BENCHMARKING RAPID-TRANSIT SERVICES WITH DATA ENVELOPMENT ANALYSIS

Antonio G.N. Novaes
Transport and Logistics Group Department of Industrial Engineering
Federal University of Santa Catarina, 88040-900 - Florianópolis, SC, Brazil
fax: (55) 48 2319770 - email: novaes@eps.ufsc.br

ABSTRACT

Data Envelopment Analysis (DEA) is a recently developed OR technique, that takes Linear Programming (both conventional and fractional) as a basis, and whose objective is to analyze and compare independent units (firms, organizations, departments, etc) with regard to its operational performance. It was originally developed by Charnes, Cooper and Rhodes (1978) and applied to the evaluation of public schools in the United States. Today, DEA constitutes one of the most investigated fields of OR, with a large number of papers presented at the specialized Conferences and published in scientific journals. In this text an application of DEA to the evaluation of rapid-transit agencies is discussed, with the purpose of defining a benchmarking strategy on the one hand, and estimating a production function, on the other hand. Forty-four subways of the world are investigated, with the objective of assessing their efficiency. A peer-group analysis is particularly performed for the São Paulo's rapid-transit, relating it to the efficient DMUs (benchmarking).

1. INTRODUCTION

The privatization of transport services, both passenger and cargo, raises the question of how to control and evaluate operators correctly, without bureaucracy and with the desired transparency to the community. The practical application of any evaluating method requires that the data be obtained sistematically and with relative facility from the operating firms. Otherwise the evaluation process will be in danger of failure. This is to say that, although transport theorists argue in favor of choosing certain inputs and outputs to represent a production scheme, very often it is necessary to assume some sort of simplification. For example, the supply of transportation capacity by the rapid transit agency is best expressed in vehicle-kilometers (Berechman, 1993). It reflects the *efficiency* of the firm, and it is related to production objectives of technical order. The consumed output, on the other hand, is related to the transportation demand, being expressed in passenger-trips or passenger-kilometers. As pointed out by Talley (1988), the public transit firms generally seek to maximize passenger kilometers or passenger trips. But one limiting factor is the availability of data, in this case related to the type of fare structure in operation. The common flat structure prevailing in many rapid-transit systems provides information on passenger trips, whereas a graduated fare structure gives way to a passenger-kilometer maximization objective. Statistics on annual passenger trips are available for any rapid-transit of the world, whereas passenger-kilometers are only available from the agencies that adopt a graduated fare system. As a consequence, one must very often adopt a simplified formulation, taking passenger-trips, for instance, to represent the production output, instead of the more general passenger-kilometers variable. But, in order that the evaluation process may effectively yield practical results in terms of improving the services under analysis, it is important that the adopted methodology furnish precise indications with regard to the lines of corrective actions to be taken.

There is an extensive literature on the estimation of productivity for the transportation industries. The basic idea of productivity is the comparison of outputs with inputs, more specifically, how output-input relationships differ across firms and/or change over time. A number of conceptual approaches for measuring productivity can be identified in the literature. They include parametric approaches, for which a functional relationship among inputs and outputs (a production function) must be specified and its parameters must be statistically estimated. Nonparametric approaches are also common, as index numbers and, more recently, Data Envelopment Analysis (DEA).

In dealing with parametric approaches, there is, of course, the problem of choosing an appropriate functional form and the relevant variables to be included in the model. In current transport productivity studies the use of generalized Cobb-Douglas production or cost functions has been substituted by second-order approximations, namely the transcendental logarithmic representation, or *translog* for short (Spady and Friedlaender, 1976). One problem that arises with parametric models of productivity is that traditional regression estimation methods, such as ordinary least squares, fit an average curve through a set of data points. However, economic theory describes the production or cost function as reflecting the behaviour of the most efficient firms in the set. This suggests that the usually estimated middle-of-the-data-set production or cost function is inappropriate (Oum *et al*, 1992). Alternative approaches, as *stochastic frontier functions*, have been applied by some authors. In this paper we show that the combination of traditional regression estimation methods with DEA can significantly improve the results.

Benchmarking is today one of the modern instruments of management, and it is used to improve the firm's performance in a comparative way. One departs from the assumption that the organizations that strive in a certain sector of the economy tend to show similar activities, with patterns that can applied to them all. Some of them succeed in attaining better results, due to better combinations of inputs, modern production processes and technology, etc thus generating outputs with more efficiency and more effectiveness. These organizations are taken as a reference to the less productive firms. The latter must then review their processes and costs as to evolve to a better rank in terms of management performance. DEA has also been used to benchmark inefficient units relating them to the peer groups of efficient units (Banker *et al*, 1986; Kao, 1994; Silva *et al*, 1994). It is a powerful tool to define a benchmarking strategy intended to upgrade inefficient firms to the efficient level.

