Efficiency of the Portuguese metros. is it different from other European metros?

Santos, Joana and Simões, Pedro and Costa, Álvaro and Cunha Marques, Rui

Center for Management Studies (CEG-IST), Technical University of Lisbon

2010

Efficiency of the Portuguese Metros. Is it Different from other European Metros?

Joana Santos

IST, Technical University of Lisbon

E-mail: cjoanasantos@gmail.com

Pedro Simões

Center for Management Studies (CEG-IST), Avenida Rovisco Pais, 1049-001 Lisbon, Portugal

Tel: +351 218417729; Fax: +351 218417979; Email: pedrotsimoes@ist.utl.pt

Álvaro Costa

School of Engineering, University of Porto, Rua Dr. Roberto Frias, 4260-465 Porto, Portugal

E-mail: afcosta@fe.up.pt

Rui Cunha Marques

Center for Management Studies (CEG-IST), Avenida Rovisco Pais, 1049-001 Lisbon, Portugal

Tel: +351 218417729; Fax: +351 218417979; E-mail: rui.marques@ist.utl.pt)

ABSTRACT

This research analyses the performance of Portuguese metros in the European context. By means of two non-parametric benchmarking techniques, respectively performance indicators and data envelopment analysis, we compute the efficiency of 37 European metros. In order to provide statistical inference and robustness to our results we apply the recent technique of bootstrap. We also use the partial frontiers (order-m) to identify outliers and the double bootstrap procedure in a second stage methodology to take into account the influence of the operational environment. The results show important levels of inefficiency both in the Portuguese metros and in other European metros.

Keywords: Metro, efficiency, Portugal, performance indicator, data envelopment analysis

1. INTRODUCTION

Nowadays metro systems represent an element of social cohesion and sustainable urban mobility, having gained an increasing importance as an alternative to private transportation in urban centers. However, this transportation system is, in general, provided in a monopolistic environment with few incentives to be efficient and innovative. Like this, measuring the performance of metros and the application of benchmarking are key tasks to provide the value for money of this service. The current research examines the performance of metros in Portugal in a set of 37 European metros.

According to ERRAC (2004), a metro system must meet four conditions: it must be electrically driven; move in a dedicated lane (surface, elevated or underground) separating itself from other traffic; present high frequencies and deal with large flows of passengers per hour and direction.

Based on those principles, the *Union Internationale des Transports Publics (UITP)* classifies metro systems into two types: a) conventional metro, that is, a tracked, electrically driven local means of transport, which has an integral, continuous track bed of its own (large underground or elevated sections); and b) light rail, that is, a tracked, electrically driven local means of transport, which can be developed step by step from a modern tramway to a means of transport running in tunnels or above ground level. This broad definition encompasses a wide range of situations, from conventional tramway to tram-train solutions.

In 2006, according to UITP, Europe had about 200 metro systems, encompassing a wide range of solutions. Of those, only 36 systems were classified as conventional metro and other 18 were classified as hybrid systems (running in a dedicated track partially underground), making a total of 54 systems, which were the subject of this review.

So, beyond the research over such important service for great urban areas, the major objective of this study is to measure the performance of Portuguese metros in the European context. To fulfill this aim, using a sample of 37 metro systems, representing 18 European countries, diverse benchmarking techniques, such as performance indicators, the non-parametric technique of data envelopment analysis (DEA), the robust non-parametric approaches of bootstrap and of order-m, and the recent methodology of double-bootstrap were applied. The latter intends to include the explanatory factors in the analysis that may influence the results obtained.

The current study represents a relevant contribution for the literature, mainly due to the lack of research that characterizes the metro system, in general, and the performance evaluation of metro service in particular. It uses for the first time (as far as we know) the concepts of bootstrap, order-m and double bootstrap in the non-parametric efficiency analysis to compute the efficiency of the European metros. This recent idea overcomes one of the major problems of the DEA method which is its deterministic nature. Furthermore, to give robustness to the results obtained we adopted the bootstrap tool and identified the outliers of the sample through the order-*m* approach. Among the contributions of the paper there is also the uniqueness of the data, since it is the only study that includes a comparison just between metros and not between (general) railway services (for instance Graham, 2008, which compares 200 urban railways), which somehow skews the possible conclusions drawn about this sector.

The current article is organized as follows. After this brief introduction, the second section provides some ideas about the matter of benchmarking in the infrastructure services. The third section presents an explanation of the different approaches used to evaluate the performance of European metros. Section four presents and analyses the results, identifies the outliers and takes into account the influence of the operational environment. Section 5 provides some policy implications. Finally, the sixth section draws the concluding remarks.

2. BENCHMARKING

Leonard (2001) defined benchmarking as a process, which seems quite enlightening. It is the means by which we try to achieve a superior level of performance, in a particular area, changing current practices in the company, leading to improvements in its performance. Benchmark is a standard of excellence, the basis of comparison to similar results. Following the best practices is the way of achieving the maximum level of performance (benchmark).

The benchmarking process can be developed inside or outside the company. It can be a systematic performance comparison between departments of the same company (internal benchmarking) or it can compare the performance of one company with other organizations or competitors in the sector (external benchmarking).

