Branko Radulović[1]
Stefan Dragutinović[2]

JEL: H11, H40, R51
DOI: 10.5937/industrija43-8846
UDC:005.336.1:352.07(497.11)"2012"
005.61
Original Scientific Paper

Efficiency of local self-governments in Serbia: an SFA approach

Article history:
Received: 2 May 2015
Sent for revision: 2 June 2015
Received in revised form: 24 August 2015
Accepted: 1 September 2015
Available online: 15 October 2015

***Abstract**: We examine the extent of inefficiency in a sample of 143 Serbian local self-governments for the year 2012. The input side is described by current expenditures of the local self-governments, and the outputs are describing provision of key public services. We use stochastic frontier approach. Our results reveal that the "average" local self-government in the Republic of Serbia produces public services i.e. output at costs that are approximately 23 percent higher than the 'best practice' peers; once we control for the exogenous constraints on public service provision. We provide a brief discussion of our results and guidance for future research.*

***Keywords**: Serbia, Local self-government, Technical efficiency, Stochastic frontier analysis*.

Efikasnost lokalne samouprave u Srbiji: SFA pristup

***Apstrakt**: Rad analizira nivo neefikasnosti na uzorku od 143 lokalnih samouprava u Republici Srbiji u 2012. godini. Kao ulaznu veličinu koristimo nivo tekućih rashoda lokalnih samouprava, dok na strani izlaznih veličina koristimo indikatore koji opisuju ključne javne usluge lokalnih samouprava. U radu koristimo parametarski pristup merenju efikasnosti. Rezultati pokazuju da primenom stochastic frontier metode, u slučaju kada kontrolišemo uticaj egzogenih ograničenja na ponudu lokalnih javnih usluga, "prosečna" lokalna samouprava proizvodi javne usluge uz troškove koji su za oko 23 procenta veće u odnosu na najefikasnije lokalne samouprave. Rad sadrži kratak osvrt na dobijene rezultate i putokaz za nastavak istraživanja.*

[1] Faculty of Law, University of Belgrade, bradulovic@ius.bg.ac.rs
[2] Economics institute, Belgrade

***Ključne reči**: Srbija, lokalna samouprava, tehnička efikasnost, stochastic frontier analysis.*

1. Introduction

The current financial and economic crisis imposed considerable constraints on government budgets. Indeed, the effects of the crisis are not limited to the national level as they also put a significant curb on spending of local governments. While local governments can consolidate their budgets either by cutting public expenditure (reducing quality and/or quantity of public services) or by increasing public revenues (imposing new taxes), they can also reduce expenditure by operating more efficiently. Local governments that operate inefficiently are not providing the same amount or quality of public services as their more efficient peers for a given level of expenditure. Such local self-governments often fail to attract private investors and scarce human capital. Besides the fiscal aspects, the significance of local government's efficiency has been increasing with the implementation of decentralized policies. In fact, the optimal degree of decentralization may determine spending efficiency at the local level (Afonso and Fernandes, 2008; Prud'homme, 1995), as "*there is little to gain by shifting tasks to inefficient levels of government*" (Geys and Moesen, 2009).

Research on efficiency of local government services provision is divided into two main approaches. The first approach includes studies that focus on the assessment of efficiency of single services delivered by local governments (e.g. waste collection, administration, road maintenance, etc.) The second (composite) approach includes studies that are aimed at assessing an overall local government efficiency score. Composite approaches to local government technical or cost efficiency became widespread particularly within European countries (Šťastná and Gregor, 2015). In this stream, scholars have conducted a number of empirical investigations that cover several European countries (or regions within countries): Belgium (Geys and Moesen, 2009a and 2009b; De Borger and Kerstens, 1996; De Borger et al., 1994), Czech Republic (Šťastná and Gregor, 2015, 2011), Germany (Geys et al. 2013, Geys et al., 2012; Kalb et al, 2012; Kalb, 2010; Geys et al., 2010), Finland (Loikkanen and Sisiluoto, 2005), France (Seifert and Nieswand, 2014), Greece (Athanassopoulos and Triantis, 1998), Italy (lo Storto, 2013; Boetti et al., 2012), Portugal (Afonso and Fernandes, 2008 and 2006), Slovenia (Pevcin, 2014), Spain (Balaguer-Coll et al., 2010; Bastida and Garcia, 2010; Benito et al., 2008; Gimenez and Prior, 2007). The researchers examined local governments' efficiency mostly by applying non-parametric and parametric estimation approaches. A comprehensive overview of the existing empirical literature suggests that there are "*substantial inefficiencies in the*

provision of public goods and services by local governments" (Kalb et al., 2012).

In this paper we examine efficiency of Serbian local self-governments (LSG) for the 2012 fiscal year. Before we proceed, we would like to point out that in this paper we mainly study only one aspect of the local government public goods and services provision, namely its technical and, to some extent, its cost efficiency. In general, efficiency can be measured as the ratio of the observed to the maximum potential output given a certain amount of input (output orientation), or the ratio of minimum potential to actual input (input-orientation). However, there are various measures of efficiency. Technical efficiency is defined in terms of production possibilities, i.e. local governments that operate on a production frontier are technically efficient. Any decision-making unit may be technically efficient, but yet allocatively inefficient because it fails to choose correct input combination. Both technical inefficiency and allocative inefficiency increase cost and lead to cost inefficiency (Kumbhakar, et al. 2015) defined as the rate of minimum to actual cost (the observed cost lies above the minimum frontier due to inefficiency).