2. THE TRANSLOG PRODUCTION FUNCTION

A major class of multiplicative models is the well-known Cobb and Douglas specification, based on principles from the neoclassical theory of the firm. Although easy to apply, the Cobb-Douglas model imposes restrictions on some economic effects of interest in transportation analysis. For example, the elasticity of substitution between different factors of production is constrained to unity. Berechman (1993), reviewing a representative collection of empirical transit studies, concluded that transit technology, both bus and rail, is not a Cobb-Douglas production function type technology. This is because elasticities of input factor substitution have important implications in transit technology, and thus cannot be held constant. Also, returns to scale cannot be computed independently of the mutual effects of output and input factors in some circumstances.

More recently, transit studies have used the translog specification, which provides a second-order approximation to an arbitrary production or cost function. A function is a second-order numerical approximation to *f(x)* at a point x^* if it accurately reproduces the value, the gradient, and the Hessian of *f(x)* at x^* (Spady and Friedlaender, 1976). The translog function satisfies these conditions and is quite flexible in distinction from restricted types, like the Cobb-Douglas representation. It places very few *a priori* restrictions on the attributes of the underlying production technology, such as the degree of factor substitution or of homogeinity (Berechman, 1993). Since the translog function provides a second-order numerical approximation to a production function at a point, it is necessary to specify such a point. Let it be the point (y^0, x^0), with $x^0 = (x_1^0, x_2^0, ..., x_n^0)$, where n is the number of inputs. We assume only one output y, but the concept can be extended to multiple output formulations (Spady et al, 1976; Christensen et al, 1973). The translog production function of $\Phi(y,x)$ is:

$$\ln\Phi(y,x) = a_0 + \sum_{i=1}^{n} a_i (\ln x_i - \ln x_i^0) + \tfrac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} b_{ij} (\ln x_i - \ln x_i^0)(\ln x_j - \ln x_j^0) \qquad (1)$$

where $b_{ij} = b_{ji}$. The production function $\Phi(y,x)$ is homogeneous of degree k in x, for fixed y, if and only if (Spady and Friedlaender, 1976):

$$(i) \quad \sum_{i=1}^{n} a_i = k \qquad (2)$$

$$(ii) \quad \sum_{j=1}^{n} b_{ij} = 0 \qquad i = 1,2,...,n \qquad (3)$$

After estimating the production function, it can be used to study possible returns to scale, and to compute production elasticities related to each input, as well as elasticities of substitution between different factors of production. Transportation economists tend to first estimate a cost function and, from it, the corresponding production structure. This approach is based on some economic concepts and properties developed by Shephard (1953, 1970). The Shephard's lemma shows that it is possible to analyze the production frontier by first estimating the "dual" cost function. Under certain conditions, the cost function will yield all the economic information of the related production function. The econometric estimation of cost functions, based on exogenously determined output level and input unit prices, generally implies low correlation between variables. Input demands, on the other hand, are likely to be to some extent determined endogenously, leading to simultaneous-equation bias when production functions are estimated directly.

The extensive use of Shephard's "duality" concept by transportation economists, however, is not free of theoretic and empirical limitations (Charnes, Cooper and Schinnar, 1982; Berechman, 1993). First, since cost minimization is not, in fact, the main objective of the transit entity, the estimated cost function parameters are likely to be biased, since Shephard's lemma presupposes cost minimization behavior of the firm (Berechman, 1993). Second, Shephard's duality concept requires input prices independent of the corresponding input quantities, a strong hypothesis frequently unobserved in real life. Furthermore, when treating the problem through a cross-section analysis of world-wide data, as in the case focussed in this paper, the input prices can hardly be taken as explanatory variables due to different national economic policies, labour regulations, construction costs, rolling stock prices, energy costs, etc. In those situations it seems to be more appropriate to base the analysis directly on production factors, instead of unit prices. By applying transformations on the input variables one can reduce correlations to acceptable levels when estimating the production function.

As a consequence of these restrictions, we decided to directly estimate the production function of rapid-transit agencies based on engineering and operational variables such as total trackage, number of stations, fleet, number of employees, etc. To do so it is necessary to avoid multicollinearity problems resulting from high correlation between independent variables. This will be explained in the next section, where the statistical analysis of the data is presented.