The cyclic process of benchmarking can be described in 9 steps, as shown in Figure 1, adapted from Hanman (1997). The first 4 steps are the scope of this research. Generally, 3 levels of benchmarking can be defined in 3 increasing degrees of commitment and cooperation, as described in Table 1 (EQUIP, 2000).

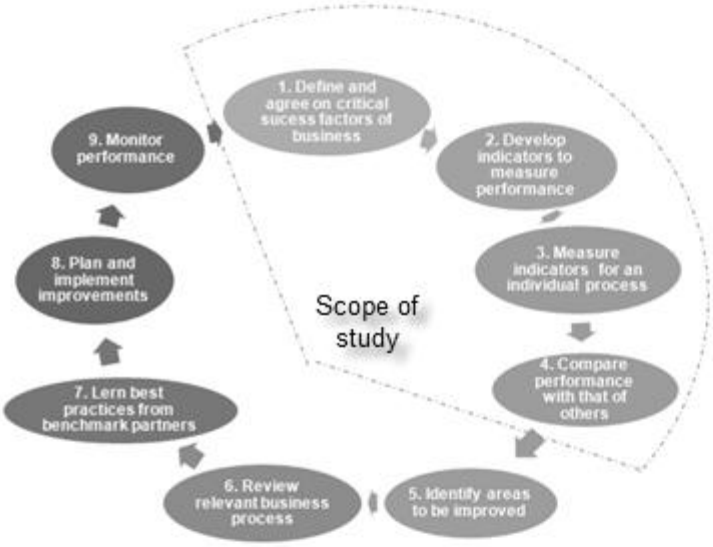


Figure 1 - The continuous improvement process of benchmarking – a nine stage model

Table 1 - Benchmarking levels

Level	Category	Elements
I	Self-Assessment	• Measure your own company's performance (over time);
II	Comparison	• Compare your performance with database of anonymous indicator value; • Identify improvement areas and best "standards";
III	Partnering	• Work with relevant partners, perhaps with some outside the direct business sector; • Exchange confidential information; • Learn best practices and the means of implementing the necessary changes; • Ideally, this should be a two-way process.

The process of external benchmarking between operators, corresponding to the level III, is not common. The main reasons for this are confidentiality, lack of efficient tools to identify comparable practices and a remarkable resistance to information dissemination. In spite of this, there are several organizations, groups such as Comet, NOVA, ALAMYS, UITP, EMTA, ERRAC, among others, that have carried out several tasks regarding benchmarking projects on public transportation, especially on metros. In this context, the European Union has supported various projects, such as BEST, BOB and the MODUrban.

Despite the various benchmarking studies published about public transportation, the majority does not focus on one specific means of transport, providing only a comparison between public and private transportation. Benchmarking studies focused on metro publicly available are rare and therefore more valuable. Nevertheless, their use was very helpful to this research and for that reason the works from Costa (1998) and Frasquilho (2005) must be pointed out.

3. PERFORMANCE EVALUATION OF METRO SYSTEMS

3.1 Introduction

In this study, as mentioned, two different non-parametric techniques were used to evaluate metro operators' performance, such as performance indicators and DEA. To fulfill this aim, 206 metro systems in operation in Europe were identified for the year 2006. Out of the 206 metros systems (including heavy conventional metro and light rail solutions), 54 systems (36 heavy conventional metro plus 18 hybrid solutions) were selected from 25 countries, but we only got data from 37 metro systems.

3.2 Performance indicators

One of the first advantages of performance indicators relates to the need for collection and compilation of information from the organization. This measure requires first and foremost a self-knowledge on the company's activity, which justifies, by itself, its implementation. Furthermore, the determination of performance indicators allows for more awareness and a proactive management (Jeon and Amekudzi, 2005). If one compares the figures recorded in different organizations it can even be as stimulus to improve the company's performance.

The computation of performance indicators begins with the indicator's definition and calculation formula or measurement criteria. After being calculated, it follows the discussion of the results. Their interpretation should include the analysis of the explanatory factors. These factors attempt to justify or reflect on results validity, based on data directly or indirectly related to the perofrmance indicator and can be classified into controllable or non-controllable (Witte and Marques, 2009). In the first group, the controllable factors, are all factors that are in some way under the operator action (for example ageing), and in the second one, those which the operator does not have intervention (for example population density or GDP *per* capita).

After the interpretation of results, the next step is the comparison with the benchmarks. These values, considered as appropriate, are associated with current practices in the sector, that is, the average values (Marques and Monteiro, 2001). In this analysis, a margin of 5 per cent was applied to the average values, resulting in a band of values corresponding to the operators with acceptable performance. Operators with superior performance are, in case where the objective is to maximize the ratio, above the reference band (for example passengers carried *per* employee), or vice versa, below the reference band when the goal is to minimize the indicator (for example operating costs *per* employee).

This research proposes a framework of performance indicators deemed to be appropriate to review a metro's performance organized into 6 groups, namely supply and demand, human resources, quality of service, efficiency, effectiveness and economic and financial indicators. The list of these performance indicators is presented in Table 2.

