Comparing the Performance of Urban Transit Bus Routes after Adjusting for the Environment, Using Data Envelopment Analysis

Darold T. Barnum, University of Illinois at Chicago
Sonali Tandon, Chicago Transit Authority
Sue McNeil, PE, M.ASCE, University of Delaware

Great Cities Institute
College of Urban Planning and Public Affairs
University of Illinois at Chicago

Great Cities Institute Publication Number: GCP-08-05

A Great Cities Institute Working Paper
April 2008

GREAT Cities
UIC'S METROPOLITAN COMMITMENT

UIC

The Great Cities Institute
The Great Cities Institute is an interdisciplinary, applied urban research unit within the College of Urban Planning and Public Affairs at the University of Illinois at Chicago (UIC). Its mission is to create, disseminate, and apply interdisciplinary knowledge on urban areas. Faculty from UIC and elsewhere work collaboratively on urban issues through interdisciplinary research, outreach and education projects.

About the Author

Darold T. Barnum, University of Illinois at Chicago is Professor of Management and Professor of Information & Decision Sciences at the University of Illinois at Chicago. He formerly was an Associate Director at the Indiana University Institute for Urban Transportation, where he participated in the training of transit managers from across the nation. His research focuses on performance measurement.

Sonali Tandon, Chicago Transit Authority is a Transit Research Analyst for the Chicago Transit Authority.

Sue McNeil, University of Delaware is Professor of Civil and Environmental Engineering at the University of Delaware. She formerly was Director of the University Illinois at Chicago Urban Transportation Center, and Professor of Urban Planning & Policy. Her research and teaching interests focus on transportation infrastructure management with emphasis on the application of advanced technologies, economic analysis, analytical methods, and computer applications.

Great Cities Institute Publication Number: GCP-08-05

Great Cities Institute (MC 107)
College of Urban Planning and Public Affairs
University of Illinois at Chicago
412 S. Peoria Street, Suite 400
Chicago IL 60607-7067
Phone: 312-996-8700
Fax: 312-996-8933
http://www.uic.edu/cuppa/gci

Comparing the Performance of Urban Transit Bus Routes after Adjusting for the Environment, Using Data Envelopment Analysis

Abstract

Urban transit managers strive to attain multiple goals with tightly constrained resources. Ratio analysis has evolved into a powerful tool for dealing with these goals and constraints. Ratio analysis provides analytical methods for comparing the performance of multiple agencies, as well as the performance of subunits within a particular agency, in order to identify opportunities for improvement. One ratio analysis procedure that has become increasingly popular is Data Envelopment Analysis (DEA). DEA yields a single, comprehensive measure of performance, the ratio of the aggregated, weighted outputs to aggregated, weighted inputs. This paper makes two contributions to the practice of urban transit performance evaluation using DEA. First, instead of using DEA to compare the performance of multiple transit systems, it uses DEA to compare the performance of multiple bus routes of one urban transit system. Second, it introduces a new procedure for adjusting the raw DEA scores that modifies these scores to account for the environmental influences that are beyond the control of the transit agency.

INTRODUCTION

Data Envelopment Analysis (DEA) is being increasingly used to analyze urban transit agency performance. De Borger, Kerstens and Alvaro (2002) identify 15 articles published prior to 2001, and since then at least 15 more have been published or are in press (Boilé 2001; Nolan et al. 2001; Novaes 2001; Odeck and Alkadi 2001; Pina and Torres 2001; De Borger et al. 2002; Nolan et al. 2002; Karlaftis 2003; Boame 2004; Karlaftis 2004; Brons et al. 2005; Sheth et al. 2006; Graham 2006; Barnum et al. 2006; Odeck 2006). Many of the techniques and applications that appear in these articles serve as the foundation and motivation for this paper. These are discussed in the following two sections.

One advantage of DEA in the analyses of urban transit performance is well-stated in one of the earliest urban transit DEA studies. Chu, Fielding and Lamar (1992, p. 224) advise that "performance analysis needs to progress from multiple measures and partial comparison to more robust indicators of performance . . . so that the achievements of one agency can be examined in reference to peer group agencies." DEA provides a single, comprehensive measure of an agency's technical efficiency, that is, the ratio of its aggregated, weighted outputs to its aggregated, weighted inputs.

Another advantage of DEA is the method by which it assigns weights to inputs and outputs. DEA uses linear programming to assign weights in an objective, economically sound manner (Färe et al. 1994; Cooper et al. 2004; Färe and Grosskopf 2004). The final DEA weights are different for every unit being analyzed. Each unit's final weights are assigned so it will attain the best possible score when it is compared to peer units that temporarily are assigned the same set of weights. If the focus unit does not score 100 percent, this tells us that its peers are still more productive even when the weights of all are set to maximize the score of the focus unit. So, not only is there sound economic justification for the weights assigned, but also no inefficient unit can complain that its score would have been better if only a different set of weights were used. Thus, as Chu, Fielding and Lamar (1992) recognized, DEA is uniquely equipped to fulfill the need for an overall, objective, summary performance indicator.

With this paper, we illustrate two techniques that we think will become common in urban transit DEA applications for both research and practice. First, we apply DEA to the actual bus route data of an urban transit agency, illustrating the value of and the methodology for using DEA to compare subunits of urban transit systems. Second, we illustrate an improved method for modifying DEA scores to account for environmental differences among those routes being analyzed.

In the next section we describe the value of analyzing subunits with DEA, and in the following section we evaluate methods that can be used to adjust DEA scores for environmental differences. Then, we discuss the analysis of subunits. We describe our inputs and outputs. We present our statistical models and use them to adjust the outputs for external influences. We then present our DEA model and conduct the DEAs. Finally, we present and discuss the results, and then conclude the paper.

VALUE OF ANALYZING ORGANIZATIONAL SUBUNITS WITH DEA

The unit of analysis usually is called a Decision Making Unit (DMU). Although treating entire transit systems as the DMUs is very helpful in comparing different agencies, it does little to help a given agency evaluate its internal activities, especially those with multiple objectives.