3. RAPID-TRANSIT PRODUCTION FUNCTION: ROUGH ESTIMATION

Let us first estimate a rapid-transit production function using the traditional multiple regression approach. The data comprise 44 rapid-transit agencies of the world, and were gathered by the São Paulo's transit agency (Cia. do Metropolitano de São Paulo). Originally 76 rapid-transit properties were reported, but good part of the raw data was discarded due to incomplete information. The 44-sample, however, is representative enough, comprising properties of all sizes and covering most important cities of Europe, the United States, Asia and Latin America. The metros of the sample are: Atlanta, Baku, Baltimore, Barcelona, Beijing, Bucarest, Buenos Aires, Caracas, Chicago, Frankfurt, Glasgow, Hamburg, Hong Kong, Kobe, Kyoto, Lille, Lisbon, London, Los Angeles, Lyon, Madrid, Marseille, Mexico City, Minsk, Moscow, Nagoya, New York, Osaka, Paris, Prague, Pusan, Rome, Rotterdam, Santiago, S. Francisco, São Paulo, Sapporo, Seul, Singapore, Tokio, Vienna, Washington, Yerevan, Yokohama. The following attributes were available: population of the served city, total passengers carried per year, peak-hour headway, total track extension, number of existing lines, total number of stations, rolling stock (passerger cars), daily period of operation, and total number of employees. Table 1 shows the average, standard deviation, minimum, and maximum values of each variable.

Table 1
Data Summary

Variable	Average	Standard deviation	Minimum	Maximum
Population (million)	4.9	4.7	0.8	20.0
Pass. carried per year (million)	431.2	610.9	4	2738
Peak-hour headway (sec.)	178.8	86.1	72	480
Trackage (km)	82.2	95.2	7.1	443.0
Number of lines	4.6	5.2	1	31
Number of stations	77.7	98.2	5	504
Rolling stock (pass. cars)	873.9	1351.1	30	6285
Daily time of operation (hrs)	19.0	1.6	15	24
Total number of employees	5407.4	6215.6	182	27544

As mentioned, since a substantial part of the world's rapid transit systems adopt a flat-fare structure, the universal data available for the output are passengers carried per year (variable *P* in the model). This is the attribute we have adopted to represent the output in our model. With regard to the inputs, instead of the traditional economic approach, in which the factors of production are generally capital, labor and energy, we focus our analysis on the technical/operational aspects of the rapid-transit production process, choosing the input variables accordingly. Furthermore, a dominant input factor in rapid-transit operations is their *trackage* (Berechman, 1993). This variable, together with the *number of stations* and the *number of lines*, are short-run fixed attributes. On the other hand, another group of attributes are formed by dynamic variables, in the sense that they can be adjusted on a short or medium-run basis in order to regulate the supply of service. The *number of cars*, the *headway*, the *daily period of operation*, and the *number of employees* are examples of this type of variable.

After a careful analysis of the correlations and of the production process we selected three explaining variables:

(a) total trackage, *T* (km);
(b) a transformed variable *D*, given by

$$D = \left(\frac{C}{T}\right) * \left(\frac{S}{T}\right) * \left(\frac{3600}{h}\right) \quad (4)$$

where *C* is total the number of cars in operation, *S* is the total number of stations, and *3600/h* is the hourly frequency of trains during the peak period (*h* is the headway, in seconds). The idea behind this formulation is that, apart from the direct effect of trackage *T*, the production of passenger trips is dependent on the fleet of cars in operation and on the number of stations, both taken as indexes, being also dependent on the service frequency. Although the number of cars per km of line is partially dependent on the frequency of trains and on the average spacing between stations, it also reflects the size of the composition, and therefore brings adittional information to the regression analysis. The number of stations, on the other hand, is related with the rapid-transit level of penetration into the urban tissue (capillarity), bringing more or less transport demand to the rapid-transit system, in a spatial competition with other modes. Therefore, the number of stations per km of track is also an important independent variable. Finally, the frequency of trains is associated with the degree of competition between the rapid-transit mode and other transport alternatives in the time domain. The effect of the daily period of operation on production was not significant. This is understandable, since the non-operating hours of the day are always coincident with very low levels of demand (usually from 11 pm or midnight to around 5 am in the morning);

(c) a transformed variable E, given by

$$L = \left(\frac{E}{S}\right) \tag{5}$$

where E is the total number of employees and S is the number of stations. We analyzed the index E/T too, but the ratio E/S has proved to be statistically more significant. The number of employees, although operationally related to the number of trains (train driving, vehicle maintenance), trackage (line maintenance), and number of stations (ticketing, control, surveillance), also reflects the management effort to serve the users better (clean premises, public attendance, supplementary services) with a positive marketing impact. Hence, the number of employees can reflect explaining factors other than the basic operational relationships, justifying its inclusion as an independent variable.