Each of the 6 groups of performance indicators have particular goals. The first group attempts to assess the levels of supply and demand experienced in each metro system. The performance of a transportation system is intimately connected to these two factors, more specifically, to the ability of adjust supply to demand. The second group intends to evaluate the performance of the current body of the organization, discussing, among other factors, the rate of absenteeism. Staff training is also addressed, with consequences on the quality of service and implications for the ability to work. The third group concerns the quality of service. This includes several factors, ranging from issues of accessibility, security, environment, realibility to customer satisfaction. The fourth group includes the efficiency indicators which are divided into three categories, allowing to examine the production levels (for example perfomance ratios of vehicle-kilometres and journeys), the levels of labor productivity (for example ratios that correlate the vehicle-kilometres and seat-kilometres produced to the staff) and the levels of capital (ratios which relate vehicle-kilometres and seat-kilometres produced to vehicles, stations or the network lenght). The fifth group corresponds to the effectiveness indicators, divided into three groups, respectively levels of occupation, labor effectiveness and capital effectiveness. Among the indicators concerning the occupancy levels, two measures of occupation (one absolute and other relative) can be distinguished. The efficiency and effectiveness ratios are usually expressed by linking the volume of passengers or the passenger-kilometres with vehicles, network lenght, stations or with staff. The last group focuses on economic and financial indicators, including ratios related to revenues, costs and other results.

Although performance indicators allow a quick and accessible reading of an operators' activity, there are some limitations associated with their results analysis. The major flaw is related to the fact of being partial measures of productivity that relate only one of the production factors (input) and one of the results (outputs) that the organization 'consumes' and 'produces'. Furthermore, the application of performance indicators in the analysis does not include the operational and institutional environment (Witte and Marques, 2009). The diverse explanatory factors (population, GDP, among others), despite influencing and explaining in some way the results for the different indicators, are difficult to be directly correlated to each indicator individually.

Table 2- Proposed framework of performance indicators

Performance indicators

Supply and Demand:

-Seats available per carriage
-Seats available per train
- Average distance travelled by each passenger in the network
-Passengers per capita
-Passengers per vehicle-km
-Passengers per seat-km

Human Resources:

-Workshop hours by employee
-% Non qualified professionals
-% Qualified professionals
-% Higher qualified professionals
-% Junior workers
-% Intermediate workers
-% Senior workers
-Absenteeism rate
-No. Annual Hours worked by employee

Efficiency:

Production Levels
-Vehicle-km made/Vehicle-km forecasted
-Train passages made/ Train passages forecasted

Human Resources' Productivity
-Vehicle-km per employee (I1)
-Seat-km per employee
-Vehicle-km per working hour
-Seat-km per working hour

Capital Productivity
-Vehicle-km per vehicle (I2)
-Seat-km per vehicle
-Vehicle-km per net km (I3)
-Seat-km per net km
-Vehicle-km per station (I4)
-Seat-km per station

Effectiveness:

Occupancy Levels
-Occupancy rate
-Passengers per vehicle

Human Resources' Effectiveness
-Passenger-km per employee (I5)
-Passenger-km per working hour

Capital's Effectiveness
-Passenger-km per vehicle (I6)
-Passenger-km per net km (I7)
-Passenger-km per station (I8)

Quality of service:

Accessibility
-Lift availability
-Escalators availability
-Presence of information systems through variable message signs with sound system

Security
-Fatalities per passenger
-Suicides per passenger
-Travel accidents per passenger
-Incidence rate of industrial accidents
-Frequency of industrial accidents

Environment
-Energy consumption per passenger-km
-Energy consumption per vehicle-km
-Energy consumption per seat-km

Reliability
-Rolling stock Availability
-Average distance between failures
-Hours between failures
-Punctuality
-Average time of delay per trip

User Satisfaction
-Complaints by passenger

Economy and Finance:

Revenues
-Traffic revenue per passenger
-Traffic revenue per passenger-km
-Operational revenues per passenger
-Operational revenues per passenger-km
-Operational revenues per seat-km

Costs
-Weight of staff costs in operational expenses
-Weight of staff costs in total costs
-Total costs per vehicle-km
-Operational expenses (OPEX) per vehicle-km
-Maintenance costs per vehicle-km
-Administrative costs per vehicle-km
-Investment expenditure per vehicle-km
-Total costs per passenger
-OPEX per passenger
-OPEX per station

Other results
-Total revenues / OPEX
-Operational Revenues/ OPEX
-Operational Revenue excluding subsidies / OPEX
-Net income/deficit for the year per passenger-km

Another issue is related to the treatment of outliers' presence (see Wilson, 1995). These atypical observations are the ones that have a large distance of the remaining sample or the ones that are inconsistent with this. In the analysis of performance indicators, the outliers concern operators whose performance is, in the various indicators, distant or out of the sample average. In such cases, we must examine what causes this disparity and whether it is justifiable or not or if they are best practices.