For example, multiple quality and quantity objectives apply to bus operations on a given route. These objectives may include directives such as minimize service interval, maximize span of service, maximize passenger trips, maximize passenger revenue, maximize on-time performance, and many others, constrained, of course, by budget (Benn 1995). A single composite indicator that fairly and objectively aggregates such various activities, so the performance of bus routes could be validly compared, would be very useful. Such an indicator would quickly pinpoint any route whose overall performance was low. This would allow management by exception, that is, management would not have to continuously monitor many indicators for all routes, but could react to those routes that clearly were troubled. Also, it would provide a valuable indicator for planners as they attempted to improve the route structure to enhance future performance. DEA indicators could fulfill this need for a summary measure of parallel internal activities, just as it has been useful in comparing entire agencies.

DEA comparisons of organizational subunits are common in other industries (Cooper et al. 2004). However, all but two of the 30 transit articles cited in our first paragraph have used entire transit systems as the unit of analysis. Sheth, Triantis and Teodorovic (2006) suggest an innovative method that combines DEA and goal programming to compare bus routes, demonstrating their procedure with artificial, simulated data. Barnum, McNeil and Hart (2006) combine DEA and Stochastic Frontier Analysis to compare park-and-ride lots, using data from the Chicago Transit Authority. Tandon (2006) uses DEA to compare bus routes using data from a large American bus system. This paper is a continuation of the fruitful trend of analyzing subunits in order to improve urban transit performance, and is based on Tandon's (2006) data, research and thesis.

CORRECTIONS FOR ENVIRONMENTAL INFLUENCES

Another important issue for DEA is the need to identify environmental influences in order to explain variations in efficiency caused by factors external to the DMUs. It is important for policy purposes, and it is needed in order to adjust for exogenous factors so one can correctly evaluate the endogenous efficiency of individual DMUs. Endogenous efficiency often is called managerial efficiency or true efficiency, because it represents the efficiency under the agency's control. Exogenous influences usually are called environmental influences or contextual influences.

The Two-Stage Method

In transit DEA applications, the "two-stage method" has been used to identify exogenous influences beyond the control of the DMUs (Nolan 1996; Kerstens 1996; Viton 1997; Nolan et al. 2001; Odeck and Alkadi 2001; Pina and Torres 2001; Boame 2004; Barnum et al. 2006; Odeck 2006). With this procedure, a DEA is first conducted using only traditional inputs and outputs. In the second stage, most commonly the DEA scores are regressed on the exogenous variables of interest. The regression outcomes are used to identify exogenous factors that influence the first-stage DEA scores to a statistically significant degree. In a few of the articles (Viton 1997; Odeck and Alkadi 2001; Odeck 2006), the second stage has been more limited, comparing the mean efficiencies of categorical groups for statistically significant differences. However, only one of these articles uses its statistical results to estimate each DMU's endogenous efficiency (Barnum et al. 2006), which is needed if one wishes to compare DMUs.

Although it has been long suspected (Grosskopf 1996), Barnum and Gleason (2006a) recently have demonstrated that the two-stage procedure can exhibit substantial bias, low precision, and low power. Barnum and Gleason's results were based on asymptotic assumptions. However, the use of bootstrapping methods with the two-stage procedure (Simar and Wilson 2007) also has been shown to have low power in detecting true relationships (Zelenyuk 2005).

The Exclusion Method

An alternate approach to adjust DEA scores for exogenous influences is a procedure we call the exclusion method. Under the exclusion method, each target DMU is compared only to those other DMUs in the set that have an equal or less favorable environment (Ruggiero 2004a; Ruggiero 2004b; Muniz et al. 2006). An index is generated that aggregates the combined effect of environmental influences, with the variables defined so that larger values denote a less favorable environment (Equation 1).

$$z_j = \sum_{i=1}^{n} \beta_i z_{ji} \tag{1}$$

In Equation 1, z_{ji} is the value of exogenous influence *i* for DMU *j*, β_i is the weight measuring the strength of the effect of exogenous influence *i* on the efficiency of any DMU, and z_j is the aggregated effect of exogenous influences on DMU *j*. The β_i can be defined by expert judgment or other means (Sheth et al. 2006). In the subsequent DEA, DMU *j* is compared only to those DMUs *k* where $z_j \leq z_k$.

Unfortunately, this procedure has even more serious shortcomings than the two-stage method. The DMU with the highest *z*-score is compared only to itself, so it will always be reported efficient, regardless of its true endogenous efficiency level. Likewise, if there are *m* inputs and *n* outputs, then the *m*+*n* DMUs with the highest *z*-scores have a very high probability of being found efficient by random chance, regardless of their true efficiency. This is because of the well-known “curse of dimensionality” that afflicts deterministic methods such as DEA (Cooper et al. 2000; Simar and Wilson 2000).

Sheth, Triantis and Teodorovic’s (2006) simulation illustrates the problem. Their randomly-generated data means that there is no relationship between the environmental index and endogenous efficiency, so each DMU should have an equal probability of being found efficient after accounting for environmental factors. In fact, for the Variable Returns to Scale model, eight of the DMUs with the top ten *z*-scores were reported efficient, including of course the first one. Of the remaining 50 DMUs in their sample, only 14 were reported efficient. For the Constant Returns to Scale model, 6 of the top 10 were reported efficient and only 3 of the bottom 50 attained efficiency.

