REGULATION, COMPETITION AND TECHNICAL EFFICIENCY IN BUS PUBLIC TRANSPORT: DISCUSSING THE CASE OF THE METROPOLITAN AREA OF SAN JOSÉ, COSTA RICA

Carlos Contreras-Montoya
Ministry of Public Works and Transport
COSTA RICA

Enilson Santos
Department of Civil Engineering
Federal University of Rio G. do Norte
BRAZIL

Rômulo Orrico Filho
Transportation Engineering Program
Federal University of Rio de Janeiro
BRAZIL

ABSTRACT

In the last two decades the relative roles of regulation and competition in promoting efficiency in urban bus industry has emerged as a central theme in both technical and academic debate. Empirical studies have concentrated in European, North American and Australian cases, as well as translog functions have been the main technical tool used in empirical analysis. In this paper, we focus on a Latin American case — San José, Costa Rica — using DEA techniques for the assessment of technical efficiency of bus service private provision under public regulation. Conclusions highlight a significant technical inefficiency level in the industry. Operators running their services efficiently are identified in order to provide industry benchmarkings.

INTRODUCTION

Beginning in the seventies, debate on efficiency of publicly or privately operated urban buses has become a major theme both in academic and technical arenas. The interest on efficiency matters regarding local bus markets gave rise in the last two decades to significant research concerning the relative roles of regulation and competition in urban bus industry. As it should be expected, empirical studies have been concentrated in European, North American and Australian cases. Furthermore, translog functions have been the main technical tool for discussing the merits of market-driven and publicly coordinated solutions in the industry development. In this paper, we focus on a Latin American case — the Metropolitan Area of San José, Costa Rica. From a methodological point of view, Data Envelopment Analysis is the methodological tool used in order to assess technical efficiency of bus service private providers operating under public regulation.

Geographical and demographic aspects of Great San José are initially described in order to characterize the environment of public transport. Following this, the existing regulatory framework is analyzed, with an emphasis on the competitive environment that regulatory design and practice should be able to induce in local public transport markets. Expected impacts of such regulation on the technical efficiency of bus operators are discussed and related to firms´ strategies addressed to reducing competitive pressures, growing market power, bettering financial results and so on.

Finally, operators´ economic performance is assessed with the application of Data Envelopment Analysis on a database containing information on inputs — fleet size, operating and administrative personal total wages, gas oil costs — and outputs measured in vehicle-km produced. The impacts of regulation, especially quantity regulation, in determining the levels of technical efficiency reached by providers is then explored and discussed.

A BRIEF DESCRIPTION OF THE METROPOLITAN AREA OF SAN JOSÉ, COSTA RICA

As in general, Latin American cities are strongly dependent upon bus public transport. Local buses are responsible by 65% to 85% of all motorized displacements in urban agglomerations, mainly due to low motorization rates (a fact related to unequal income distribution) and poor performance or absolute absence of rail transit options.

Nevertheless, the importance of local bus industry to urban economic efficiency has not been largely studied. In effect, bus industry production costs remains largely unknown although fare regulation is generally a central item in regulatory framework as well as in public and political debate regarding local bus provision. In the sense that bus public transport is a very important input to urban economic dynamics, it is a surprising fact that production cost structure and operators´ productivity and technical efficiency are not carefully analyzed yet. Furthermore, it should be highlighted that both State and incumbents are not aware of knowing in-depth about production factors interrelationships, or about impacts of different vehicular sizes or technologies, or about real economic meaning of regulatory decisions respect to financial and operational equilibrium of bus networks.

Population and territory

These conditions apply to urban transport system in San José Metropolitan Area. This is the main urban agglomeration in Costa Rica, a country with a population around 3.85 million inhabitants distributed over a little more than fifty thousand square kilometers and an yearly growth of 2%. The demographical census of 2000 registered an urbanization rate about 46% and economic surveys point out an GDP per capita around US$ 2,573 (exchange rate: US$ 1= 233 *colones*) in the fiscal year of 1996. Inflation annual rates are between 10-15% (13% in 1997, the year to which database used in this paper was constructed).