The adoption of only one mass variable (T), together with input indexes, has to do with the reduction of correlation between independent variables. For instance, the number of metro lines is highly correlated with T, C and S. Its introduction in the regression model did not improve the results, and thus it was eliminated. The index $L = E/S$, on the other hand, generated better statistical results than the index E/T. This means that the labor effect on rapid-transit production is better associated with station-located activities as ticket sales, station patrol, etc than with trackage.
The adopted translog production function is:

$$\begin{aligned}
\ln P_i - \ln P_0 = {} & a_0 + a_1 (\ln T_i - \ln T_0) + a_2 (\ln D_i - \ln D_0) + a_3 (\ln L_i - \ln L_0) + \\
& + \tfrac{1}{2} b_{11} (\ln T_i - \ln T_0)^2 + \tfrac{1}{2} b_{22} (\ln D_i - \ln D_0)^2 + \tfrac{1}{2} b_{33} (\ln L_i - \ln L_0)^2 + \\
& + b_{12} (\ln T_i - \ln T_0)(\ln D_i - \ln D_0) + b_{13} (\ln T_i - \ln T_0)(\ln L_i - \ln L_0) + \\
& + b_{23} (\ln D_i - \ln D_0)(\ln L_i - \ln L_0)
\end{aligned} \tag{6}$$

where we have substituted b_{12}, b_{13}, and b_{23} for b_{21}, b_{31}, and b_{32}. As the translog form provides a second-order approximation of the production function at a point, one has to specify this point. Following current practice we used the sample geometric mean of each variable (P_0, T_0, D_0, and L_0) to constitute such an approximation point. Moreover, in order to guarantee that the production function be homogeneous (Spady and Friedlaender, 1976) one must have:

$$\begin{aligned}
b_{11} + b_{12} + b_{13} &= 0 \\
b_{12} + b_{22} + b_{23} &= 0 \\
b_{13} + b_{23} + b_{33} &= 0
\end{aligned} \tag{7}$$

Equations (7) allow for the elimination of three coefficients, namely b_{12}, b_{13}, and b_{23}:

$$b_{12} = \tfrac{1}{2} (-b_{11} - b_{22} + b_{33}) \qquad (8)$$

$$b_{13} = \tfrac{1}{2} (-b_{11} + b_{22} - b_{33}) \qquad (9)$$

$$b_{23} = \tfrac{1}{2} (+b_{11} - b_{22} - b_{33}) \qquad (10)$$

Substituting (8), (9), and (10) in (6) and simplifying we get:

$$\ln P_i - \ln P_0 = a_0 + a_1 (\ln T_i - \ln T_0) + a_2 (\ln D_i - \ln D_0) + a_3 (\ln L_i - \ln L_0) + \\ + \tfrac{1}{2}\, b_{11}\, Z_{i,1} + \tfrac{1}{2}\, b_{22}\, Z_{i,2} + \tfrac{1}{2}\, b_{33}\, Z_{i,3} \qquad (11)$$

with:

$$Z_{i,1} = (\ln T_i - \ln T_0)^2 - (\ln T_i - \ln T_0)(\ln D_i - \ln D_0) - (\ln T_i - \ln T_0)(\ln L_i - \ln L_0) + \\ + (\ln D_i - \ln D_0)(\ln L_i - \ln L_0) \qquad (12)$$

$$Z_{i,2} = (\ln D_i - \ln D_0)^2 - (\ln T_i - \ln T_0)(\ln D_i - \ln D_0) + (\ln T_i - \ln T_0)(\ln L_i - \ln L_0) + \\ - (\ln D_i - \ln D_0)(\ln L_i - \ln L_0) \qquad (13)$$

$$Z_{i,3} = (\ln L_i - \ln L_0)^2 + (\ln T_i - \ln T_0)(\ln D_i - \ln D_0) - (\ln T_i - \ln T_0)(\ln L_i - \ln L_0) + \\ - (\ln D_i - \ln D_0)(\ln L_i - \ln L_0) \qquad (14)$$

The coefficients of equation (11) were obtained via multiple regression analysis, the results being displayed in Table 2.