The study is the first one on Serbian LSG data. The contribution of this paper is threefold. Firstly, in Serbia the need to significantly reduce the amount of public expenditure at all government levels and (vertical) fiscal misbalance between the central and local government levels (Fiscal Council, 2013) makes the issue of measuring efficiency of local governments even more pertinent. In addition, examining the relationship between the efficiency of local self-governments and exogenous factors and identifying causes of inefficiency is of particular interest to local policy makers. Secondly, results enable comparison of the efficiency of Serbian LSGs with results for local governments from other countries. However, as noted by Kalb et al. (2012) we need to be very careful with comparison due to various heterogeneity issues (different input and output variables, diverse analytical tools, etc.). Thirdly, in Serbia a further decentralization of responsibilities from the central to the local self-government level is under consideration. Our results provide information to guide political decisions on whether LSGs efficiently carry out the responsabilities bestowed upon them.

Finally, in our study efficiency is a relative not an absolute concept. The efficiency of any local government is evaluated relative to the efficiency of other local governments in the sample. Thus, the local government that is deemed to be the most efficient need not necessarily be efficient in absolute terms; as in reality the actual efficiency could be even significantly higher.

The paper is organized as follows. Section 2 provides a brief overview of the methods used to determine the efficiency of LSGs. Section 3, describes data and measurement issues. This section also provides a short introduction of the institutional framework of the local self-government in Serbia. In section 4,

the SFA method is used to assess different aspects of the efficiency of LSGs in Serbia, using data presented in section 3. In this section we present and discuss the empirical results that show that there is a substantial divergence in efficiency scores. Our results reveal that the average local government in the Republic of Serbia "produces" public services i.e. output at costs that are approximately 23 percent higher than the 'best practice' peers; once we control for the exogenous constraints on public service provision. Section 5 concludes.

2. Methodology

The first step is to select an appropriate input–output combination that can determine efficient behavior (best practice frontier) of what are usually called decision-making units (here local self-governments). Then, in the second step, one attributes deviations from the best practice frontier to inefficiency. However, there are several issues related with the attempts to implement these two steps: Firstly, how to derive the best practice frontier from a data set of local self-governments. Secondly, how to establish the extent to which deviations from the best practice frontier are attributable to inefficiencies and not to other factors including measurement errors. Overall, in the empirical literature, best practice frontiers have been derived using two general approaches – the non-parametric and parametric methods. In addition, we can distinguish two types of deviations from the best practice frontiers – deterministic and stochastic. A deterministic approach interprets any deviation from the best practice frontier as inefficiency. This may be problematic since observed levels of inputs and outputs may be subject to measurement errors or stochastic influences. Consequently, we can distinguish four classes of method - non-parametric deterministic, parametric-deterministic, parametric-stochastic and non-parametric stochastic methods.

In non-parametric deterministic methods - Data Envelopment Analysis (DEA) and Full Disposal Hull (FDH), the frontier is estimated using a linear programming approach as a piecewise envelopment of the data subject to specific assumptions. More specifically, the DEA assumes that the production frontier is convex and that there is strong free disposability in inputs and outputs, while the FDH differs from the DEA in that it drops the convexity assumption. These non-parametric methods are fully deterministic, and the entire deviation from the best practice frontier is attributed to inefficiency.

Parametric methods determine the frontier on the basis of a specific functional form using econometric techniques (Greene, 2008). Parametric-deterministic methods include corrected ordinary least squares (COLS), modified ordinary least squares (MOLS). COLS assume that any deviation from the frontier is

interpreted as inefficiency (Kumbahar, et al 2015). COLS first estimates a linear regression through the cloud of input-output points, and then to obtain the frontier it shifts the regression line downwards until all of the residuals are positive and at least one is zero (De Borger and Kerstens, 1996).

Stochastic frontier analysis (SFA) is parametric-stochastic approach where the deviation from the best practice frontier is decomposed into two components i.e. the residual is represented by a two-component term, which consists of (a normally distributed) component that captures measurement errors, and (a half-normally distributed) component that captures inefficiency. Consequently, SFA is more precise compared to previously described methods with respect to the definition of the deviation from the best practice frontier (Hjalmarsson et al., 1996). The major drawback of the parametric approach is that one has to make assumptions about the specific functional form and the distribution of the inefficiency term. However, as LSGs differ in various socio-economic, political and other aspects, the stochastic approach enables researchers to 'separate' errors from inefficiency.

Finally, non-parametric-stochastic methods do not require assumptions on functional form and they try to increase robustness associated with stochastic approaches. These methods include the order-m and order-α approaches. While these methods are relatively new they exhibit good results, but only if outliers are presented and if variances of measurement errors are large (Krüger, 2012).

Besides opting for a specific method, researchers need to choose between input and an output perspective. On the one side input-oriented efficiency perspective measures how inputs can be reduced while output levels are held fixed. On the other side, output-oriented efficiency perspective maintains input levels and examines possible output expansion. In the literature there is a wide consensus to use the input-oriented approach. In the case of LSGs, the policy-makers are more likely to influence spending levels (inputs) than the size of infrastructure, number of public facilities and the services related to the number of residents (outputs). Hence, we follow the literature that assumes that input-oriented efficiency is more appropriate and proceed with estimates with current expenditures as input.

