrated on the stimulation of competition on local markets (measured by Herfindahl-Hirschman Index in the analysis) and investment in computerization of community schools influence Internet access more than direct government expenditures.

5 Conclusions

In our paper we propose a two-step procedure for the performance evaluation of regions according to specified criteria. The analysis of communities performed in respect to the advancement of fixed-line broadband provides the Polish telecommunications regulator with valid policy-making implications, which are important for the fulfillment of regulatory requirements and objectives specified in Gaj (2012). The method proposed enables the regulator to compare communities and prepare their ranking. In particular, it allows us to identify communities with both highest and lowest performance of fixed location broadband technological advancement. High performing communities might serve as benchmarks for less successful communities experiencing technological gap, so that UKE can try to replicate good patterns and practices from well performing communities in the poorer ones. The identification of poor performing communities enables the regulator to concentrate its tailor-made activities only in communities of great need. This feature of DEA method fulfills one of UKE's goals set by Gaj (2012), which states that "(…) identification of areas with unsatisfied demand for fixed line services and broadband internet access is crucial for consumer-oriented policy consisting in improving service accessibility in rural areas and widening the choice of service providers in urbanized areas".

Furthermore, regulatory measures aimed at improving poor performing communities may be specified according to the second step of the procedure proposed. We show that the effective regulation may foster the advancement of fixed location broadband Internet access. For instance, the analysis performed reveals that the level of citizens' education influences significantly the broadband advancement in urban and urban-rural communities, whereas in rural communities the level of school computerization plays a vital role. An active regulator might direct its resources into incentivizing local governments to increase their level of computerization in schools as well as introduce innovative multimedia-based classes in schools. All these measures will increase the awareness among pupils which will be propagated and incorporated into their parents' decisions regarding Internet subscription.

The analysis presented enables telecommunication regulators to set precise and attainable, short- and medium-term goals for communities. The results of the analysis highlight that public policy concentrated on stimulation of competition on local markets (measured by the Herfindahl-Hirschman Index in the analysis) and investment in computerization of community schools influence Internet access more than direct government expenditures. Additionally, the method allows us to compare units from various periods so that the progress of performance of communities can be assessed as well. Also the effectiveness of regulatory activities can be evaluated by their impact on a community standing over time. All these advantages of the two-step procedure proposed make it into a versatile and robust tool that enables to assess the communities' advancement with regard to fixed location broadband deployment and the effectiveness of regulatory actions.

References

Alfonso A., Fernandes S. (2008), *Assessing and Explaining the Relative Efficiency of Local Governmen*, The Journal of Socio-Economics, Vol. 37.

Badasyan N., Shideler D., Silva S. (2011), *Broadband Achievement Index, Moving Beyond Availability.* Telecommunications Policy.

Bogetoft P., Otto L. (2011), *Benchmarking with DEA, SFA, and R*, Springer.

Bresnahan T., Trajtenberg M. (1995), *General Purpose Technologies 'Engines of Growth'?*, Journal of Econometric, Vol. 65, No. 1.

Brockett P., Charnes A., Cooper W., Huang Z., Sun D. (1997), *Data Transformations in DEA Cone Ratio Envelopment Approaches for Monitoring Bank Performances*, European Journal of Operational Research, Vol 98, No. 2.

Charnes A., Cooper W., Rhodes E. (1978), *Measuring the Efficiency of Decision MakingUunits*, European Journal of Operational Research.

Commision of EC (2008), *Indexing Broadband Performance*, The European Communities.

Cooper W., Seiford L., Tone K. (2006), *Introduction to Data Envelopment Analysis and Its Uses*, Springer.

Czernich N., Falck O., Kretschmer T., Woessmann L. (2011), *Broadband Inrastructure and Economic Growth*, The Economic Journal, Vol. 121, No. 552.

Fernandez-Menedez J., Lopez-Sanchez J., Rodriguez-Duarte A., Sandulli F. (2009), *Technical Efficiency and Use of Information and Communication Technology in Spanish Firms*, Telecommunications Policy, Vol. 33, No. 7.

Fornefeld M., Delaunay G., Elixmann D. (2008), *The Impact of Broadband on Growth and Productivit*, . A study on behalf of the European Commision.

Gaj M. (2012), *Strategia regulacyjna do roku 2015*, Urząd Komunikacji Elektronicznej, Warszawa.

Giokos D., Pentzaropoulos G. (2008), *Efficiency Ranking of the OECD Member States in the Area of Telecommunications, A Composite AHP/DEA study*, Telecommunications Policy, Vol. 32, No. 9-10.

Guzik B. (2009), *Podstawowe modele DEA w badaniu efektywności gospodarczej i społecznej*, Wydawnictwo Uniwersytetu Ekonomicznego w Poznaniu.