Table 2
Estimation Results for the Translog Production Function
Situation A (44 observations, $R^2 = 0.878$)

Coeff.	Coeff. Value	St. Error	t statistic
a_0	0.4191	0.1360	3.082
a_1	0.8737	0.0887	9.847
a_2	0.6300	0.0992	6.351
a_3	0.4253	0.0940	4.523
b_{11}	- 0.5409	0.1847	- 2.928
b_{22}	- 0.2756	0.1139	- 2.420
b_{33}	- 0.3408	0.1040	- 3.277

These results correspond to a rough estimation of the production function, including in the sample all the available rapid-transit data regardless of the efficiency status of the operating agency. We call this case "situation A" (Table 2). The R^2 value is satisfactory, and the t statistic with 37 degrees of freedom, is 99% significant for all coefficients, except b_{22} which is significant at the 95% level. We will see that these results can be considerably improved when one upgrades the inefficient rapid-transit firms with DEA, and further taking the resulting sample as a basis for obtaining the production function via multiple regression analysis.

4. EFFICIENCY EVALUATION WITH DATA ENVELOPMENT ANALYSIS

Data Envelopment Analysis (DEA) was developed by Charnes, Cooper, and Rhodes (1978) as a process for measuring the relative efficiency of a group of public and/or non-profit decision making units (DMUs). Each DMU is represented by a set of S outputs and a set of M inputs. In contrast to parametric approaches whose objective is to optimize a single regression curve through the data, DEA optimizes on each individual observation in order to calculate a discrete piecewise frontier determined by the Pareto-efficient DMUs. In parametric analysis, the single optimized regression function is applied to each DMU. In contrast, DEA optimizes the performance measure of each DMU. In other words, the focus of DEA is on the individual observations as represented by the n required optimizations (n being the number of DMUs under analysis). This is in contrast to the focus on the average behavior that is associated with single-optimization statistical approaches (Charnes *et al*, 1994).

Charnes, Cooper, and Rhodes (1978) extended Farell's approach linking the estimation of technical efficiency and production frontiers. Their initial model generalized the single output/input ratio measure of efficiency for a single DMU due to Farrel. The multiple output/input characterization of each DMU was transformed into a single virtual output and a single virtual input. Then, the problem was treated with a fractional linear-programming formulation. The relative technical efficiency of any DMU is calculated by forming the ratio of a weighted sum of outputs to a weighted sum of inputs. These weights are selected calculating the Pareto efficiency measure of each DMU subject to the constraint that no DMU can have a relative efficiency score greater than unity. The DMUs which attain a relative efficiency score equal to one form the frontier efficient set of DMUs (Charnes *et al*, 1994).

Two basic DEA models are generally used in the applications. The first, called the *CCR ratio model* (Charnes et al, 1978) allows for the evaluation of overall efficiency, identifies the efficient and non-efficient DMUs, and determines how far from the efficient frontier are the non-efficient units. The *BCC* model (Banker *et al*, 1984) , which is the dual of the former, is normally used for benchmarking. It yields the projected efficiency point on the envelopment surface for each non-efficient DMU, as well as the peer-group of efficient units associated with each non-efficient DMU.

We first analyze model CCR. Let $y_i = \{y_{1i}, y_{2i}, \ldots y_{Si}\}$ and $x_i = \{x_{1i}, x_{2i}, x_{3i}, \ldots, x_{Mi}\}$ be the vector of outputs and inputs respectively for DMU i (i = 1, 2,.., n), where S and M are respectively the number of outputs and inputs considered in the analysis. Outputs and inputs are transformed into single virtual entities by weighting the values of the attributes. The single virtual output, for DMU i, is:

$$Y_i = u_1 y_{1i} + u_2 y_{2i} + \ldots + u_S y_{Si} \tag{15}$$

and the single virtual input:

$$X_i = v_1 x_{1i} + v_2 x_{2i} + \ldots + v_M x_{Mi} \tag{16}$$

In order to reach maximum efficiency, DMU i will strive to maximize the ratio Y_i/X_i. Although the ratio Y_i/X_i can take on any non-negative value, it is costumary to normalize it, assigning the unit value to the maximum attainable ratio, with all other situations in between. If DMU i is under analysis, the CCR mathematical model is:

$$\varphi_i = max \frac{u_1 y_{1i} + u_2 y_{2i} + \ldots + u_S y_{Si}}{v_1 x_{1i} + v_2 x_{2i} + \ldots + v_M x_{Mi}}$$

$$\text{subject to:} \quad \varphi_j = \frac{u_1 y_{1j} + u_2 y_{2j} + \ldots + u_S y_{Sj}}{v_1 x_{1j} + v_2 x_{2j} + \ldots + v_M x_{Mj}} \leq 1 \qquad \text{for } j = 1,2,\ldots, n \tag{17}$$