Most studies use non-parametric data envelopment analysis to assess efficiency of public goods provision, while several studies utilize the parametric estimation approaches; mainly using stochastic frontier analysis. Both approaches are suitable for estimating efficiency of local self-governments and both have their own advantages and disadvantages. The DEA approach can easily handle multiple outputs representing different LSGs services. DEA does not require input prices or specific assumptions regarding the technology i.e. functional form. This is very useful since governmental behaviour does not have well-established production or cost functions in

microeconomic theory. Several studies based on non-parametric methods use two stage procedures such that the efficiency analysis itself and the second as an evaluation of its determinants (Afonso and Fernandes, 2008). However, this approach has been criticized because of the contradictions between the assumptions made in the first and the second stages (Greene, 2008). In addition, this approach cannot correct the efficiency estimates for the influence of these determinants. By introducing both error terms and control variables we are able to draw more accurate inferences and we emphasize the importance of including socio-economic characteristics in studies of local government efficiency (Kalb et al. 2012).

3. Data

This section describes the institutional background, presents stylized facts for the LSGs analyzed, explains input, output and exogenous variables, and provides descriptive statistics. The choice of relevant input, output and exogenous variables is largely based on previous empirical literature, though some variables are specific to the Serbian LSGs framework.

2.1. Local self-government in Serbia

Serbia has a mono-type, single-tier local government consisting of 174 local self-governments including 150 municipalities, 23 cities and the city of Belgrade. The current system is based on the Constitution of the Republic of Serbia and the Law on Local Self-Government. LSGs in Serbia are relatively large by continental European standards. The average resident population in 2012 was 49,895, with the vast majority of LSGs in the 10,000-100,000 range with only eleven municipalities with resident populations under 10,000. While 174 LSGs exist currently in Serbia, there are also 28 urban municipalities. We exclude urban municipalities from the analysis for several reasons - they are subordinated to city administrations, have few independent functions and their expenditures to a large extent are not comparable to other municipalities in our sample. In addition, urban municipalities are not defined by the law i.e. the division of cities into municipalities is determined by city statutes. Table 1 presents stylized facts for the LSGs for 2012.

Table 1. LSGs descriptive statistics

	Area (sq km, 2012)*	Resident population (2012)*	Population per sq km*	Average current expenditures per capita (2012) RSD**
Average	535	49.649	77.1	24.820
Max	3.234	1.664.218	514.6	60.557
Min	51	1.608	5.1	12.116
Total	77.592	7.199.077	-	-

Source: RSO "Municipalities and Cities in the Republic of Serbia", 2012 and LSG Financial statements database.

The primary responsibilities of LSGs in Serbia are mainly confined to infrastructure services including urban water supply and sewerage, local road maintenance, solid waste collection and, in urban areas, district heating and public transport. These services are typically provided by public enterprises owned by LSGs, which for some services cover all or part of their operating costs through tariffs (World Bank, 2014). Besides this, LSGs also provide a set of social services (Table 2). These include basic education i.e. the operation of primary schools and kindergartens and the maintenance of school buildings (although not the payment of teachers' salaries). Finally, a significant share of LSGs expenditures are devoted to social protection; (supplementing the principal social assistance programs financed by central government) and so-called general administrative services.

Table 2. LSGs expenditure responsibilities

Area	Expenditure responsibility
Basic education	Implementation of legislation and financing of non-wage spending
Social assistance	Implementation of legislation
Infrastructure	Local roads
Communal services	Water, sanitation, and other specific needs of citizens
Other	Culture, sports and environment

Serbia's current LSG financing framework dates back to 2006, when the Law on local self-government financing was adopted (Kecman, 2010). According to the current financing framework, LSGs obtain their revenues from four principal sources: 1) shared taxes (centrally-administered taxes imposed at centrally-determined rates and shared with the municipality in which they are collected); 2) block transfers (central government transfers that are not earmarked for specific purposes); 3) local taxes (including the property tax); and 4) fees and charges (World Bank, 2013).

2.2. Inputs

The majority of the literature approximates inputs either by total spending or by total current spending. Most studies opt for the second approach as more suitable "*given that capital spending is highly volatile*" (Šťastná and Gregor, 2015), and that current spending represents realistic LSG input measure as local governments have limited discretion. Indeed, in 2012, the current expenditures of LSGs represented 79.9 percent of total expenditures (235.1 billion RSD). In 2013, the real current expenditures remained almost exactly same, while capital expenditures significantly decreased (-31.1%). Ministry of Public Administration and Local Self-Government provided data on LSGs budgets. We should note that there are significant differences with respect to average current expenditures per capita (Table 1.)

Following Šťastná and Gregor (2015) we also account for diverse cost conditions in municipalities. Almost all other studies assume that labour costs are quite similar across the local governments. In most countries this is a justified assumption since salary scales of municipal employees are mainly regulated by the central government (e.g. Kalb et al., 2012; Boetti et al, 2012; Seifert and Nieswand, 2014). We should note that most studies consider local governments only within one region or province that makes wage homogeneity an assumption rather plausible. In addition, the availability of information on input prices at the municipal level is typically limited.

In Serbia, according to the existing legislative framework, the size of the local wage bill is to a large extent controlled by the central government (World Bank, 2014) which sets out procedures for determining individual salaries and sets ceilings on LSGs wage bill spending. However, in practice, differences in average salaries among individual LSGs are considerable. Therefore, we obtain wage adjusted inputs by dividing total current costs with average salaries at the local level. In other words, following Šťastná and Gregor (2015) we assume that "*the labor cost difference across LSGs serve as a good proxy for the overall cost difference*". This variable contains substantial imperfections since we were not able to obtain data on gross wages on the municipal level. With respect to capital expenditures, it is often argued in the empirical literature that local governments have access to the same capital markets and thus face rather similar capital related inputs prices, which can also be assumed for the Serbian LSGs.