It is not only those DMUs with the worst environments whose DEA scores will be biased toward greater efficiency than truth. A favorable bias from a different source will be present for all of the DMUs. Suppose there are two DMUs that have equal endogenous efficiency. But, DMU *B* operates in a "Bad Environment" while DMU *G* enjoys a "Good Environment." So, although they have equal endogenous efficiency, *B* will need more resources than *G* to produce the same amount of outputs. If *G* were to be compared to *B*, then because *B* uses more resources, *G*'s reported efficiency score would be biased in the favorable direction. The amount of bias will vary, depending on the size of the environmental differences between a target DMU and its benchmark peers. And, the bias will increase as the impact of the environment grows. Thus, the exclusion method will be most biased and the bias will be most inconsistent when it is most needed, where the environment has a large effect. A final problem is that because the score of each DMU is computed from a different set of competitors, the scores of any two DMUs cannot be validly compared to each other.

The Reverse Two-Stage Method

Without doubt, it is necessary to account for exogenous influences in order to explain variations in DEA scores caused by factors external to the DMUs. Because the conventional two-stage method is suspect and the exclusion method is invalid, finding a better alternative method is important.

In this paper, we apply a new method, which in essence reverses the order of the steps of the conventional two-stage procedure. The reverse two-stage method has been shown with simulated data to yield estimates without the bias, precision and power problems that compromise the conventional two-stage method (Barnum and Gleason 2006a). Perhaps of equal importance, the reverse two-stage procedure yields individual DEA scores already adjusted for exogenous influences, so DMUs' managerial performance can be more easily compared. Below are the steps of the reverse two-stage procedure.

In the first stage the outputs are adjusted for environmental differences. Each output is regressed on all inputs and the exogenous factors expected to influence it. For example, suppose there is one output, y, M traditional inputs x_m, and N exogenous factors z_n. Then, the regression to be estimated would be:

$$y = \alpha + \beta_1 x_1 + ... + \beta_M x_M + \gamma_1 z_1 + ... + \gamma_N z_N \qquad (2)$$

To finish the first stage, each output is adjusted for environmental differences by removing the marginal influence of the statistically significant exogenous variables, that is:

$$y_{adj} = y - (\gamma_1 z_1 + ... + \gamma_N z_N) \qquad (3)$$

for γ_n different from 0 to a statistically significant degree

Note that only the expected value of the influence of the exogenous variables is removed; left in is the error term and any inefficiency in converting inputs to output.

In the second stage, the adjusted outputs and inputs are analyzed by DEA, with the resulting scores being independent of exogenous effects. The same procedure could be applied to inputs, and more complex regression models should be used where appropriate.

COMPARISONS OF ORGANIZATIONAL SUBUNITS vs. ENTIRE ORGANIZATIONS

There are important differences when the unit of comparison is an organizational subunit instead of an entire organization. They involve the influence of environmental variables, and the impact of other organizational levels on the subunits being analyzed.

Influence of Environmental Variables in Comparisons of Organizational Subunits

It is well known that statistical analysis of multiple organizational levels must proceed differently than the analysis of a particular level (Wooldridge 2002; Raudenbush and Bryk 2002;

Skrondal and Rabe-Hesketh 2004). For our case, the factors influencing the response variables may differ among levels, and even factors that are the same may have differing effects.

For example, population density may have a substantial influence on efficiency when comparing agencies in different cities, and there is little management can do to affect this difference. But, within a given city, management has control over all of the routes, so it can match service levels to the numbers of riders, so even large differences in density might cause no differences among routes in efficiency. That is, management has the ability to adjust service for density when all of the routes are under its control, and, if it does so, then all routes could be equally efficient regardless of density differences. Therefore, density might appear to have a small empirical effect on efficiency when comparing routes, but might seem to have a large empirical effect on efficiency when comparing systems in different cities.

Of course, public policy often requires that service not fall below a specified minimum in any part of a service area regardless of environmental differences, with subsidy revenue expected to cover the inevitable losses. Then, within a given system, we would expect routes in unfavorable environments to have lower unadjusted efficiency than routes in favorable environments. If environmental differences were the sole reason for differing efficiency levels, then after those environmental differences have been adjusted for, all of the resultant efficiency ratios should be equal.

Thus the meaning of, reasons for, and managerial response to efficiency levels may differ when comparing subunits within a transit organization than when comparing entire organizations.

Influence of Other Organizational Levels in Comparisons of Organizational Subunits

"Managerial efficiency" generally refers to outcomes resulting from the decisions and activities of personnel within the entire organization, because typically the organization itself is the DMU. By analogy, when a DMU is a subunit of an organization, it would seem that the efficiency of personnel within each subunit is involved.

However, when subunits are the DMUs, decisions and activities may occur both within the subunit and at higher organizational levels. Responsibility for managerial performance depends on who controls the subunit decisions and activities, and this can depend on the nature of the inputs and outputs involved.

For the particular set of bus route inputs and outputs that we use, the efficiency of each route is primarily the result of decisions by the agency's planners and schedulers, although partly the result of the performance of those responsible for the supervision of a given route. Informed by the agency's service standards, the planners and schedulers determine span of service, average frequency and maximum frequency, thereby also influencing seat kilometers and seat hours. And, because of these decisions, they also are partly responsible for ridership levels and on-time performance, although responsibility for on-time performance is also affected by the activities of personnel within each DMU.

Based on the DEA, planners and schedulers could adjust the inputs and most outputs to change a route's future efficiency score. If a route is performing badly on the variables used herein, it is not the supervisors of a given route or garage who are primarily responsible, but the planners and schedulers who can adjust a route's inputs and controllable outputs to attain the efficiency level desired for that route.

The main purpose of the DEAs herein, therefore, is to inform agency planners and schedulers about efficiency differences, so they can compare the effects of their decisions on the various routes, and change inputs and controllable outputs accordingly. However, on-time performance also is influenced by personnel at the garage level. In sum, on-time performance is likely to be influenced by (1) scheduling decisions, (2) garage supervision, and (3) environmental conditions.

In order to better measure the impact of environmental conditions, it would be worthwhile to control by garage. The effect of garage does not tell us the source of performance differences, but it does allow us to more validly measure the effect of environment. Therefore, for the on-time performance regression, we have added dummy variables ($G_2, \ldots, G_6$) to identify the garage from which each route originates. The garage with the best on-time performance serves as the base, with the dummy variable coefficients reflecting the percentage by which the other garages' on-time performance is lower than the best garage.