Model (17) has constant returns to scale. In order to represent variable returns to scale, one adds an "intercept" coefficient u_0 to the numerators of (17), or a v_0 coefficient to its denominators. The former corresponds to the *input orientation* case, in which one focuses on maximal movement toward the frontier through proportional reduction of inputs, whereas the latter, the *output orientation* form, focuses on maximal movement via proportional augmentation of outputs (Charnes *et al*, 1994). We adopted the input orientation approach in our study, since it is easier to improve efficiency in rapid-transit operation by acting at least on part of the inputs.

Model (17) is applied separately to each DMU. In each case, the value of φ_j for each DMU is recorded, yielding a square matrix $\{\varphi_{ij}\}$ whose element is the efficiency level of DMU j as seen by DMU i. A DMU is only efficient if the correponding value of φ is equal to unit. In other words, when $\varphi = 1$, the DMU is in the maximum production frontier. Thus, a first measure of efficiency is the number of cases that a given DMU has got a score $\varphi = 1$. As a secondary measure, the average value of all the φ_{ij}'s obtained by DMU i may be also considered in the comparisons.

We applied the logarithmic version (Charnes *et al*, 1994) of the CCR model (17) to the 44 rapid-transit DMUs, taking the logarithms of T, D, and L as inputs, and the logarithm of P as output, as explained in section 3, resulting in 20 efficient metros, as shown in Table 3.

Table 3
Efficient Rapid-Transit Ranking Obtained with DEA

Ranking	City Name	Number of obtained "ones"	Average φ value	Ranking	City Name	Number of obtained "ones"	Average φ value
1	Kobe	30	0.978	11	Singapore	11	0.947
2	Seul	21	0.958	12	Prague	8	0.944
3	Viena	20	0.968	13	Beijing	8	0.927
4	H. Kong	19	0.954	14	Yokohama	6	0.954
5	Toquio	18	0.963	15	Lille	6	0.907
6	Kyoto	17	0.951	16	L. Angeles	6	0.845
7	Yerevan	14	0.933	17	Paris	5	0.912
8	Atlanta	14	0.913	18	Osaka	4	0.955
9	Rotterdam	13	0.961	19	Glasgow	3	0.891
10	Pusan	13	0.956	20	Baltimore	1	0.837

Next, we applied the dual model, called the *BCC* DEA model (Banker *et al*, 1984), to the non-efficient DMUs:

$$min \quad \theta_E - \varepsilon[\sum_{i=1}^{S} s_i^+ + \sum_{j=1}^{M} s_j^-] \qquad (18)$$

subject to:

$$-\sum_{k=1}^{n} \lambda_k x_{ik} + \theta_E x_{iE} - s_i^- = 0 \qquad i = 1,2,..M \qquad (19)$$

$$\sum_{k=1}^{n} \lambda_k y_{jk} - s_j^+ = y_{jE} \qquad j = 1,2,..S \qquad (20)$$

$$\sum_{k=1}^{n} \lambda_k = 1 \qquad (21)$$

where E represents any non-efficient DMU, θ_E is the efficiency level of DMU E (with $0 \le \theta_E < 1$), and ε is a non-Archimedean (infinitesimal) constant (Charnes *et al*, 1994). The factor λ_k, with k=1,2..n, represents the participation fraction of the efficient DMU k in the benchmarking of the non-efficient DMU E. In other words, taking the resulting non-zero λ_ks, one is able to define the peer group formed by the efficient DMUs that will serve as benchmark to upgrade DMU E to the production frontier. Furthermore, the corresponding value of λ_k gives the level of participation of DMU k in the benchmarking process of unit E. Of course, for all non-efficient DMUs the corresponding value of λ_k will be zero. Variables s_i^- e s_j^+ are linear programming slacks variables.