2.2. Outputs

The selection of LSGs output indicators is mainly driven by data availability, existing empirical literature, and the need to match expenditures with the most appropriate set of variables. We use several output variables with quantitative

characteristics that describe key responsibilities (core services) of the Serbian LSGs with respect to the infrastructure (communal), social, administrative and educational services. Apart from the total resident population administered by the LSG, that represents a good proxy for administration expenditures, we decided not to use rather far proxies for the provision of public services (e.g. the share of residents older than 60 or 65 years). In general, we use similar outputs to other studies. For example, Geys and Moesen (2009b) use the number of subsistence grants beneficiaries, the number of students in local primary schools, the size of public recreational facilities, the total length of municipal roads and the share of municipal waste collected through door-to-door collections.

Table 3 presents the selected output indicators used in the study to proxy the results of LSGs services provision.

Table 3. Selected LSGs services (*output*) *indicators* (*2012*)

Variable	Indicators
Administration	Total resident population
Social protection	The share of social protection users in total resident population
Education	The number of school and preschool institutions
Road maintenance	The length of total or total high quality roads maintained by the LSG
Communal services and sanitation	Number of water connexions - and/or number of connexions - sewerage

Obviously, there are a number of issues with respect to the selected local government outputs. First, some of these indicators do not take into account the differences in quality that may exist across LSGs (Worthington, 2000). Second, the lack of available statistical data on these services at the LSG level prevented us from adding more variables that are related to public service provision. Missing variables are mostly related to the provision of public transportation and cultural and recreational services (e.g. there is no reliable number of total cultural facilities - museums, libraries, theaters, cinemas, children's centers, etc. or sport facilities within the LSGs). However, the small amount of available variables means that we do not have to deal with the the dimensionality problem that leads to biased results, large variance and wide confidence intervals. Šťastná and Gregor (2011) utilize the principal components analysis and narrow 19 output variables into six principal components (using "the 80% rule"). Yet in our case, depending on the mix of outputs we use, the first component would explain more than two thirds of the variance.

Table 4. Selected LSGs services (output) indicators (*2012*), *N=143*

	Total resident population	The share or social protection users in total population	The number of school and preschool institutions	Average current expenditures per capita (2012) RSD	The length of modern roads in km maintained by the LSG	Number of water connexions
Average	50.131	10.03%	46	24.821	102.7	14.247
Max	1.664.218	39.22%	881	60.558	2.433.0	560.256
Min	1.608	1.6%	4	12.117	4.0	94
Total	7.168.792	-	6623	-	14.690	2.037.292

Source: RSO "Municipalities and Cities in the Republic of Serbia", 2012 and LSG Financial statements database.

Most of our variables are highly correlated and contain in some cases basically identical information related to the size of LSG. Thus, we decide to use only the number of water connexions instead of using data on sewerage connexions as well, as they proxy basically the same type of output.

2.3. Determinants – exogenous variables

A researcher may not only want to know the levels of inefficiency for each LSG, but also the factors that can explain inefficiency. Some LSGs may be unable to be on the "best-practice frontier" due to objective constraints i.e. "relative harsh environment" (Afonso and Fernandes, 2008). Therefore, we use several variables to proxy the effect of the environment (socio-economic, demographic and location factors) on the LSGs efficiency. Based on the data available, we use:

1) *Education* – share of population with secondary education and share of university graduates (tertiary education) to proxy for the general educational level. Literature often assumes that local governments with a higher share of educated residents should be more efficient (e.g. De Borger, Kerstens 1996). Still as emphasized by Šťastná and Gregor (2011), the effect of education may be ambiguous as human capital may raise productivity and raise the reservation wage for the public sector as well as increase the demand for (unobservable) high-quality services. Afonso and Fernandes (2008) refer to a broad literature stating that monitoring local government performance and costs is influenced by the local residents' education level;

2) *Population density* – proxies the urban/rural divide and the assumption that population density enables scale economies a well as the ability to cluster provision of local public services which increases its efficiency;

3) *Distance to Route E75* (Motorway A1). Location matters for various reasons (Loikkanen and Susiluoto, 2005) including varying transport costs and possibilities to attract better human capital, etc. To proxy for location effects we calculated the geographical distance between the LSG and the European route E75. The motorway spans approximately 590 kilometers and it spans the country from north to south. We expect that municipalities closer relative to the E75 would be more efficient;

4) *Aging index* or the ratio between the number of people over 65 and the number of people under 18 proxies for the structure of public services provision);

5) *Unemployment* measured as number of unemployed in 1000 residents. Geys et al. (2009a) state that this variable may have an ambiguous effect "*as it implies both higher spending (a 'cost effect') and lower demand for public services which increase with income levels (a 'preference effect')*."

4. Empirical results

4.1. Stochastic frontier analysis

The SFA method assumes a specific functional form describing the relationship between costs and outputs i.e. the cost function representation of a given production technology of LSGs. We follow a consensus in the literature and use the input-oriented approach (i.e., with the goal being to minimize the cost for given output levels). More specifically, the dependent variable is the level of LSG current spending and independent variables are selected outputs. For the i^{th} LSG, the cost function $C_i(y_i, p_i, \beta)$ defines a lower bound for total current costs C_i, necessary to obtain output levels y_i at given input prices p_i, and β represents the set of technological parameters to be estimated.

The SFA allows one to distinguish between the effects of measurement error (noise) and inefficiency (Coelli, et al 2005). This is achieved by introducing a composite error term consisting of a component v (white noise) and an inefficiency as a one-sided non-negative component ($u \geq 0$) .