To circumvent all these weaknesses, other approaches of performance evaluation were computed, such as DEA, which has the advantage of dealing with several inputs and outputs, identifying the best practices and being empirically-based, even though extremely influenced by outliers, along with other three recent approaches (DEA-bootstrap, order-m and double-bootstrap). DEA-bootstrap allows for the correction of DEA efficiencies and for the statistical inference of the results (Simar and Wilson, 1998) and the order-m approach identifies the sample outliers (Casals et al., 2002 and Simar, 2003). At last, the double bootstrap procedure is used to investigate the influence of the operational environment on efficiency (Simar and Wilson, 2007).

3.3 Data Envelopment Analysis

3.3.1 Introduction

DEA is a non-parametric technique which uses mathematical programming, aimed at assessing the relative performance of decision making units (DMU), in the presence of a uniform set of multiple inputs and multiple outputs (Ozbek et al. 2010). As a non-parametric approach, instead of assuming a predefined function to the production frontier (as in parametric methods), it constructs the frontier by the best practices observed in the available sample. It does not need, therefore, a prior specification of the weights of each input/output, neither does it require judgments on the production function form. The efficiency of each DMU (metros in this case) is measured by the distance between the DMU and the efficient frontier (created by the best practices).

The elementary DEA model using constant returns to scale (CRS) and strong disposability of inputs was developed by Charnes, Cooper and Rhodes in 1978 (Charnes et al., 1978) based on the previous studies of Farrell in the 50's (Farrell, 1957). Algebraically, the problem of DMU efficiency (h_k) can be stated as follows:

$$Max : h_k = \frac{\sum_{i=1}^{l} a_i y_{ik}}{\sum_{i=1}^{l} b_j x_{jk}} \tag{1}$$

subject to

$$\frac{\sum_{i=1}^{l} a_i y_{ik}}{\sum_{i=1}^{l} b_j x_{jk}} \leq 1 \quad , m = 1,\ldots,k,\ldots,M$$

$$a_i, b_j > 0 \quad , i = 1,\ldots,l;\ j = 1,\ldots,J$$

where,

y_{ik} - output i of DMU k; x_{ik} - input j of DMU k; a_{i} - output i weight; b_{i} - input j weight; M - number of DMUs; l - number of outputs; J - number of inputs.

This approach has the underlying principle of CRS, which does not always correspond to the reality of the industries studied. Later, Banker et al. (1984) introduced the possibility of variable returns to scale (VRS), where the aggregate or overall efficiency of a DMU can be decomposed into two components, respectively the pure technical efficiency and the scale efficiency. Scale efficiency determines the degree of savings that would occur if the DMU was operating at the optimal scale. Scale efficiency is computed by comparing the technical efficiency obtained by the models of DEA assuming CRS and VRS.

3.32 Model specification and data collection

In this research, at a first stage, the efficiency of European metros was measured applying the traditional DEA model. Within this, we developed a model that encompasses as inputs the net length, the number of vehicles and the number of staff and as outputs the passengers transported and the number of vehicles *per* km. It should be noted that the selection of inputs and outputs were considered as the best to characterize the dynamics of the industry and was constrained by the sample and data available. Figure 2 presents the specification of the model adopted.

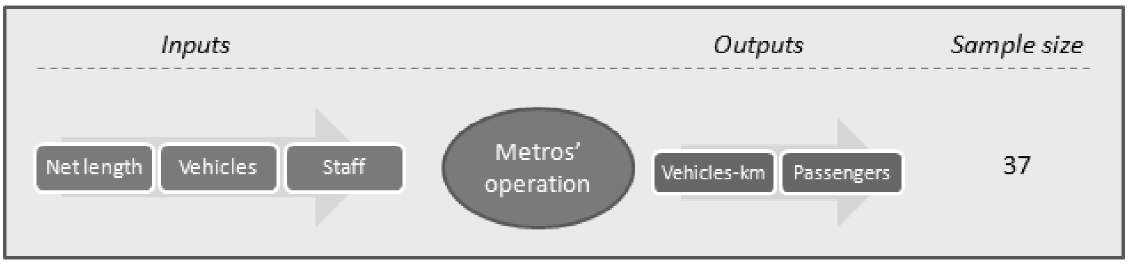


Figure 2 - Model specification

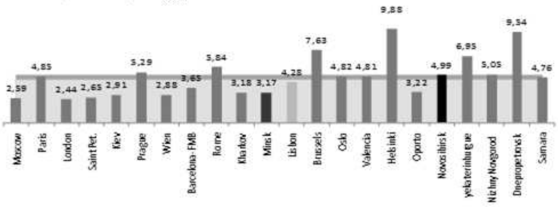
All data were collected directly from operators either through their annual account reports or by telephone contacts with the metro’s office. Regarding the orientation of the DEA model, we adopted an input minimization orientation, since in metro service we considered the provision of a public service and the objective of minimizing the production factors more important than profits and expansion of outputs.