The Province of San José has a territorial extension of 4,959 km^2 — almost 10% of the country's area. The San José Metropolitan Area is inserted in the Province, occupying an area of 365 km^2. It comprises eleven *cantones* (municipalities) which have an aggregate population around 1.2 million inhabitants — or 30% respect to population of entire Costa Rica. These figures highlights the meaning of Great San José as a nucleus of population concentration: two of three urban inhabitants of Costa Rica live in this urban region with a density around 3,000 inhabitants by square kilometer. Population of Great San José showed an annual rate of growth of 2.3 % (INICEM, 1995) in the period 1970-1990, but this rate may be greater for the last decade due to immigration from neighbor countries, mainly Nicaragua. As in other Latin American cases, in Great San José the growth of population is greatly absorbed in peripheral metropolitan areas with poor urban infrastructure, except for the social dwelling big projects developed by State — it is the case of the communities of *Alajuelita*, *Hatillo* and *Pavas*, in the south.

Public Transport System

Public transport system in Great San José comprises exclusively road transport. Mainly bus lines form the network although minibuses (26 to 44 seats) and microbuses (9 to 25 seats) are also used. Irrespective of vehicle size, private operators under public concession or permission run bus routes and there is no State owned operator. In the last days of 1996, according to data from the Ministry of Public Works and Transport (*MOPT*, by Ministerio de Obras Públicas y Transportes), there were 1,038 urban buses in Great San José. These buses were distributed in 130 routes conceded to 47 private operators that registered monthly 16.65 million passengers in 226 thousand round trips. By their turn, 187 minibuses and microbuses — distributes in 30 routes — were producing monthly 49 thousand round trips and registering 1.89 million passengers.

It is worthy to note that more recently some supply of irregular services using microbuses and vans is being developed in those peripheral communities above mentioned. Also the emergence of private cars operated as taxis but without the necessary authorization may be noted. These *informal* services are being run in direct competition (on-street competition) with regulated regular buses, but they are not yet as much as to supersede regulated buses in any particular market. So, in the year 1976 for what empirical analysis is developed in this paper it can be stated that the only competitors that regulated buses were facing were the authorized taxis and private automobile.

Institutional Arrangement and Regulation

Institutional arrangement for organizing public transport (including in urban areas) in Costa Rica is based upon central Government tutelage (charged to MOPT) over an activity legally defined as a public service in the French sense. Acts that were providing legal support for public transport organization and management were, in 1996:

- Act n. 3503 (1965, May 10th) that regulates road passenger transportation,
- Act n. 5406 (1973, October 31st) that regulates taxi transport,
- Act n. 6324 (1979, May 24th) that defines road provision and administration, and

- Act n. 5930 (1976, December 13th), lately reformed by Acts n. 6249 (1978, May 2nd) and n. 6250 (1993, March 30th), which relates to traffic.

The Act 3503 establishes the public service character of passenger transport and defines the MOPT as the State organism responsible for regulation, control and monitoring the activity. The same Act defines that service provision may be delegated by the MOPT to private operators under the national normative. It is important to remark here that MOPT tutelage applies to any geography of road passenger transport, from urban to international routes. Furthermore, the Act creates a Transport Technical Commission in order to define the administrative procedures that MOPT shall follow in the delegation process. The Act 6324 enlarged the competences of Transport Technical Commission that became responsible by the entire delegation procedure, comprising operator choice, operational monitoring, delegation renewing or suspension, as well as defining fare calculation and revision.

Later, in 1996, September 5th, Act n. 7593 instituted the Regulatory Authority for Public Services that meant a real reduction of the attributions of Transport Technical Commission. After that, the Commission no more decided ultimately respect to fares and other regulatory aspects but the Regulatory Authority should approve its decisions. Completing the picture, in 2000, Transport Technical Commission was substituted by the Public Transport Council: while the Commission was a representative of MOPT, the Council is composed by representatives of MOPT and other instances of Government, private operators and users.

As it can be seen, legal situation of urban public transport in Costa Rica follows French traditional Administrative Law, although the growing importance of an independent regulatory agency means some intrusion of the *public utility* concept. Nevertheless, network planning and fare definitions remain as Government functions and private firms that awarded concession contracts face operational features.

The choice of the private operator should be ever made through a public procurement process that comprises competitive bidding for the contract. Contract duration is 7 years but it can be renewed. In exceptional situations it is possible to temporarily permit a route operation by means of a negotiation but just while a public procurement process is being developed and implemented. However, regulatory practice has not followed the formal and legal processes. The absence of clear rules regarding to tendering as well the political pressures made by incumbents have contributed to a situation where exceptional permissions are used almost as a rule and contract renewing is always granted without any formal evaluation. An exception to this may be noted in the period 1991-93 when a great quantity of routes formerly permitted were subject to tendering processes despite strong incumbents reactions and some opposition within the State bureaucracy.