The SFA has the following log form for the i^{th} LSG:

$$lnC_i = f\left(lny_{r,i}, lnp_{n,i}\right) + \epsilon_i, \quad \epsilon_i = v_i + u_i \tag{1}$$

where C_i designates the input indicator (current expenditures) for LSG *i*, $y_{r,i}$ indicates the r-th output of LSG i (Table 3) and $p_{n,i}$ measures the price of the

n-th input of LSG i. The error term ϵ_i is decomposed into noise v_i which is assumed to be i.i.d., $v_i \sim N(0, \sigma^2)$ and a non-negative inefficiency term u_i having usually half-normal or truncated normal distribution. It is also assumed that cov(v_i, u_i) = 0 and u_i and v_i are independent of the regressors.

More formally, a stochastic frontier model can be specified in a Cobb-Douglas form:

$$lnC_i = \beta_0 + \sum_{r,i} \beta_r lny_{ri} + \sum_{n,i} \beta_i lnp_{ni} + v + u \tag{2}$$

whereas literature most often assumed no variability in input prices:

$$lnC_i = \beta_0 + \sum_{r,i} \beta_r lny_{ri} + v + u \tag{3}$$

In general, total costs and input prices described in the equation (2) are usually normalized by the price of labor replacing lnC_i with $ln(C_i/p_w)$. Note that due to duality, equation (2) has to satisfy several parameter restrictions. Detailed expositions of the Cobb-Douglas model with variability in input prices (homogeneity constrained model) are given in Coelli et al. (2005) and Kumbahar et al. (2015). In empirical literature on the efficiency of local government Šťastná and Gregor (2011) treat wage differentials in two ways - by adjusting expenditures by wage differences or estimating cost function by including wages directly among "outputs".

Translog form that represents a generalization of the Cobb-Douglas function (equation 3) adds cross-products:

$$lnC = \beta_0 + \sum_{r=1}^{s} \beta_r lny_r + \frac{1}{2} \sum_{r=1}^{s} \sum_{q=1}^{s} \lambda_{rq} lny_r lny_q + v + u \tag{4}$$

s points to the number of outputs incorporated in the model and β_r and λ_{rq} are parameters to be estimated. In general, the Translog form requires estimation of many parameters requiring a relatively large sample and it is more difficult to interpret. We limit our discussion to the Cobb-Douglas specification due to multi-collinearity and degrees of freedom issues of the Translog form.

Finally, we follow the approach applied in Geys and Moesen (2009b) and Kalb et al. (2012) who use the single stage approach suggested in Battese and Coelli (1995) and assume that the inefficiency term is a function of the set of control variables (non-discretionary influences) z_i.

$$u = \gamma + \sum_{i=1}^{s} \delta_i z_i + w \quad (5)$$

where γ represents technical efficiency in the error term, z represents a vector of control or background variables (for LGS i), δ represents a vector of parameters to be estimated and w is defined by the truncation of the normal distribution with zero mean and variance (Battese and Coelli, 1995).

We present the summary results in Table 5. Results are obtained using Frontier 4.1c software (Coelli, 1996). For several specifications we obtained results using Stata commands *frontier* and *sfcross* as well as relevant post-estimation options. More specifically, Table 5 shows two sets of results. The first set of results omits the effect of the exogenous (control) variables on efficiency scores. Model (1) represents a 'baseline' Cobb-Douglas model for all LSGs. However, for the sake of homogeneity, in Model (1A) we exclude the capital city of Belgrade as well as the three other largest Serbian cities with total populations above 150,000, i.e. Nis (260,000), Novi Sad (344,000) and Kragujevac (179,000). As large cities provide an idiosyncratic mix of public services that is not properly captured by our data, our preferred specifications are models that exclude the four big cities. The inclusion of wage differences is introduced in Model (1B). Apart from Šťastná and Gregor (2011), all other studies assumed identical labour (and capital costs), however, this is not a valid assumption in the case of Serbia. Hence we prefer specifications that take into account wage differences. Our results show that Serbian LSGs on average could reduce their current expenditures by 25.2% (Model 1B) without reducing their current output levels to achieve the result of the LSG on the 'best practice' frontier.

The next three models (2, 2A and 2B) examine the effect of determinants on (in)efficiency. To the extent that we were able to collect relevant data these models provide insights about socio-economic and demographic influences (production environment) on Serbian LSGs efficiency. As expected, the inefficiency declines once we include exogenous variables. We should note that LSGs that are on the best practice frontier should receive scores that equal *one*, while all other inefficient LSGs receive scores larger than *one* (as a result of their excessive expenditures).

Compared to Model (1B), once we include the exogenous variables, estimated inefficiency decreases and the average LSG expenditures could be reduced by 22.7% (Model 2B). This represents a reduction by about 12% compared to Model (1B). However, Geys and Moesen (2009b) rightfully warn that due to data limitations, we must be "*cautious in equating observed 'inefficiencies' with realizable cost cuts*". Nevertheless, overlooking the

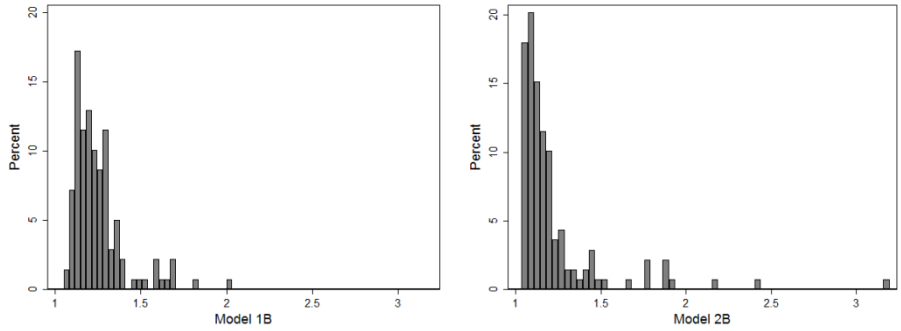
potential effect of exogenous variables can significantly affect results of the model.