DATA SET, AND INPUTS AND OUTPUTS

This paper analyses data from 46 bus routes of a United States urban transit agency, treating each route as a DMU. Data on the inputs and outputs are for the Spring 2005 weekday trips.

Inputs are the resources that supply the transit service. Outputs are the variables that measure the use and the quality of transit service. Only the variables used in this analysis are described here, with all of the considered variables being discussed in Tandon (2006).

Inputs

The seat kilometers (*SK*) and seat hours (*SH*) of each route are used as proxies for the route's use of energy, maintenance, labor and capital resources. These proxies were used for several reasons. First, data on the underlying physical resources were not available at the route level. Second, as discussed by Chu, Fielding and Lamar (1992), the physical resources are used to produce transit service (as measured by such variables as vehicle hours or kilometers), and transit service is used to produce riders and other consumed outputs. This flow would require two DEAs, one concerning the ratio of produced output to physical inputs, and another concerning the ratio of consumed outputs to produced inputs (with the outputs produced in the first step becoming inputs in the second step). In this paper, we focus on the consumed outputs to produced inputs ratio. Third, DEA assumes that there is substitutability among the inputs (Petersen 1990; Barnum and Gleason 2006b). For transit, such substitutability would mean that for a given level of output, a DMU could freely substitute labor for vehicles, or vehicles for fuel. In truth, there is extremely limited substitutability in this industry; inputs have to be combined in virtually fixed ratios, with any excess wasted. These facts, combined with the “curse of dimensionality” and random noise in the data, mean that disaggregating inputs by physical type biases efficiency scores upward.

Seat kilometers and hours are highly but not perfectly correlated because of speed differences and each measures a somewhat different aspect of input usage. Generally, even if two key inputs are highly correlated, both should be included in a DEA (Nunamaker 1985). Seat hours and kilometers are used in place of vehicle hours and kilometers to account for the different lengths of buses operated by the agency.

Seat hours are the number of seats on a bus multiplied by the total revenue and non-revenue hours traveled by all the buses on a particular route during a weekday.

Seat kilometers are the number of seats on a bus multiplied by the total revenue and non-revenue kilometers traveled by all the buses on a particular route during a weekday.

Outputs

The purpose of this study was to assist the agency management in identifying differences in route performance. DEA scores will be useful to management only if the outputs are chosen to reflect the goals of the transit agency, which may not be the goals that interest groups, public policy makers, or researchers feel the system should be pursuing (Gleason and Barnum 1982; Barnum 1987). It is important to note that outputs considered key for individual route performance may be different than those that are key for the entire system’s performance, another potential difference between organizational and subunit analysis.

Therefore, agency management chose the outputs. Five output variables are included in the analysis: ridership, span of service, average frequency, maximum frequency, and on-time performance. Whereas transit ridership is a service usage measure, all others are service quality measures.

Ridership (*Riders*) is the number of unlinked passenger trips. It also serves as a proxy for the farebox revenue generated by the route, because the urban transit agency told us that mean fares would be relatively consistent across routes. Revenue, although sometimes considered an input, is an important output according to many transit agencies, given the emphasis often placed on farebox recovery rates.

On-time performance (*OTP*) is the proportion of observed trips that depart the starting point of the trip on time, where “on time” is less than 1 minute early or less than 5 minutes late.

Span of service is the total minutes per day of transit service provided on a route in one direction. Span of service is a measure of service availability and therefore service quality. A shorter span of service indicates degraded quality of service and can discourage people from making certain types of off-peak trips by transit like shopping and entertainment trips.

Average Daily Service frequency is a measure of how well the route is served throughout the day. It is measured as average number of buses serving a stop every hour and is calculated by dividing the total number of runs on a route in both directions by the total span of service on that route.

Maximum Daily Service frequency is the number of buses servicing a stop in the hour with the highest frequency. This measure reflects how well a route is served during peak hours.

STAGE ONE: ADJUSTING THE OUTPUTS FOR ENVIRONMENTAL INFLUENCES

Under the reverse two-stage method, influences of the environment are removed from the outputs before a DEA is conducted. The applicable outputs are each regressed on the traditional input variables and those environmental factors expected to influence them. The outputs are adjusted to remove the marginal effect of the statistically significant environmental variables.

Dealing With Environmental Influences on Ridership and On-time Performance

Average Frequency, Maximum Frequency and Span of Service are under the direct control of management, and therefore should not be adjusted for exogenous influences.

On the other hand, the number of passenger trips (*Riders*) is influenced by environmental factors not under management control, as is on-time performance (*OTP*), so these two outputs must be adjusted for environmental effects. The environmental factors used herein are the ones that both the agency and the researchers agreed had the most influence on *Riders* and *OTP* in this case, given the particular set of outputs being used, the available data, and the city and the routes involved. They include the following.

<u>Population Density</u> (*PopDen*) is the mean population density within 0.4025 km of a route over its entire length. Population density is expected to influence the on-time performance, because denser areas are likely to cause more frequent, unexpected delays. (The variation in density along a route might also be a factor, especially for routes that run through both city and suburbs. Our routes are solely in the city.)

<u>Population</u> (*Pop*) is the population within 0.4025 km of the route over its entire length. Because it encompasses the number of potential riders, it should be related to the actual number of riders.

<u>Title 6 Routes</u> (*T6*), as commonly defined by the Federal Transit Administration (FTA), have

at least 1/3 of their total route length in census tracts or traffic analysis zones where the percentage of minority population greater than the percentage of minority population in the entire transit service area. These routes serve those areas where incomes and car ownership often are lower, so people in the area are more likely to be dependent on transit and therefore are likely to ride in greater numbers. Title 6 routes therefore serve as a proxy for routes with higher proportions of captive riders. They are coded as a dummy variable (Title 6 = 1, Otherwise=0), and the interaction between population and Title 6 routes $(Pop \times T6)$ is included as an environmental variable.