After solving problem (18) to (21), the corrected level of input j, necessary to upgrade DMU E to the production frontier, is given by (Banker and Morey, 1986):

$$\hat{x}_{jE} = \theta_E^* x_{jE} - s_j^{-*} \qquad (j = 1,2, ...M) \qquad (22)$$

where superscript * stands for optimal values and $\hat{x}_{jE}$ is the revised level of input j. We used a slight modified version of the *BCC* DEA model because one of the inputs, trackage T, was assumed to be exogenously fixed in a short/medium run period (Banker and Morey, 1986). In the case of the São Paulo's Metro, the set of efficient DMU's that form its benchmarking peer group is the following: (a) Hong Kong (69.5 % participation), (b) Seul (17.9 %), (c) Yerevan (7%), and (d) Kyoto (5.6%). The resulting value of θ_E^* for the São Paulo's Metro is 0.96, meaning it is not too far from the production frontier.

5. RAPID-TRANSIT PRODUCTION FUNCTION: REFINED ESTIMATION

After applying the modified *BCC* model to the 24 non-efficient rapid-transit DMU's, the short-run modifiable inputs (variables D and L, given by expressions 4 and 5) were revised via expression (22) in order to upgrade the corresponding DMU to the efficient frontier. The only DMU for which it was not possible to reach the production frontier with a proportional reduction in D and L was the metro of Moscow. Thus, we got a set of 23 efficient-transformed DMUs which, together with the 20 efficient rapid-transit DMU's, served as our sample to re-estimate the translog production function (situation "B", Table 4).

By comparing the results presented in Tables 4 and 2, one can see that the R^2 value has been significantly improved. The same happens with the t statistic of the seven coefficients of the translog function. The regression coefficients furnish a way of obtaining scale factors, elasticities and cross effects between variables. Fig. 1, on the other hand, compares the observed values of P (passengers carried per year) with the ones produced by the model. We intend to further compare the explained methodology with other methods, such as stochastic frontier techniques (Oum *et al*, 1992).

Table 4
Estimation Results for the Translog Production Function
Situation B (43 observations, $R^2 = 0.927$)

Coeff.	Coeff. Value	St. Error	t statistic
a_0	0.4363	0.0986	4.426
a_1	0.8518	0.0685	12.432
a_2	0.7523	0.0762	9.879
a_3	0.5056	0.0728	6.940
b_{11}	- 0.5439	0.1329	- 4.091
b_{22}	- 0.2690	0.0852	- 3.159
b_{33}	- 0.3903	0.0770	- 5.069

5. REFERENCES

Banker, R.D., A. Charnes and W.W. Cooper (1984) Some models for estimating technical and scale inefficiencies in data envelopment analysis. **Management Science 30**, 1078-1092.

Banker, R.D. and R.C. Morey (1986) Efficiency analysis for exogenously fixed inputs and outputs. **Operations Research 34**, 513-521.

Berechman, J. (1993) **Public Transit Economics and Deregulation Policy**, North-Holland, Amsterdam, Holland.

Charnes, A., W.W. Cooper and E. Rhodes (1978) Measuring the efficiency of decision making units. **European Journal of Operations Research 2** (6), 429-444.

Charnes, A., W.W. Cooper, L. Seiford and J. Stutz (1982) A multiplicative model for efficiency analysis. **Socio-Economic Planning Sciences 16**, 223-224.

Charnes, A., W.W. Cooper and A.P. Schinnar (1982) Transforms and approximations in cost and production function relations. **Omega 10**, 207-211.

Charnes, A., W.W. Cooper, L. Seiford and J. Stutz (1983) Invariant multiplicative efficiency and piecewise Cobb-Douglas envelopments. **Operations Research Letters 2**, 101-103.

Charnes, A., W.W. Cooper, B. Golany and L. Seiford (1983) Foundations of data envelopment analysis for Pareto-Koopmans efficient empirical production functions. **Journal of Econometrics 30**, 91-107.

Charnes, A., W.W. Cooper, A.Y. Lewin and L.M. Seiford (1994) **Data Envelopment Analysis: Theory, Methodology and Applications**, Kluwer Academic, Boston, Massachusetts.

Chu, X., G.J. Fielding and B.W. Lamar (1992) Measuring transit performance using data envelopment analysis. **Transportation Research 26A**, 223-230.

Kerstens, K. (1996) Technical efficiency measurement and explanation of french urban transit companies. **Transportation Research 30A**, 431-452.

Oum, T.H., M.W. Tretheway and W.G. Waters,II (1992) Concepts, methods and purposes of productivity measurement in transportation. **Transportation Research 26A**, 493-505.

Pozdena, R. and L. Merewitz (1978) Estimating cost functions for rapid transit properties. **Transportation Research 12**, 73-78.