As expected there is substantial heterogeneity in efficiency scores. Fig. 1 shows the number of LSGs and their efficiency scores based on the results of Models 1B and 2B. It is evident that in the presence of exogenous control variables efficiency scores shift towards the left (i.e. towards lower relative inefficiency). However, the tail of the distribution gets longer and somewhat thicker.

Table 5. Summary statistics on the efficiency of LSGs in 2012

Cobb-Douglas Model			N	Mean	StdD	Min	Max
No control variables	(1)	All LSGs	143	1.338	0.273	1.038	2.533
	(1A)	No big cities	139	1.268	0.179	1.042	2.167
	(1B)	No big cities – Wage adjusted	139	1.252	0.155	1.051	2.032
Control variables	(2)	All LSGs	143	1.157	0.244	1.021	3.186
	(2A)	No big cities	139	1.214	0.260	1.034	3.137
	(2B)	No big cities – Wage adjusted	139	1.227	0.284	1.037	3.196

Figure 1. Distribution of the SFA efficiency scores Model 1B and Model 2B



In general, economies of agglomeration and economies of scale make the larger LSGs relatively more efficient. In that respect, our results are similar to results obtained by Kalb et al. (2012) for Germany as most LSGs have relatively modest degrees of inefficiency, but some (especially the smallest LSGs) once exogenous variables are taken into account, are remarkably inefficient. Fig.2 reveals that after the introduction of control variables the LSGs with some extremely low and relatively high scores move in opposite direction. Hence, besides inability to exploit economies of scale in the provision of public services, other factors significantly contribute to inefficiency of the smallest LSGs (Table 6). These results provide strong support for the potential amalgamation of the smallest LSGs (below 15,000 residents).

However, smaller and medium sized LSGs that are fiscally vulnerable and have scarce human capital are not necessarily always less efficient compared to larger LSGs.

Figure 1. The SFA efficiency scores and the population size of LSGs

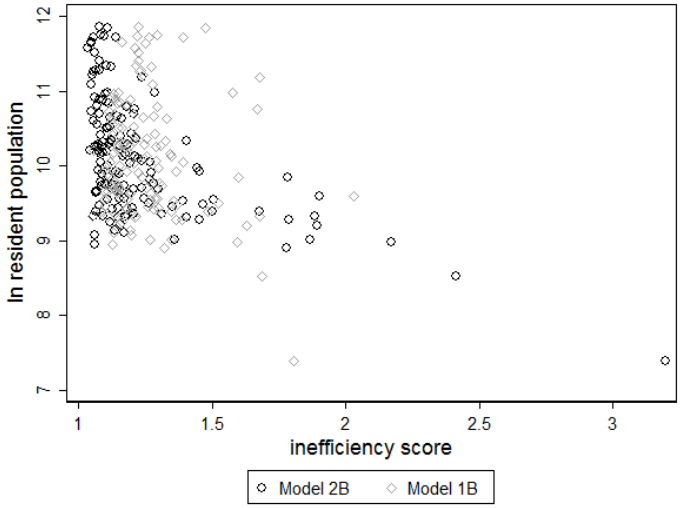


Table 6. Summary statistics on the efficiency for large and small LSGs

	Observations	Mean	Std. Dev.	Min	Max
Below 15.000 residents					
Model 1B	42	1.311	0.213	1.051	2.032
Model 2A	42	1.386	0.401	1.048	3.137
Model 2B	42	1.420	0.435	1.056	3.197
More than 100.000 residents					
Model 1B	10	1.272	0.093	1.163	1.473
Model 2A	10	1.066	0.030	1.034	1.140
Model 2B	10	1.076	0.031	1.037	1.139

Finally, we will briefly provide our results for the exogenous variables (Table 7). Exogenous variables enable a 'corrected' measure of LSG cost (in)efficiency. In general, models provide similar results for the sign and the significance of most coefficients, showing that our findings are relatively robust to alternative specifications. Apart from Model 2, which is the least preferred model, education level is not statistically significant. Yet, education always has a positive sign (increases inefficiency). Thus, we cannot claim that the larger is the highly educated share of LSG population, the more effective it is in demanding more efficient local government. With respect to the

remaining results, we find that geographical location proxied by the distance to the E75 is statistically significant.