Regarding fare definition, this is made in a route-by-route basis: a standard cost sheet is applied to route conditions respect to fleet and patronage and a pre-defined rate of profit marks up average costs in order to calculate the route fare (Aragão *et al.*, 2000). There was no awareness regarding parameters and coefficients used in fare calculations until 1997 when the Regulatory Authority contracted a private consultancy to develop a fare-calculating model in which service quality criteria were considered. No concrete results of this attempt may be showed until now: although a *Rule for Evaluating and Qualifying Passenger Public Transport Services* has been published there is no clear understanding about how evaluations should affect fares. Furthermore, there is no security about the possibility of applying the rules once legal framework does not refer to a systematic evaluation. The fact is that Public Transport Council approved an administrative decision (in 2000, August) by which every public transport concession in the Metropolitan Area of San José was renewed. The arguments used were the high costs and complexities of a tendering process as well as the need for consolidate the bus sector (La Nación, 2000, August 31st).

This political decision shall be understood as a result of a long time period of applying a regulatory regime based on the absence of competition. The incumbent operators conform now a closed group able to develop regulatory capture: if not in a formal sense at least in a practical sense. That is why tendering does not occur and contract renewing is done without any assessment of operator's merits.

TECHNICAL EFFICENCY OF PUBLIC TRANSPORT IN GREAT SAN JOSÉ

In order to examine technical efficiency revealed in urban bus industry of San José Metropolitan Area this paper uses Data Envelopment Analysis — DEA. DEA techniques are mathematical programming approaches that permit simultaneously to construct, given a technology and a set of observations, the efficiency frontier and to assess, to each individual observation, how far it is from the efficiency frontier. DEA approach is based on the best practices in combining inputs and outputs — efficiency frontier —, and on the optimization of individual decision-making units — DMUs. So, in DEA, production frontier is defined as the maximum quantity of product possible to obtain from a given set of resources or inputs: a piece-wise linear frontier is determined that contains all Pareto-efficient observations.

There are three basic models used to perform Data Envelopment Analysis:

- Constant returns to scale, also called the CCR model (due to Charnes, Cooper and Rhodes; see Charnes *et al.*, 1978) or CRS/E/I/A (Lins and Meza, 2000);
- Variable returns to scale, also called the BCC model (due to Banker, Charnes and Copper; see Banker *et al.*, 1984) or VRS/E/I/A (Lins e Meza, 2000)
- Non-increasing returns to scale, usually called NIRS, or NIRS_RAD_IN in this paper.

In the work described in this paper, we used the software EMS — *Efficiency Measurement System* —, Version 1.3, 2000, developed by Holger Scheel in the University of Dortmund. The denomination given to the three basic models is compatible with software outputs (Scheel, 2000).

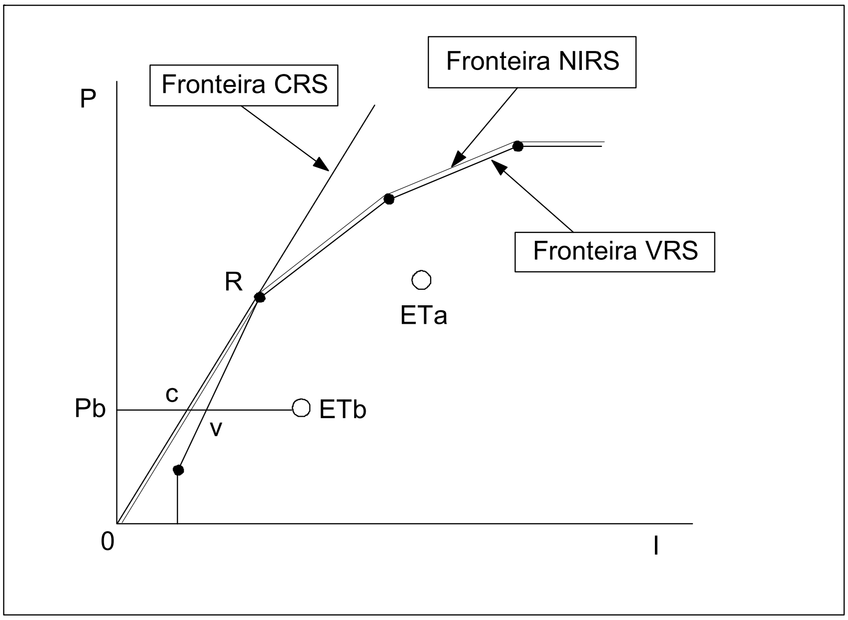

Fig. 1 - Frontiers CRS, VRS, NIRS and the determination of economies of scale

In Fig. 1, considering one input and one output, we show frontiers obtained with the models CRS, VRS and NIRS, as well as the positions of inefficient DMUs ET_a and ET_b.