Key Routes (*Key*) are designated by the transit agency as those routes that are most productive as well as some routes necessary to meet geographic coverage standards. Routes not designated as key routes are called service routes. Service Routes are mainly routes that connect key routes or serve rail stations. The levels of service on service routes are tied to demand coming from other public transportation as well as supplemental demand coming from the population adjacent to the route.

Because service routes provide connections to other public transportation, they contribute to the demand for the other transportation. Hence, their value cannot be measured solely by the number of passengers they carry, because many of those passengers connect to other routes from the service routes. Thus, it is important to consider their dual role of carrying their own riders and carrying riders contributing to other routes' performance.

Here is how we deal with the differences between key routes and service routes. The dummy variable (*Key*) is coded 1 when the route is a key route and 0 when it is a service route. The interaction between population and key routes $(Pop \times Key)$ is included as an independent variable to estimate the difference between the two route types in the effect of population on ridership. This variable allows us to identify any differences in ridership between the two types of routes, and, if differences are found, to remove those differences in the adjusted scores. In this way, service routes are neither advantaged nor disadvantaged by their dual purpose.

Interaction among Key Routes, T6 Routes, and Population $(Pop \times Key \times T6)$. Because some T6 routes are not key routes, and many key routes are not T6 routes, we also need to check for a three-way interaction when a route is both a key route and a T6 route.

Let us identify the expected effects of the preceding environmental variables on ridership. As a result of the dummy variables and their interactions with population and each other, the regression coefficient of the variable *Pop* identifies the effect of population on ridership for routes that are neither key nor T6 routes, that is, non-T6 service routes. These routes are the subset of the service routes that do not traverse T6 areas, so the adjacent population plays a less important role in supplying riders. Therefore, we would expect the regression coefficient of *Pop* to be positive, but relatively small. The regression coefficient of $Pop \times Key$ estimates the difference between the slope coefficients of the non-T6 service routes and non-T6 key routes. The regression coefficient $Pop \times T6$ estimates the difference between the slope coefficients of T6 and non-T6 service routes. If any or all of the three regression coefficients are statistically significant, we expect them to be positive, because we expect the effect of population on ridership for routes that are non-T6 service routes to be the smallest of all. We don't predict direction of the regression coefficient of the three-way interaction, $Pop \times Key \times T6$. If the joint effect of Key and T6 routes is greater than the sum of their individual effects, then the coefficient will be positive; if their joint effect is less than the sum of their individual effects, then the coefficient will be negative; and if their joint effect equals the sum of their individual effects, the coefficient will be zero.

Additional environmental variables not used here might be important in different circumstances. Some unused variables such as “parking availability” are roughly equal for all of our routes. Other common variables such as employment and employment density did not have any measurable influence for these routes, perhaps because the routes’ configuration did not result in significant employment variations. For our study, often-important influences on *OTP* such as time of day and day of week would not be factors, because our *OTP* variable was the mean value from all weekday runs, that is, the values from all times of day and all five weekdays were aggregated. Also, we think that traffic intensity and type of streets would be significantly correlated with population density, given that *OTP* is the mean value for all times of day and all weekdays; a substantial part of their effect should be picked up by the regression coefficient of population density and thereby will be corrected for in the adjusted DEA scores (Wooldridge 2002; Greene 2003). Although these exogenous factors are not expected to influence our outputs, in cases where they vary independently they should be used in statistical analyses.

Regressing the Response Variables on Service Inputs and Environmental Influences

The two equations to be estimated are as follows:

$$Riders_j = \alpha + \beta_1 SH_j + \beta_2 SK_j + \beta_3 Pop_j + \beta_4 (Pop \times Key)_j + \beta_5 (Pop \times T6)_j + \beta_6 (Pop \times Key \times T6)_j \quad (4)$$

$$OTP_j = \alpha + \beta_1 SH_j + \beta_2 SK_j + \beta_3 PopDen_j + \beta_4 G_2 + \beta_5 G_3 + \beta_6 G_4 + \beta_7 G_5 + \beta_8 G_6 \quad (5)$$

Before proceeding to the estimations of Equations 4 and 5, we report on tests of the hypothesis that the two equations are related, and the hypothesis that the service levels (*SH* and *SK*) are econometrically endogenous.

First, we tested the hypothesis that Equations 4 and 5 are related, because there may be correlation between unobserved factors common to both equations but not included as independent variables. Such correlation would cause the estimated residuals of the two equations to be correlated. If this hypothesis is true, then the equations can be estimated jointly using Zellner's Seemingly Unrelated Regression Estimation (SURE) model for systems of equations, which will increase the statistical efficiency of the estimates (Zellner 1962). Unfortunately, there was neither a statistically significant correlation between the residuals of the two equations nor an increase in model fit. Using Feasible Generalized Least Squares, ρ = 0.0271, and Breusch-Pagan test of independence of the errors for the two equations was: χ^2 = 0.034, Prob ($\chi^2(1)$ > 0.034) = 0.8541. Likewise, use of SURE did not increase the model fit over that provided by Ordinary Least Squares (OLS). The SURE and OLS Root Mean Squared Errors were the same to five places for both Equation 4 (2792.7) and Equation 5 (0.06214), and the likelihood ratio test of the increase in fit from the SURE model was χ^2 = 0.04, Prob ($\chi^2(1)$ > 0.04) = 0.8405.