Table 7. Results of the SFA estimation

Cobb-Douglas						
ln Current expenditures	No control variables			Control variables		
	All	No big cities	Wage adjusted	All	No big cities	Wage adjusted
	Model 1	Model 1A	Model 1B	Model 2	Model 2A	Model 2B
Stochastic frontier						
Constant (b_0)	11.013***	11.597***	1.817***	12.019***	10.325***	0.453
	-0.434	(0.45)	(3.913)	(0.467)	(0.539)	(0.532)
ln Total resident population	0.747***	0.689***	0.679***	0.666***	0.832***	0.835***
	(0.091)	(0.090)	(0.092)	(0.088)	(0.100)	(0.097)
ln Connexions - Water	0.143***	0.150***	0.091*	0.147***	0.167***	0.093*
	(0.051)	(0.049)	(0.051)	(0.053)	(0.051)	(0.053)
ln Quality roads (km)	0.006	-0.001	0.017	-0.018	-0.060	0.026
	(0.030)	(0.030)	(0.031)	(0.030)	(0.064)	(0.301)
ln Schools	0.025	0.031	0.028	0.005	-0.007	-0.080
	(0.063)	(0.061)	(0.063)	(0.062)	(0.031)	(0.061)
ln Social protection users	-0.0057	-0.0003	0.031	0.021	-0.024	-0.024
	(0.045)	(0.044)	(0.045)	(0.046)	(0.050)	(0.050)
Inefficiency model						
Constant (δ_0)				-1.226**	-0.289	-0.598
				(0.521)	(0.350)	(0.401)
Education				15.696***	3.461	4.704
				(6.098)	(3.403)	3.885
Population density				-0.001	-0.011**	-0.013**
				(0.001)	(0.005)	(0.006)
Distance to E75				0.001	-0.001	-0.003*
				(0.001)	(0.001)	(0.001)
Ageing index				0.004*	0.002**	0.002***
				(0.002)	(0.001)	(0.001)
Unemployment				-0.003	0.003**	0.005***
				(0.005)	(0.001)	(0.002)
σ^2	0.139***	0.108***	0.106***	0.089***	0.108***	0.133***
	(0.025)	(0.024)	(0.025)	(0.002)	(0.012)	(0.244)
γ	0.831***	0.724***	0.664***	0.520**	0.656***	0.736***
	(0.073)	(0.136)	(0.166)	(0.223)	(0.225)	(0.103)
LL	-4.777	1.345	-2.359	8.410	9.018	8.700
LR test of the one-sided error	9.39***	3.14**	2.100*	35.760***	18.482***	24.210***
N	143	139	139	143	139	139
Mean efficiency score	1.338	1.268	1.252	1.157	1.214	1.227

Standard errors in brackets; *** denotes significance at 1% level, ** at 5% and * at 10%; LL represent the loglikelihood of the model, LR test of the one-sided error assesses the null hypothesis of no inefficiencies against the alternative hypothesis that inefficiencies are present. (Kodde and Palm, 1986) All models were obtained using FRONTIER 4.1 software. We do not report results of models assuming truncated-normal distribution as parameter μ was not statistically significant.

The agglomeration effects (the ability of the LSGs to cluster local public service provision as a result of high population density) are statistically significant. As expected, higher density is negatively related to inefficiency

(i.e. positively related to efficiency) as LSGs with a dense settlement structure can organize and utilize their networks more efficiently. The structure of service provision proxied by the age index also affects efficiency. We observe a positive relationship between the age index ratio and inefficiency (the number of people over 65 divided by the number of people under 18). Finally, the unemployment rate has a statistically significant positive sign, suggesting that this aspect of production environment plays a significant role and that the cost effect (need for special services and various subsidies) probably outweighs the preference effect (i.e. reduced demand for high quality public goods). We can conclude that there is an obvious significance of the exogenous constraints on the efficiency. On the one side, the possibility to substantially improve efficiency is somewhat limited, as LSGs cannot effectively resolve issues such as demographic and socio-economic constraints. On the other side, results show that there is an evident prospect to improve current state of affairs without resorting to tax increases.

Results highlight the importance of the inefficiency u in comparison to noise v in determining the error term ϵ. The parameter γ is high (0.736 in Model 2B) and much of the variation in the composite error term is due to the inefficiency component. We also report the test of the hypothesis of no inefficiency by comparing the log-likelihood of the OLS model to that of the half-normal model. More specifically, if the LSGs are fully efficient, the model reduces to an OLS (Kumbhakar, et al., 2014). LR test of inefficiency (LR test of one-sided error) supports rejection of the null hypothesis of no inefficiency.

5. Conclusions

We examined the extent of inefficiency in a sample of 143 Serbian local self-governments for year 2012. The input side is described by current expenditures of the local self-governments, and the outputs are describing key public services provided. Our results reveal that the average LGS in the Republic of Serbia "produces" public services i.e. output at costs that are between 21 and 23% higher than the 'best practice' once we control for the exogenous constraints on public service provision. Results show a modest to strong reduction in measured inefficiency once controlling for exogenous constraints illustrates the adequacy of our approach. In general, our findings are in line with results from other countries, though direct comparison is limited due to various heterogeneity issues. Finally, our results provide strong support for the amalgamation of the smallest municipalities.

There are several ways that could improve our findings. First, there are a number of other determinants that may influence the efficiency of local governments. For example, political characteristics of a local governments

such as political fragmentation, electoral cycles, or fiscal variables such as revenues structure, fiscal independence or soft budget constraints. Due to lack of data, we refer these issues for a future research agenda and further analysis is necessary before definitive conclusions can be made. Secondly, collecting additional data to account more precisely labor cost disparity across LSGs to proxy for the overall cost difference could improve our results. However, at the moment, due to lack of data, we were unable to fully take into account the possible heterogeneity of factor costs. Third, our research is based on a cross-section approach. However, currently due to data issues, a panel approach does not represent a viable option. Finally, as this is the first study to examine the efficiency of LSGs in Serbia it would be beneficial to obtain and compare the results on the estimated efficiency levels using other approaches.

References

Afonso, A. and Fernandes, S. (2006). Measuring Local Government Spending Efficiency: Evidence for the Lisbon Region. Regional Studies, 40(1), 39–53.

Afonso, A. and Fernandes, S. (2008). Assessing and Explaining the Relative Efficiency of Local Government. Journal of Socio-Economics, 37, 1946–1979.

Athanassopoulos, A.D. and Triantis, K.P. (1998), "Assessing aggregate cost efficiency and the related policy implications for Greek local municipalities", INFOR, Vol. 36 No. 3, pp. 66-83.