With the models VRS and CRS it is possible to obtain a measure of scale efficiency for each DMU. This implies to decompounding the measurement of CRS technical efficiency into two parts: the first refers to scale inefficiency and the second relates to “pure” technical inefficiency. If a single DMU registers different technical efficiencies as measured by CRS or VRS model, then scale inefficiency exists that may be measured on the basis of the difference. In Fig. 1, considering the DMU ET_b, its CRS technical inefficiency is the distance ET_b-C. Regarding VRS model, the technical inefficiency of ET_b is given by ET_b-V and V-C is the scale inefficiency of ET_b.

The Database

Data were gathered in four secondary sources:

- Ministry of Public Works and Transport, mainly the reports of Technical Studies and of Public Transport Departments;
- Registers of Operators Accountancy;

- Consultancy Report contracted by Regulatory Authority in 1997;
- Ministry of Treasury.

The built database is composed by monthly information (year 1997) respect to 31 operators, e.g. two thirds of total incumbents of the system. The sampled operators were responsible by 81 from the 130 existing routes and their aggregated fleet summed up 70% of total operating fleet in San José Metropolitan Area. An analysis of all available information regarding the 47 operators in the system was performed and it was possible to suppose that conclusions obtained with empirical analysis of the sampled operators may be generalized as the operators not being in the sample may be represented by linear combinations of the sampled ones.

Used data were the following:

- Monthly produced vehicle-km (Q), a measure of service produced (in thousand km)
- Fleet size (fleet), an input measured in quantity of available vehicles to firm production (number of seats is approximately homogeneous in the sampled operators)
- Operational manpower (desop), an input defined as the total wages paid monthly to operative personnel (in ten thousand *colones*)
- Administrative manpower (desad), an input defined as the total wages paid monthly to administrative personnel (in ten thousand *colones*)
- Fuel consumption (combus), an input defined as the total expenses regarding monthly fuel acquisitions (in ten thousand *colones*)

Operators are represented by codes randomly defined, from ET1 to ET31. In Tab. 1, each variable is presented in terms of sample median, mean, standard deviation and coefficient of variation (standard deviation/mean).

Tab. 1 – General statistics of variables

Sample characteristics	**Q**	**fleet**	**desop**	**desad**	**combus**
Median	60,79	16	204,28	23,65	210,58
Mean	70,96	22,55	242,91	36,30	287,01
Standard deviation	47,67	17,46	200,90	37,61	249,59
Coefficient of variation	0,67	0,77	0,83	1,04	0,87

It can be observed that there is an expressive variability in the sample, both in what respect inputs and outputs. The differences in fleet size and the fact that variables are measured absolutely are the reasons for that the sample presents these characteristics. Also worthy to note is the independence between inputs as can be demonstrated in the correlation matrix shown below.

Tab. 2 – Correlation matrix between inputs and output

	Q=v-km	fleet	desop	desad	combus
Q=v-km	1,000				
fleet	0,931	1,000			
desop	0,787	0,827	1,000		
desad	0,643	0,666	0,831	1,000	
combus	0,951	0,942	0,715	0,611	1,000

It may be observed in Tab. 2 that chosen inputs are relevant to explain the production Q. Expenses with administrative personnel (desad) is the input with lower correlation with production. High correlation value between fleet and fuel expenses (combus) was expected. Nevertheless it was decided to maintain the two variables due to the fact that fuel consumption is linked to operational characteristics more strictly than fleet. So, it is supposed that they can have differentiated contributions to explain output variations.