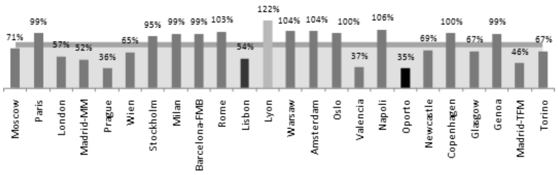
4. ANALYSIS OF THE RESULTS

4.1 Performance indicators application

As an example of the performance indicators methodology application, Figure 3 presents the results of two performance indicators, encompassing 35 metro operators, corresponding to 18 European countries. A "fact sheet" was made for each indicator identifying and characterizing them with a graph of benchmarking which sets the values observed for the various operators. The range where it is considered that the operator has an acceptable performance (range of reference) appears in a light color, and the Portuguese metros are given in black. Moreover, Table 3 provides a sumary of the scores of other 8 performance indicators which had been identified in Table 2.

Indicator: Energy consumption (Kwh) per Vehicle-km
Reference value (benchmark): 4.79 Kwh/vehicle-km





Indicator: Operational revenues / Operational expenses (%)
Reference value (benchmark): 78.5%

Figure 3 - Example of performance indicators

Table 3 - Best and worst practices for indicators of efficiency and effectiveness (see table 2)

Rank	I1	I2	I3	I4	I5	I6	I7	I8
1st	Helsinki	Frankfurt	Moscow	Moscow	Helsinki	St. Pet.	Moscow	Moscow
2nd	Berlin	Moscow	St. Pet.	St. Pet.	Milan	Moscow	St. Pet.	St. Pet.
3rd	London	London	London	London	Moscow	Helsinki	Paris	Paris
33th	Genova	Genova	Valencia	Valencia	Dnepro.	Glasgow	Valencia	Valencia
34th	Samara	Stockholm	Oporto	Oporto	Samara	Samara	Samara	Glasgow
35th	Brussels	Glasgow	Stockholm	Genova	Yeka.	Yeka.	Oporto	Oporto

Concerning the outliers' identification, the direct analysis of the performance indicators benchmarking figures suggests three possible outliers, namely Moscow, St. Petersburg and London. Anyway, to ascertain the reliability of these conclusions it is necessary to analyze further these operators using other statistical methods, and to know them better in detail.

4.2 Efficiency measurement

4.2.1 DEA

The model was designed towards the minimization of inputs. Figure 4 illustrates the technical efficiency and the potential gains from scale economies to 37 operators of metros in Europe. Table 4 presents the traditional statistics of DEA results.

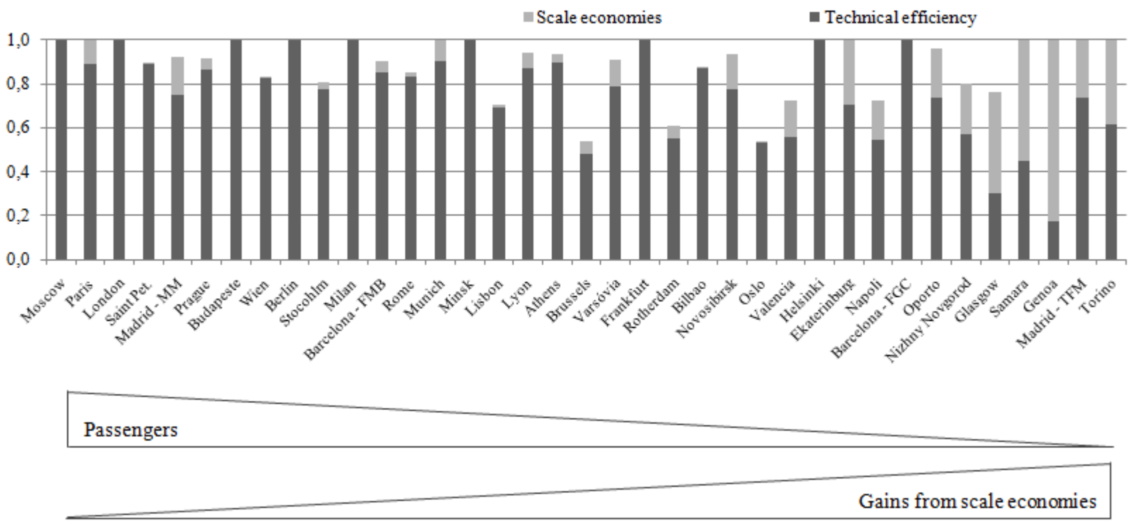


Figure 4 - Technical efficiency and potential scale economy gains

The previous figure shows that there are scale economies in passenger transportation, generating greater potential gains in efficiency for the operators with a lower volume of passengers.

From Table 4, it is possible to observe that if European metros operated efficiently they would have conditions to save about 23 per cent of the inputs consumed (providing the same level of outputs). From this, 14 per cent are associated with their inappropriate operation scale. As it is presented in the table, the sample shows, in general, increasing returns to scale (IRS). Only 7 metros have decreasing returns to scale.