Balaguer-Coll, M.T., Prior, D. and Tortosa-Ausina, E. (2013). "Output complexity, environmental conditions, and the efficiency of municipalities", Journal of Productivity Analysis, Vol. 39 No. 3, pp. 303-324.

Battese, G.E. and Coelli, T.J. (1995), "A model for technical inefficiency effects in a stochastic frontier production function for panel data", Empirical Economics, Vol. 20 No. 2, pp. 325-332.

Benito, B.,Bastida, F, and Garcia, J.A. (2010). "Explaining Differences in Efficiency: An Application to Spanish Municipalities". Applied Economics 42(4), 515-528

Boetti, L., Piacenza, M. and Turati G. (2012). "Decentralization and Local Governments' Performance: How Does Fiscal Autonomy Affect Spending Efficiency? ". Finanz Archiv / Public Finance Analysis, Vol. 68 (3) 269-302.

Coelli, T.J. (1996), "A guide to Frontier Version 4.1: a computer program for stochastic frontier production and cost function estimation", Working Paper No. 96/07, Centre for Efficiency and Productivity Analysis, The University of Queensland, Armidale.

Coelli, T.J., Prasada Rao, D.S., O'Donnell, C. and Battese, G.E. (2005), "An Introduction to Efficiency and Productivity Analysis", 2nd ed., Springer Science & Business Media.

De Borger, B. and Kerstens, K. (1996). Cost Efficiency of Belgian Local Governments: A Comparative analysis of FDH, DEA, and econometric approaches. Regional Science and Urban Economics, 26, 145–170.

Fiscal Council (2013) “Assessment of Local Government Finances in 2013”.

Geys, B., Heinemann, F. and Kalb, A. (2008). “Local Governments in the Wake of Demographic Change: Economies of Scale in German Municipalities, Finanzarchiv, Vol. 64 (4), 434-457.

Geys, B., Heinemann, F. and Kalb, A. (2010), “Voter involvement, fiscal autonomy and public sector efficiency: evidence from German municipalities”, European Journal of Political Economy, Vol. 26 (2), 265-278.

Geys, B. Heinemann, F. and Kalb. A. (2013). “Local government efficiency in German municipalities“, Raumforschung und Raumordnung, 71(4), 283–293

Geys, B. and Moesen, W. (2009a). Exploring Sources of Local Government Technical Inefficiency: Evidence from Flemish Municipalities. Public Finance and Management, 9(1), 1–29.

Geys, M. and Moesen, W. (2009b). “Measuring Local Government Technical (In) Efficiency an Application and Comparison of FDH, DEA, and Econometric Approaches“. Public Performance & Management Review, 32 (4), 499–513.

Gimenez, V. M. and Prior, D. (2007). Long- and Short-Term Cost Efficiency Frontier Evaluation: Evidence from Spanish Local Governments. Fiscal Studies, 28(1), 121–139.

Greene. W. (2008). “The econometric approach to efficiency analysis“, in H. O. Fried, C. A. K. Lovell, and S. S. Schmidt, eds., The Measurement of Productive Efficiency and Productivity Growth, Oxford: Oxford University Press, 92–250.

Kalb, A., Geys, B., Heinemann, F. (2012). “Value for money? German local government efficiency in a comparative perspective“. Applied Economics, 44 (2): 201–218

Kalb, A. (2010). The Impact of Intergovernmental Grants on Cost Efficiency: Theory and Evidence from German Municipalities. Economic Analysis and Policy, 40(1), 23–48.

Kecman M, (2010), Koncept finansiranja lokalne samouprave u Republici Srbiji, Industrija vol. 38, br. 3155-3175.

Krüger, J. (2012), “A Monte Carlo study of old and new frontier methods for efficiency measurement”, European Journal of Operational Research, Vol. 222 No. 1, 137–148.

Kumbhakar S. C., Wang, H., A. Horncastle, (2015) “A Practitioner's Guide to Stochastic Frontier Analysis Using Stata”, Cambridge University Press.

Loikkanen, H. and Susiluoto, I. (2005). Cost Efficiency of Finnish Municipalities in Basic Service Provision 1994–2002. Urban Public Economics Review, 4, 39–63.

Lo Storto, C. (2013), “Evaluating technical efficiency of Italian major municipalities: a data envelopment analysis model”, Vol. 8, pp. 346-350.

Pevcin,P. (2014),"Efficiency levels of sub-national governments: a comparison of SFA and DEA estimations", The TQM Journal, 3, 275 – 283.

Prud’homme, R. (1995). The Dangers of Decentralization. World Bank Research Observer, 10, 201–220.

Seifert, S. and Nieswand, M. (2014) “Some Determinants of Intermediate Local Governments’ Spending Efficiency: The Case of French Départements“. Local Government Studies, Vol. 40(5), pp.766-790.

Šťastná, L. and Gregor, M. (2011), "Local government efficiency: evidence from the Czech municipalities", Working Paper No. 14, Institute of Economic Studies, Prague.

Šťastná, L. and Gregor, M. (2015). "Public sector efficiency in transition and beyond: evidence from Czech local governments". Applied Economics, 47(7), 680-699.

World Bank (2014) "Municipal Public Finance Review Options for Efficiency Gains", Report no. 87944-YF.

Worthington, A.C. (2000). "Cost Efficiency in Australian Local Government: A Comparative Analysis of Mathematical Programming and Econometric Approaches", Financial Accountability and Management, Vol. 16, No. 3, pp. 201-223.