Results of the application of CRS and VRS models

For each DMU in the sample, results were obtained under the models of constant returns to scale – CRS – and variable returns to scale – VRS –. They are:

- Individual scores
- Reference DMUs to those that are not in the efficiency frontier — reference DMUs are benchmarks for inefficient ones — and intensity of influence of an efficient DMU over inefficient ones
- Quantity of DMUs to which an efficient DMU serves as reference**.**

In the case of CRS model, the reference set is formed by four DMUs. In the case of VRS, the set comprises 11 DMUs. So, considering as given the scale of production, aroun a third of sampled DMUs were using inputs efficiently. The DMU ET7 is highlighted as a benchmarking and this conclusion implies its production practices should be studied in-depth.

In Tab. 3, below, a synthesis is shown of results obtained for short and log term efficiency. Lowest values of efficiency scores are also shown. It is remarkable that ET12 operates with an inefficiency around 50%. One sixth of sampled DMUs operates with a short run efficiency level below 70%. The Tab. 3 also remarks those DMUs operating upon the efficiency frontier.

Tab. 3 - Technical Efficiency and Reference Sets

	Technical Efficiency		Reference Set
	Mean	Minimal	
Short term	74,99%	42,37% (ET12)	ETs: 7, 8, 9, 10, 16, 18, 19, 20, 21,26, 29
Long term	85,33%	51,78% (ET12)	ETs: 7, 16, 18, 20

As discussed above, this kind of analysis permits the calculus of scale efficiency of each DMU. But it does not provide a classification of the kind of returns to scale: decreasing, constant or increasing ones. This information may be obtained solving NIRS model and comparing the results with those obtained for VRS model. Tab. 4 shows the results.

Tab. 4 – DEA Classification of DMUs according to scale efficiency and returns to scale (DEA with VRS e NIRS models)

DMU	***SCORE*** **CRS**	***SCORE*** **VRS**	***SCORE*** **NIRS**	**Scale efficiency**	**Returns to scale** **crs: constant** **irs: increasing** **drs: decreasing**
ET1	80,51%	87,61%	80,51%	91,90%	irs
ET2	57,35%	82,69%	57,35%	69,36%	irs
ET3	45,67%	57,84%	45,67%	78,96%	irs
ET4	69,83%	70,25%	70,25%	99,40%	drs
ET5	73,28%	74,99%	73,28%	97,72%	irs
ET6	57,97%	81,35%	81,35%	71,26%	drs
ET7	100,00%	100,00%	100,00%	100,00%	crs
ET8	90,81%	100,00%	90,81%	90,81%	irs
ET9	95,23%	100,00%	95,23%	95,23%	irs
ET10	68,94%	100,00%	100,00%	68,94%	drs
ET11	75,79%	76,34%	75,79%	99,28%	irs
ET12	42,37%	51,78%	51,78%	81,83%	drs
ET13	65,09%	67,26%	65,09%	96,77%	irs
ET14	64,83%	64,87%	64,83%	99,94%	irs
ET15	87,50%	89,40%	89,40%	97,87%	drs
ET16	100,00%	100,00%	100,00%	100,00%	crs
ET17	68,39%	71,31%	68,39%	95,91%	irs
ET18	100,00%	100,00%	100,00%	100,00%	crs
ET19	91,37%	100,00%	91,37%	91,37%	irs
ET20	100,00%	100,00%	100,00%	100,00%	crs
ET21	75,46%	100,00%	75,46%	75,46%	irs
ET22	54,49%	81,25%	54,49%	67,06%	irs
ET23	96,12%	96,25%	96,12%	99,86%	irs
ET24	57,16%	61,86%	61,86%	92,40%	drs
ET25	73,34%	73,63%	73,63%	99,61%	drs
ET26	72,42%	100,00%	100,00%	72,42%	drs
ET27	70,49%	93,43%	93,43%	75,45%	drs
ET28	54,90%	94,62%	94,62%	58,02%	drs
ET29	96,50%	100,00%	96,50%	96,50%	irs
ET30	82,07%	82,87%	82,07%	99,03%	irs
ET31	56,78%	85,70%	85,70%	66,25%	drs
Means	74,99%	85,33%	81,13%	87,88%	

It should be noted that when technical efficiencies obtained with CRS and VRS models are equal then operator is running under constant returns to scale (Coelli *et al*., 1998). In Tab. 4, it may be seen that 4 operators are producing in a situation of constant returns to scale. In a region of increasing returns to scale it was possible to find 16 sampled operators. And eleven of them were found to be producing in a region of decreasing returns to scale.

Four cases are highlighted in Tab. 4: those of DMUs ET3 and ET12 — with the largest distances to efficient frontier — and those of DMUs ET23 and ET28 — with the lowest distance to efficient frontier — both from the point of view of inputs.