Second, we tested the hypothesis that the service inputs (*SH* and *SK*) are econometrically endogenous. Such endogeneity would result if the number of passengers on a route or a route's on-time performance significantly influenced the level of service on that route. If such endogeneity is present, it would be necessary to replace *SH* and *SK* with Instrumental Variables (IVs). The conventional statistical test involves comparing the original OLS equations with IV equations in which the suspected variables are replaced with IVs, and accepting the hypothesis of endogeneity if the equations differ to a statistically significant degree. Applying the Hausman Specification Test, the OLS and IV equations with Riders as the response variable did not differ to a statistically significant degree [χ^2 = 1.11, Prob ($\chi^2(6)$ >1.11) = 0.9811]. Again applying the Hausman Specification Test, the OLS and IV equations with On-time Performance as the response variable did not differ to a statistically significant degree [χ^2 = 1.25, Prob ($\chi^2(8)$ >1.25) = 0.9961]. Therefore, the hypotheses that econometric endogeneity is present cannot be accepted.

It may seem curious that the number of riders could not be shown to affect level of service. This may have occurred because the agency's service standards support equity for all areas of the city, which results in a fairly high minimum level of service on all routes, regardless of ridership levels. Although this agency increases service if ridership increases to the point that the buses are over-crowded, usually any increases in ridership can be absorbed by the current frequency and size of buses, without increasing service. Likewise, the minimum level of service requirement generally means that decreases in ridership will not lower service levels.

Let us return to estimation of the parameters of Equations 4 and 5. We used OLS regression, with robust variance-covariance estimation in order to adjust for arbitrary heteroskedasticity and serial correlation. For Equation 5, the proportion of runs on time (*OTP*) did not involve censored data (0.56 < *OTP* < 0.90), so it was unnecessary to use a censored-data regression model (Breen 1996). By using robust estimation procedures to deal with potential heteroskedasticity and serial correlation, the procedure yields more valid tests of statistical significance (Wooldridge 2002).

The results of the regressions are presented in Tables 1 and 2. The regression coefficients and statistical significance of seat hours and seat kilometers are immaterial. The two variables are highly multicollinear, but, more importantly, they are merely serving as control variables. At the 0.05 level, in Equation 5 only $Pop \times Key \times T6$ is statistically significant, and in Equation 6 $PopDen$ and the entire garage dummy variables are statistically significant.

Table 1. Robust Regression of Riders on Endogenous and Exogenous Variables

Riders	Coefficient	Robust Standard Error	t-ratio	P(*t*)>t
SH	0.0214	0.0071	3.00	0.005
SK	-0.0295	0.0199	-1.48	0.147
Pop	-0.0071	0.0122	-0.58	0.783
*Pop*Key*	0.0241	0.0161	1.50	0.071
*Pop*T6*	0.0173	0.0116	1.49	0.073
*Pop*Key*T6*	0.0464	0.0205	2.26	0.030
Constant	-331.1825	476.1243	-0.70	0.491

R^2 = 0.8846; F(6, 39) = 48.58; P[F(6,39)>48.58] < 0.00005;
Probabilities for *Pop*, *Pop*Key*, and *Pop*T6* are one-tail.

Table 2. Robust Regression of On-time Performance On Endogenous and Exogenous Variables

OnTimePerformance	Coefficient	Robust Standard Error	t-ratio	P(*t*)>t
PopDen	-0.0563	0.0271	-2.08	0.023
SH	0.1837	0.1099	1.67	0.103
SK	-0.7164	0.3496	-2.05	0.048
Garage 2	-0.1495	0.4251	-3.52	0.001
Garage 3	-0.0862	0.3653	-2.36	0.024
Garage 4	-0.0836	0.1945	-4.30	0.000
Garage 5	-0.0424	0.1883	-2.25	0.030
Garage 6	-0.1060	0.0286	-3.70	0.001
Constant	0.8884	0.0223	40.09	0.000

Note: *SH* and *SK* values in millions.

R^2 = 0.4863; F(8, 37) = 17.51; P[F(8,37)>17.51] < 0.00005

Probability for *PopDen* is one-tail

Adjusting the Outputs

We use the expected values of the regression coefficients of the environmental variables that were statistically significant at the 0.05 level, $Pop \times Key \times T6$ for ridership and *PopDen* for on-time performance, to remove the effects of these variables, as shown in Equations 6 and 7. Note that these adjusted values retain any inefficiency and random noise contained in each DMU's residual error. Only the expected marginal effects of the exogenous variables are removed.

$$\text{Adjusted } Riders_j = Riders_j - \beta_6 Pop \times Key \times T6_j \tag{6}$$

$$\text{Adjusted } OTP_j = OTP_j - \beta_3 PopDen_j \tag{7}$$

STAGE TWO: DATA ENVELOPMENT ANALYSIS

The DEA procedure has been well-explained in most of the 30 transit DEA studies cited in our first paragraph, including the paper by Boilé (2001) in this journal, and in general texts (Färe et al. 1994; Cooper et al. 2004; Coelli et al. 2005). Therefore, we do not repeat those explanations here.

Our DEA scores are computed using Linear Program (8), which is output oriented and assumes constant returns to scale. The DEAs were conducted with Scheel's EMS software (2000).

$$\begin{aligned}
&\max_{\boldsymbol{\lambda}} \theta \\
&\text{subject to} \quad \sum_{j=1}^{46} y_{jm}\lambda_j \geq \theta y_{km} && m = 1,2,3,4,5 \\
&\qquad\qquad\quad \sum_{j=1}^{46} x_{jn}\lambda_j \leq x_{kn} && n = 1,2 \\
&\qquad\qquad\quad \lambda_j \geq 0 && j = 1,2,\dots,46
\end{aligned} \tag{8}$$

For each of the j DMUs (j = 1,…,46), there are data on the n = 2 inputs x $(x_{11},...,x_{jn})$, and on the m = 5 outputs y $(y_{11},...,y_{jm})$. The DEA score θ identifies the technical efficiency of the target DMU k. The program was run twice for each DMU, once with unadjusted outputs and once with adjusted outputs, so the marginal effects of the environment on efficiency could be identified.