Growiec J. (2012), *The World Technology Frontier, What Can We Learn from the US States?*, Oxford Bulletin of Economics and Statistics, Vol. 74, No. 6.

Grubesic T. (2008), *Spatial Data Constraints, Implications for Measuring Broadband*, Telecommunications Policy, Vol. 32, No. 7.

Grubesic T. (2010), *Efficiency in Broadband Service Provision, A Spatial Analysis*, Telecommunications Policy, Vol. 34.

Hastie T., Tibshirani R., Friedman J. (2009), *The Elements of Statistical Learning, Data Mining, Inference, and Prediction*, Springer.

Ho C., Zhu D. (2004), *Performance Measurement of Taiwan's Commercial Banks*, International Journal of Productivity and Performance Management, Vol. 53, No. 5.

Holt L., Jamison M. (2009), *Broadband and Contributions to Economic Growth, Lessons from the US Experience*, Telecommunications Policy, Vol. 33, No. 10-11.

Hothorn T., Hornik K., Zeileis A. (2006), *Unbiased Recursive Partitioning, A Conditional Inference Framework*, Journal of Computational and Graphical Statistics.

Huggins R. (2003), *Creating a UK Competitiveness Index, Regional and Local Benchmarking*, Regional Studies, Vol. 37, No. 1.

ITU (2009), Measuring the Information Society – The ICT Development Index, International Telecommunications Union.

Kamiński B., Zawisza M. (2012), *Receptury w R. Podręcznik dla ekonomistów*, Oficyna Wydawnicza Szkoły Głównej Handlowej, Warszawa.

Kang Ch. (2009), *Privatization and Production Efficiency in Taiwan's Telecommunications Industry*, Telecommunications Policy, Vol. 33.

Katz R. (2009), *Estimating Broadband Demand and Its Economic Impact in Latin America*, Proceedings of the 3rd ACORN-REDECOM Conference Mexico City.

Katz R., Vaterlaus S., Zenhäusern P., Suter S. (2010), *The Impact of Broadband on Jobs and the German Economy,* Intereconomics, Vol. 45, No. 1.

Kelly T., Rossotto C. (eds.) (2012), *Broadband Strategies Handbook*, World Bank, Washington.

Koronacki J., Ćwik J. (2008), *Statystyczne systemy uczące się*, Akademicka Oficyna Wydawnicza EXIT, Warszawa.

Koutroumpis P. (2009), *The Economic Impact of Broadband on Growth: A Simultaneous Approach*, Telecommunications Policy, Vol. 33, No. 9.

Lam P., Lam T. (2005), *Total Factor Productivity Measures for Hong Kong Telephon,* Telecommunications Policy, Vol 29, No. 1.

Lam P., Shiu A. (2008), *Productivity analysis of the telecommunications sector in China*, Telecommunications Policy, Vol. 32, No. 8.

Liaw A., Wiener M. (2002), *Classification and Regression by Random Forest*, R News.

OECD (2008), *Broadband and the Economy*, Ministerial Background Report DSTI/ICCP/IE(2007)3/FINAL.

R Core Team (2012), *R, A language and Environment for Statistical Computing*, R Foundation for Statistical Computing, URL: http,//www.R-project.org/.

Sastry P. (2009), *Identifying Leaders and Laggards – A Method and Application to US Local Telephone Companies*, Telecommunications Policy, Vol. 33, No. 3-4.

TechNet (2003), *The State Broadband Index*.

Tsai H., Chen Ch., Tzeng G. (2006), *The Comparative Productivity Efficiency for Global Telecoms*, International Journal of Production Economics, Vol. 103, No. 2.

Uri N. (2003), *Technical Efficiency in Telecommunications in United States and the Impact of Incentive Regulation*, Telecommunications Policy, Vol. 27, No. 1.

Walesiak M., Gatnar E., red. (2009), *Statystyczna analiza danych z wykorzystaniem program R*, Wydawnictwo Naukowe PWN, Warszawa.

Zawisza M., Kamiński B. (2012), *Oligopolistic Competition in Geographical Regions with Product Differentiation*, in: *Proceedings of the 8th European Social Simulation Association Conference*, Ed. A. Koch, Salzburger Geographische Arbeiten, Band 48.

Zhang Q., Cheng L., Boutaba R. (2010), *Cloud Computing, State-of-art and Research Challenge*,. Journal of Internet Services and Applications, Vol. 1, No. 1.

Zhu J. (2000), *Multi-factor Performance Measure Model with an Application to Fortune 500 companies*, European Journal of Operational Research, Vol. 123, No. 1.

Zhu J. (2004), *Imprecise DEA via Standard Linear DEA Models with a Revisit to a Korean Mobile Telecommunication Company*, Operations Research, Vol. 52, No. 2.

Zhu J. (2009), *Quantitative Models of Performance Evaluation and Benchmarking*, Springer.