Tab. 5 shows the 11 DMUs which compound the reference set according to VRS analysis, as well as their classification respect to the kind of returns to scale they present. We added a column with the DMU fleet in order to provide the scope of reference set (Range of fleet in sample: 5-to-74 vehicles).

Tab. 4 – Reference set and fleet size by DMU

DMU	*SCORE*	RETORNOS	FROTA
ET7	100,00%	crs	16
ET8	100,00%	irs	7
ET9	100,00%	irs	10
ET10	100,00%	drs	74
ET16	100,00%	crs	12
ET18	100,00%	crs	14
ET19	100,00%	irs	5
ET20	100,00%	crs	10
ET21	100,00%	irs	5
ET26	100,00%	drs	33
ET29	100,00%	irs	10

In the reference set, of course, we find those 4 DMUs operating with a scale efficiency of 100%. These DMUs are on the efficiency frontier and producing in a point of optimal scale. These DMUs are ET7, ET16, ET18 and ET20. The mean efficiency score of inefficient DMUs — those not in the reference set — is around 77% with a standard deviation of 12,5%.

FINAL COMMENTS

An extended and rigorous performance analysis concerning to public transport industry in developed countries is needed. In metropolitan areas of developing world, public transport is responsible by carrying people to work and other urban activities in a greater proportion than in developed ones. So, as public transport is a general input to urban production, greater costs of production inefficiency in the bus industry result in greater production costs for general economic activity.

As Latin American cities, almost in general, has adopted the public service concept as a basis for organizing production of urban transport services, public regulation must behave in order to obtain the maximum of technical efficiency and regulatory framework should be designed with this objective in mind. Application of DEA techniques may help this public effort, in the sense that they may identify and quantify good production practices, as well as highlight unacceptable levels of technical inefficiency.

In what concerns to empirical analysis showed in this paper, it is possible to conclude that, given the scale of production, just one third of all operators in the bus industry of San José Metropolitan Area are running services efficiently. In a short-term perspective, technical efficiency in the industry is around 85% while in the long run the figure goes down to 75%. These figures show significant technical inefficiencies in the industry and points out the need for public intervening in order to achieve higher levels of efficiency.

Another interesting result is the identifying of benchmarking firms whose production practices should be examined by public regulators and inefficient incumbents in order to improve the economic performance of the whole system. Incentive regulation may be used in order to induce inefficient firms to become closer to efficiency frontiers, as well as to guarantee that efficient ones will stand on ideal production points.

Competitive pressures may be posed to incumbents during the contract period in the sense that performance evaluation may be used as a criteria to be observed when the decision moment about contract renewing arrives. DEA techniques may be an efficient technical aid for regulators and providers in a more competitive regulatory environment of local bus industry.

Referências Bibliográficas

Aragão, J.J., Contreras-Montoya, C. and Lima Neto, O. (2000) Co-ordination and competitiveness levels of transit services in the metropolitan areas of Recife (Brazil) and San José (Costa Rica). In: *Proceedings of the International Conference CODATU IX*, pp. 609 - 616 , Mexico City, April. Rotterdam: A.A. Balkema

Banker, R. D., Charnes, A., and Cooper, W. (1984) Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis. In: *Management Science*, v. 30, n.9, pp. 1078-1092.

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

Coelli, T., Rao, D. and Battese, G. (1998) *An Introduction to Efficiency and Productivity Analysis*, Kluwer Academic Publishers, Norwell, MA, USA.

Contreras-Montoya, C. (2000) Estudo da Forma de Organização, Eficiência Técnica e de Escala na Indústria de Ônibus Urbano de San José com Análise Envoltória de Dados. Curso de Análise Envoltória de Dados, Programa de Engenharia de Produção, COPPE, UFRJ.

INICEM (1995) Costa Rica: datos e indicadores básicos, San José: INICEM-Market Data.

LA NACIÓN (2000) "MOPT listo para concesiones", Sección de noticias nacionales, 31 de agosto.

Lins, M. P. E. and Meza, L.D., eds, (2000) Análise Envoltória de Dados e perspectivas de integração no ambiente de Apoio à Decisão. Rio de Janeiro, COPPE, UFRJ.

Scheel, H. (2000) EMS: Efficiency Measurement System, User's Manual. Version 1.3. Dortmund.