THE RESULTS

There are two things worth looking for in the following scores. First are the differences between the unadjusted and adjusted scores. If the difference is relatively large for a given route, then it would be wise to carefully study that route. The second thing we need to look for is adjusted scores that are substantially higher than 1. These routes should be thoroughly analyzed for correctable problems, including comparisons with their best-practice benchmarks.

Best-practice routes are those found to be efficient, and those best-practice routes whose input and output proportions most closely mirror an inefficient route become its benchmarks. Those best-practice routes that are benchmarks for large numbers of inefficient routes are especially worthy of study, because they are the efficient routes that have most in common with many other routes.

In some cases management may find that there are uncorrectable factors that result in low efficiency, or that there are other justifiable reasons for the high adjusted scores. In other cases, the high adjusted scores may identify routes that can and should be improved. Methods for improvement may be discovered by looking at the practices of an inefficient route's best-practice benchmarks.

The unadjusted and adjusted DEA efficiency scores are shown in Table 3. Of the 46 routes, 20 became more efficient, 12 did not change, and 14 became less efficient.

Table 3. Data Envelopment Analysis Results, by Decision Making Unit

ID	Unadj. Score	Adj. Score	Change	Change	Benchmarks
1	1.22	1.20	better	-0.02	10, 23, 33, 44
2	1.70	1.72	worse	0.02	15, 23, 28
3	1.20	1.30	worse	0.10	10, 23, 28
4	1.00	1.06	worse	0.06	23
5	2.40	2.88	worse	0.49	23, 28, 29
6	1.07	1.06	better	-0.01	23, 29, 33, 44
7	1.38	1.76	worse	0.37	10, 23, 25
8	1.10	1.24	worse	0.14	23, 29
9	2.19	2.06	better	-0.13	23, 28
10	1.00	1.00	same	0.00	Benchmark for 14
11	1.37	1.25	better	-0.12	10, 23
12	1.95	1.88	better	-0.07	15, 23, 28
13	1.58	1.54	better	-0.04	15, 23, 25, 33
14	1.14	1.39	worse	0.25	23
15	1.00	1.00	same	0.00	Benchmark for 13
16	1.35	1.25	better	-0.11	10, 23, 29, 33, 44
17	1.87	1.87	worse	0.00	15, 23, 28, 29, 33
18	1.27	1.21	better	-0.06	10, 23, 25
19	1.36	1.64	worse	0.28	10, 23, 25, 29
20	1.75	2.07	worse	0.32	23, 28
21	1.17	1.18	worse	0.01	15, 23, 29
22	1.06	1.16	worse	0.10	10, 23, 28
23	1.00	1.00	same	0.00	Benchmark for 29
24	1.09	1.31	worse	0.22	10, 23, 25, 29
25	1.00	1.00	same	0.00	Benchmark for 16
26	1.63	1.65	worse	0.01	15, 25, 28, 32
27	1.00	1.00	same	0.00	Benchmark for 1
28	1.00	1.00	same	0.00	Benchmark for 16
29	1.00	1.00	same	0.00	Benchmark for 13
30	1.18	1.18	same	0.00	15, 25, 29, 32
31	1.24	1.24	same	0.00	23, 25, 28, 29
32	1.00	1.00	same	0.00	Benchmark for 5
33	1.00	1.00	same	0.00	Benchmark for 7
34	1.02	1.01	better	-0.01	15, 23, 25, 28
35	2.01	1.98	better	-0.03	10, 25, 28
36	2.01	1.98	better	-0.03	15, 23, 28
37	1.15	1.10	better	-0.05	15, 25, 32, 33
38	2.37	2.36	better	-0.01	15, 23, 25, 28
39	2.06	2.04	better	-0.03	10, 23, 25, 29
40	1.99	1.96	better	-0.03	10, 23, 25, 29
41	2.01	1.94	better	-0.07	23, 28
42	1.68	1.53	better	-0.15	10, 25, 28
43	2.39	2.27	better	-0.12	10, 23, 25, 29
44	1.00	1.00	same	0.00	Benchmark for 3
45	1.65	1.64	better	-0.01	15, 27, 32
46	2.27	2.23	better	-0.04	15, 32, 33

Decision Making Units listed in ID order. Higher scores are worse. Benchmark Decision Making Units are those for adjusted scores.

Identification of route efficiencies, with both original and adjusted scores, is the first step. Next the “red flag” test should be used to identify routes that should receive further attention. Suppose management decides to analyze both (1) routes whose reported efficiency changed by at least |0.30| when adjusted for environment, and (2) routes whose adjusted scores were greater than 2. These criteria would flag 8 of the 46 routes, namely DMUs 5, 7, 9, 20, 38, 39, 43, and 46.

Management needs to look at these eight routes more closely. For example, consider DMU 5, which has been flagged above for low efficiency. The benchmark routes for DMU 5 are 23, 28, and 29 (Table 4). Comparison of the inputs, outputs and environmental variables of these four DMU's shows that DMU 5 has outputs comparable to DMU 23 but DMU 5 needs almost 2.5 times the inputs as DMU 23. Further comparison reveals that DMU 5 is an express service and serves longer distances as compared to DMU 23. DMUs 28 and 29 are both express routes and still are efficient. This indicates that it might be possible to improve the efficiency of DMU 5 by reducing its service hours like DMU 28 and 29 or making it a regular service like DMU 23. However these suggestions may or may not be practical, if non-efficiency considerations also are involved.

Table 4. Decision Making Unit 5 and Its Benchmark Systems, by Outputs and Inputs

ID	Service Span	Mean Freq	Max Freq	Riders	OTP	SK	SH
5	1233	6.11	10.00	11298	0.74	311635	1110833
23	1271	5.71	7.50	10949	0.59	118245	446880
28	291	10.61	15.00	2038	0.57	31788	111510
29	808	4.00	4.00	1106	0.78	20291	145236

Figure 1 is a representation of the benchmark units for each of the 8 DMUs of concern identified in Table 3. DMU 23 serves as a benchmark for all but one of the DMUs of concern. And, DMUs 25 and 29 each serve as benchmarks for half of the DMUs of concern, and together are benchmarks for all but one of these DMUs. This suggests that these common benchmarks warrant closer attention to learn why they are efficient and whether their experiences are transferable to inefficient units.