Table 4 - Statistic summary of the DEA results

Variables	CRS	VRS	Scale efficiency	Returns to scale
Mean	0.769	0.894	0.862	
Str. Dev.	0.212	0.136	0.193	9 CRS
Median	0.827	0.933	0.942	7 DRS
Min.	0.172	0.531	0.172	21 IRS
Max.	1.000	1.000	1.000	

Figure 5 presents the efficiency of the Portuguese metros, where it is possible to observe that they have quite similar performances when compared to those in other European countries. Table 5 shows the targets of Lisbon and Oporto metro operators, assuming CRS. In terms of peers for the Portuguese metros, the benchmarks identified are Moscow, Helsinki, Barcelona-FMB and Berlin metros.

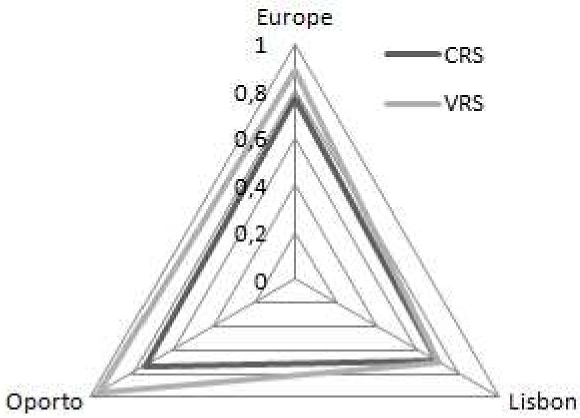


Figure 5 - Performance of Portuguese metro operators in European context

Table 5 - Targets of metro operators of Lisbon and Oporto, considering CRS

Variables	Lisbon		Oporto	
	Actual	Target	Actual	Target
Net length	35.60	35.60	58.88	58.88
Vehicles (no.)	338	338	72	72
Staff (no.)	1,702	1,179	426	314
Peers	Berlin, Helsinki, Budapest, Moscow		Helsinki, Budapest, Moscow	

Table 6 provides a comparison between light and conventional metros and between public and private management. The analysis shows analogous performance of light and conventional metros, and that private metros are slightly more efficient than the public ones. However, the results are not statistically significant.

4.2.2 DEA-bootstrap approach

DEA is characterized by not allowing statistical inference. Opposing this view, Simar and Wilson (1998 and 2000) developed a new non-parametric methodology, more robust, which uses a bootstrap smooth algorithm, based on a data generating process (DGP). The basic idea of bootstrapping is simply to simulate the sampling distribution of interest by mimicking the DGP. DGP follows the principle that restricted to the relationship between inputs and outputs, the stochastic elements in the productive process are totally encompassed by the random inputs efficiency measures. The procedure for the algorithm in each re-sample follows two phases. In the first one, the inputs frontier is estimated and the bootstrap pseudo-inputs are created through the DGP application at the estimated frontier of inputs and pseudo-efficiencies. The algorithm makes use of a smoothed re-sampling procedure**,** based upon the consistency argument. In the second phase, the bootstrap efficiency estimate is obtained by accounting the distance of the original input from the bootstrap frontier estimate (for more details see Simar and Wilson, 1998 and 2000).

Table 6 - Targets of metro operators of Lisbon and Oporto, considering CRS

	Light metro	**Conventional metro**	**Public**	**Private**
Mean	0.438	0.607	0.591	0.595
Str. Dev.	0.278	0.170	0.246	0.178
Median	0.438	0.617	0.591	0.617
Min.	0.242	0.320	0.417	0.242
Max.	0.635	0.856	0.765	0.856
Mean	0.438	0.607	0.591	0.595
Lisbon	-	0.707	0.707	-
Oporto	0.962	-	0.962	-

Given the fact that DEA-bootstrap is, in general, more robust than the simple DEA model, the development of rankings is more reliable with this methodology. Simar and Wilson (2000) provide a test to prove this. The results are presented in Table 7. Figure 6 shows the confidence interval obtained for the corrected efficiencies of DEA-bootstrap.

Table 7 - Rankings from DEA and DEA-bootstrap models

Ranking	**DEA-VRS**		**DEA-bootstrap-VRS**	
	Metro system	Efficiency	Metro system	Efficiency
1	Turin	1.000	Frankfurt	0.922
2	Madrid-TFM	1.000	Barcelona-FGC	0.916
3	Barcelona-FGC	1.000	Samara	0.913
…	…	…	…	…
33	Valencia	0.721	Napoli	0.680
34	Lisbon	0.707	Lisbon	0.665
35	Rotterdam	0.608	Rotterdam	0.572
36	Brussels	0.539	Brussels	0.509
37	Oslo	0.531	Oslo	0.495

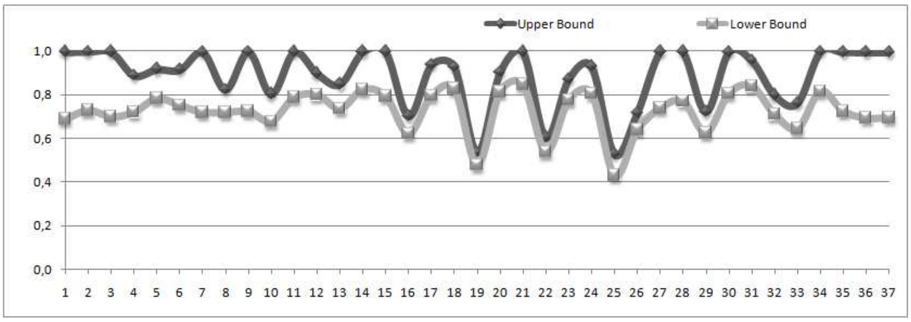


Figure 6 - Confidence intervals of DEA-bootstrap

4.3 Outlier detection

There are several procedures able to deal with the issue of outliers' presence (Witte and Marques, 2009). However, we opted here to apply the recent non-parametric parametric approach of order-m. Basically, this methodology, being a partial frontier analysis, since it compares each DMU with *m* DMU benchmarks, allows for getting efficiencies greater than 1, so, above a certain level, an outlier can be identified (for more details see Simar, 2003). The results are presented in Table 8.