Shephard, R.E. (1953) **Cost and Production Functions**, Princeton University Press, Princeton, New Jersey.

Shephard, R.E. (1970) **Theory of Cost and Production Functions**, Princeton University Press, Princeton, New Jersey.

Silva, H.B. and R.Y. Qassim (1994) Data envelopment analysis: uma ferramenta para o benchmarking. **Pesquisa Operacional 15**, 13-26.

Spady, R. and A.F. Friedlaender (1976) Econometric estimation of cost functions in the transportation industries, Report 76-13, Center for Transportation Studies, M.I.T.

Talley, W.K. (1988) An economic theory of the public transit firm. **Transportation Research 22B**, 45-54.

Viton, P.A. (1980) On the economics of rapid-transit operations. **Transportation Research 14A**, 247-253.

Viton, P.A. (1981) A translog cost function for urban bus transit. **The Journal of Industrial Economics XXIX**, 287-304.

Viton, P.A. (1992) Consolidations of scale and scope in urban transit. **Regional Sciences and Urban Economics 22**, 25-49.

Wunsch, P. (1996) Cost and productivity of major urban transit systems in Europe: an exploratory analysis. **Journal of Transport Economics and Policy XXX** (2), 171-186.

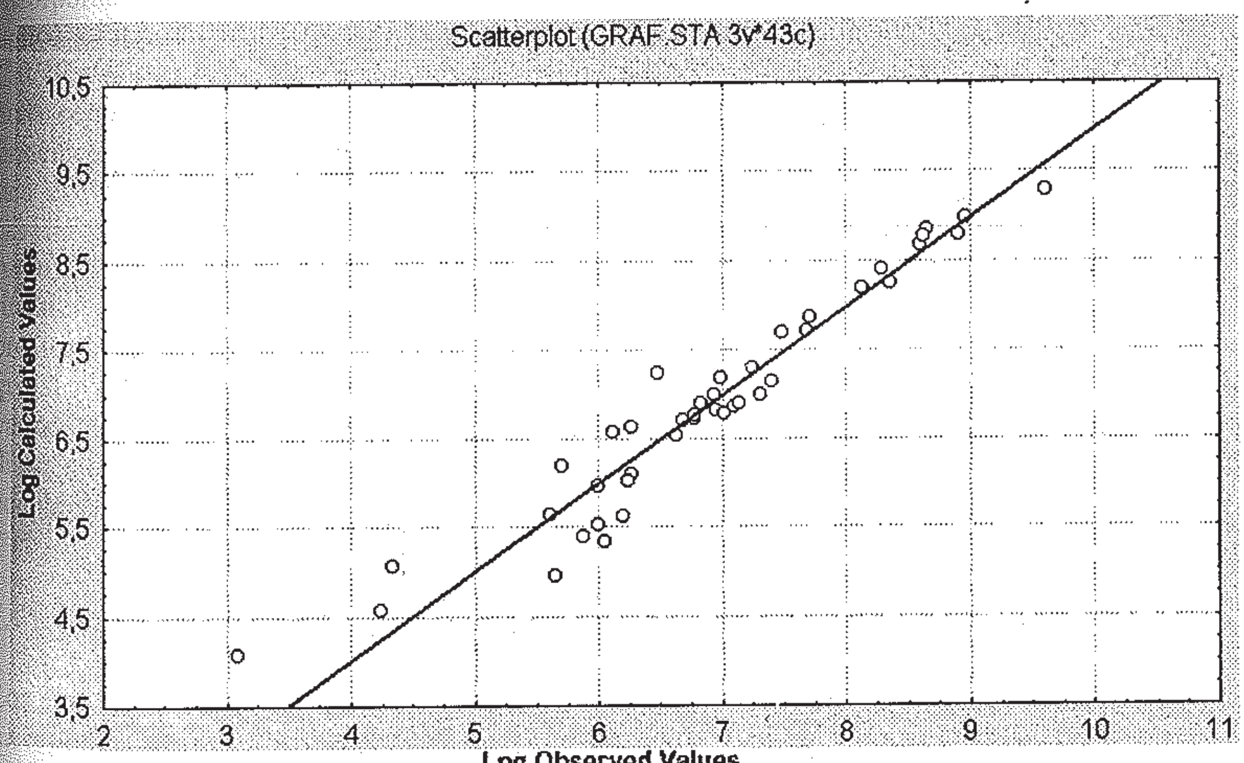

Fig. 1