Table 8 - Order-m results

M=15		**M=20**		**M=25**	
DMU	EFF	DMU	EFF	DMU	EFF
London	1.700	Paris	1.557	Paris	1.449
Paris	1.686	Moscow	1.517	London	1.396
Moscow	1.678	London	1.504	Moscow	1.349

Considering these results, the metros of Paris, Moscow and London are strong candidates to be outliers, corroborating, globally, the performance indicators results.

4.4 Operational environment analysis

In order to incorporate the analysis of the operational environment, this research also applies a second stage methodology, labeled as double-bootstrap. Following this methodology, a (semi-parametric) regression analysis (after the DEA-VRS model in the first stage) is carried out to determine the influence of environmental variables on the bias-corrected efficiency scores (Simar and Wilson, 2007).

In this case, the study encompasses the Gross Domestic Product (GDP) by region, the population of the cities, the number of metro stations and the net length. The results are presented in Table 9.

From these results it is possible to conclude that that GDP and population have a positive influence on metros' efficiency. Concerning the number of stations and the net length, they have an opposite influence. However, GDP and net length variables are not statistically significant.

Table 9 - Double bootstrap results

Variables	Estimate	Lower bound	Upper bound	t-value
Intercept	1.4763	1.2086	1.7169	168.4554
Stations (no.)	0.0019	-0.0020	0.0058	12.8693
Net (km)	0.0002	-0.0042	0.0045	1.2151
GDP *per capita*	-0.0001	-0.0018	0.0015	-0.921
Population	-0.0001	-0.0002	0.0000	-29.5056
St. Deviation	0.0545	0.0278	0.0947	42.8674

5. POLICY IMPLICATIONS

In general, metro operators show a large dependence on operating subsidies. The fact that they are not valued monetarily neither contracted leads to the allocation of subsidies randomly, granted according to the financial availability of the State, which usually results in an insufficient value in relation to investments, forcing the bank debt.

To control the indebtedness of the industry and ensure an improved quality of service, it is necessary to promote measures to improve the effectiveness and efficiency. This consists firstly in matching supply to demand, by increasing the supply in times of increased demand by users and cutting off those periods.

Regardless of whether or not entrance metropolitan transport authorities are set, the large volumes involved in the financing of the system require the definition of new strategies and funding schemes. Operators must be able to attract higher levels of revenue through alternative means of revenue from traffic, should be eligible for more funds, coming, for instance, from taxes on companies that benefit directly from the metro network, as in France through the *Versement du Transport*. So that the need for funding is the lowest possible, it is necessary to reduce the operational costs of the system. This can be achieved in different ways, including a reduction of the operator's staff and expecting increased productivity through training and qualification of employees. Monitoring the average age of the fleet also helps to reduce the costs of maintenance and increase the levels of reliability (lower number of failures, greater punctuality).

Information about the performance of metro systems is very scarce. Although, there are some sector associations that apply benchmarking and carry out studies on benchmarking, among metro operators the results are not publicly available. The pressure of stakeholders on operators in particular of media, users and political forces could bring into the light positive effects. The bad performers would stay embarrassed and feel the need to correct their deviations in the future and the good ones would have additional incentives to continue outperforming. As this policy of “name and shaming” is very successful in different sectors some authors propose the implementation of a European Observatory whose major functions will be the comparison of performance and the sharing and publicizing information of the sector (Marques and Brochado, 2008). Even acknowledgment the local nature of metro systems a creation of a body of this type could be a good decision in the European context.

6. CONCLUSIONS

The current paper evaluated the performance of the two Portuguese metros in the European context. To fulfill this aim non-parametric benchmarking methodologies were applied, such as performance indicators, DEA and DEA-bootstrap, using a sample of 37 metros from 18 countries for the year 2006. Moreover, the study proposed to analyze critical issues like the presence of outliers and the influence of operational environment, employing the methods of order-m and double-bootstrap, respectively.

From the analysis of performance indicators we can globally observe which is the standard of Portuguese metros in Europe. It highlighted some inefficiency for the Portuguese ones. For instance, regarding the operational costs, we verified that the average rate of coverage of the operating costs by operating income is around 78.5 per cent. Regarding the Portuguese situation, we noticed that the revenues, on average, do not cover more than 50 per cent of the operational costs. This results is emphasized by the poor overall performance of the Oporto metro, in both efficiency and effectiveness indicators, although we must bear in mind its recent start (2003). This analysis also stood out the good performance of the London, Russian and St. Petersburg metros, although possibly they are outliers.

Through the DEA technique results, as expected, globally corroborated with the conclusions drawn from the performance indicators analysis. The Portuguese metros still reveal some problems. However, through the model encompassed, Lisbon show worse results than the Oporto metro. In global terms, the European metros show relevant inefficiencies, from which, considering the high expenses of the European States on this sector, several millions of Euros could be saved if they operated efficiently.

The computation of the bootstrap approach allowed us to develop more robust ranking, where Frankfurt, Barcelona-FGC and Samara revealed the best scores. The computation of order-m highlighted the metros of Paris, Moscow and London as possible outliers in our sample. Finally, the analysis of the operational environment showed that GDP and population have a positive influence on metros efficiency, in opposition of the number of stations and the net length.

REFERENCES

Banker, R., Charnes, A. and Cooper W. (1984). "Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis', *Management Science*, 30(9), 1078-1092.

Benito, B., Bastida, F. and Garcia, J. (2005). "*The Determinants of Efficiency in Municipal Governments"*, Fundación Séneca of the Autonomous Community of the region of Murcia, Murcia.

Cazals C., Florens, J. and Simar, L. (2002). "Nonparametric Frontier Estimation: A Robust Approach", *Journal of Econometrics*, 106(1), 1-15.

Charnes, A. Cooper, W. and Rhodes, E. (1978). "Measuring the Efficiency of Decision Making Units", *European Journal of Operational Research*, 2(6), 429-444.

Costa, Á. (1998). "Public Transport Efficiency and Effectiveness: Metro de Madrid", in Button, K., Nijkamp, P., Priemus, H., (ed.) *Transport Networks in Europe: Concepts, Analysis and Policies*, Edward Elgar Publishing, pp. 248-264.

EQUIP-Consortium (2000). "*The Benchmarking Handbook"*. Deliverable D4. Brussels and NewCastle upon Tyne: European Commission DG-TREN and EQUIP Consortium.

ERRAC (2004). "*Light Rail and Metro Systems in Europe - Current market, perspectives and research implication"*. Brussels.

Farrell, M. (1957). "The Measurement of Productive Efficiency", *Journal of the Royal Statistical Society*, 120(3), 253-281.

Frasquilho, M. (2005). "*Eficiência dos Transportes Urbanos e Autoridades Metropolitanas de Transporte"*. Grupo Banco Espírito Santo, Espírito Santo Research.

Graham, D. (2008). "Productivity and Efficiency in Urban Railways: Parametric and Non-Parametric Estimates", *Transportation Research Part E*, 44(1), 84-89.

Hanman, R. (1997). "Benchmarking your Firm's Performance with Best Practice", *International Journal of Logistics Management*, 8(2), 1-8.

ISOTOPE (1998). "*Improved Structure and Organization for Urban Transport Operations of Passengers in Europe"*. Office for Official Publications of the European Communities, Luxembourg.

Jeon, C. and Amekudzi, A. (2005). “Addressing Sustainability in Transportation Systems: Definitions, Indicators and Metrics”. *ASCE Journal of Infrastructure Systems*, 11(10), 31-50.

Marques, R. and Monteiro, A. (2001). “Application of Performance Indicators in Water Utilities Management - A Case-Study in Portugal”, *Water Science and Technology*, 44(2-3), 95-102.

Marques, R. and Brochado, A. (2008). “Airport Regulation in Europe. Is There Need for a European Observatory?”. *Transport Policy*, 15, 163-172.

Ozbek, M., Garza, J. and Triantis, K. (2010). “Data and Modeling Issues Faced during the Efficiency Measurement of Road Maintenance Using Data Envelopment Analysis”. *Journal of Infrastructure Systems*, 16(1), 21–30.

Simar, L. (2003). “Detecting Outliers in Frontier Models: A Simple Approach”, *Journal of Productivity Analysis*, 20, 391-424.

Simar, L. and Wilson, P. (1998). “Sensitivity Analysis of Efficiency Scores: How to Bootstrap in Nonparametric Models”, *Management Science*, 44(1), 46-61.

Simar, L. and Wilson, P. (2000) “A General Methodology for Bootstrapping in Nonparametric Frontier Models”, *Journal of Applied Statistics*, 27(6), 779-802.

Simar, L. and Wilson, P. (2007). “Estimation and Inference in Two-Stage, Semi-Parametric Models of Production Processes”, *Journal of Econometrics*, 136(1), 31-64.

Wilson, P. (1995) “Detecting Influential Observations in Data Envelopment Analysis”, *Journal of Productivity Analysis*, 6(1), 27-45

Witte, K and Marques, R. (2009). “Gaming in a Benchmarking Environment. A Non-Parametric Analysis of Benchmarking in the Water Sector”. *Water Policy* (forthcoming).

Witte, K. and Marques, R. (2009). “Influential Observations in Frontier Models. A Robust Non-Oriented Approach to the Water Sector”. *Annals of Operations Research*, 181(1), 377-392.