Figure 1. Selected Inefficient Decision Making Units, by Benchmark Decision Making Units

		Benchmark Decision Making Units							
Inefficient Decision Making Units		10	15	23	25	28	29	32	33
	5								
	7								
	9								
	20								
	38								
	39								
	43								
	46								

Source: Table 3.

CONCLUSIONS

Using DEA to evaluate the efficiency of the subunits of an urban transit system is a promising procedure for improving urban transit performance. This paper demonstrates an improved method for adjusting DEA scores for the environment, and illustrates DEA performance analyses when organizational subunits instead of entire organizations are the focus of study.

ACKNOWLEDGEMENTS

We thank the three anonymous reviewers, whose detailed comments and perceptive suggestions greatly improved our work. Although we made many changes based on reviewer suggestions, the ones that were particularly significant included (1) controlling for each route's home garage in the Equation 5 regression, (2) adding the subsection entitled

"Influence of Other Organizational Levels on Organizational Subunits," (3) conducting tests for Seemingly Unrelated Regressions and econometric endogeneity, leading to stronger confirmation of the validity of the statistical models, and (4) adding a description of the method by which DEA assigns weights, which is one of its most important strengths. We also gratefully acknowledge the support and cooperation of the urban transit agency that provided the data and worked with us to suggest appropriate inputs and outputs.

References

Barnum, D. T. (1987). "Evaluating transit: the performance tree method." *Managing Urban Transportation As a Business*, E. Bers and C. Hendrickson, eds., American Society of Civil Engineers, New York, 216-230.

Barnum, D. T. and Gleason, J. M. (2006a). "Bias and precision in the DEA two-stage method." Forthcoming in *Applied Economics*.

Barnum, D. T. and Gleason, J. M. (2006b). "Biases in technical efficiency scores caused by intra-input aggregation: Mathematical analysis and a DEA application using simulated data." *Applied Economics*, 38(14), 1593-1603.

Barnum, D. T., McNeil, S., and Hart, J. (2006). "Comparing the efficiency of public transportation subunits using data envelopment analysis." Forthcoming in the *Journal of Public Transportation.*

Benn, H. P. (1995). *Bus Route Evaluation Standards, Transit Cooperative Research Program, Synthesis of Transit Practice 10,* Transportation Research Board, Washington, DC.

Boame, A. K. (2004). "The technical efficiency of Canadian urban transit systems." *Transportation Research Part E-Logistics and Transportation Review*, 40(5), 401-416.

Boilé, M. P. (2001). "Estimating technical and scale inefficiencies of public transit systems." *Journal of Transportation Engineering*, 127(3), 187-194.

Breen, R. (1996). *Regression models: Censored, sample-selected, or truncated data,* Sage Publications, Thousand Oaks, Ca.

Brons, M., Nijkamp, P., Pels, E., and Rietveld, P. (2005). "Efficiency of urban public transit: A meta analysis." *Transportation*, 32(1), 1-21.

Chu, X., Fielding, G. J., and Lamar, B. W. (1992). "Measuring transit performance using data envelopment analysis." *Transportation Research Part A: Policy and Practice*, 26(3), 223-230.

Coelli, T. J., Rao, D. S. P., O'Donnell, C. J. and Battese, G. E. (2005). *An introduction to efficiency and productivity analysis,* Springer, New York, NY.

Cooper, W. W., Seiford, L. M. and Tone, K. (2000). *Data envelopment analysis: a comprehensive text with models, applications, references and DEA-solver software,* Kluwer, Boston.

Cooper, W. W., Seiford, L. M. and Zhu, J. (2004). *Handbook on data envelopment analysis,* Kluwer Academic Publishers, Boston, MA.

De Borger, B., Kerstens, K., and Costa, A. (2002). "Public transit performance: What does one learn from frontier studies?" *Transport Reviews*, 22(1), 1-38.

Färe, R. and Grosskopf, S. (2004). *New directions: Efficiency and productivity,* Kluwer Academic Publishers, Boston.

Färe, R., Grosskopf, S. and Lovell, C. A. K. (1994). *Production frontiers,* Cambridge Univ. Press, Cambridge, Eng.

Gleason, J. M. and Barnum, D. T. (1982). "Toward valid measures of public sector productivity: Performance indicators in urban transit." *Management Science*, 28(4), 379-386.

Graham, D. J. (2006). "Productivity and efficiency in urban railways: Parametric and non-parametric estimates." *Transportation Research Part E: Logistics and Transportation Review*, In Press, Corrected Proof

Greene, W. H. (2003). *Econometric Analysis, 5th ed.,* Prentice Hall, Upper Saddle River, NJ.

Grosskopf, S. (1996). "Statistical inference and nonparametric efficiency: A selective survey." *Journal of Productivity Analysis*, 7(2-3), 161-176.

Karlaftis, M. G. (2003). "Investigating transit production and performance: A programming approach." *Transportation Research Part A-Policy and Practice*, 37(3), 225-240.

Karlaftis, M. G. (2004). "A DEA approach for evaluating the efficiency and effectiveness of urban transit systems." *European Journal of Operational Research*, 152(2), 354-364.

Kerstens, K. (1996). "Technical efficiency measurement and explanation of French urban transit companies." *Transportation Research Part A: Policy and Practice*, 30(6), 431-452.

Muniz, M., Paradi, J., Ruggiero, J., and Yang, Z. (2006). "Evaluating alternative DEA models used to control for non-discretionary inputs." *Computers & Operations Research*, 33(5), 1173-1183.

Nolan, J. F., Ritchie, P. C., and Rowcroft, J. E. (2002). "Identifying and measuring public policy goals: ISTEA and the US bus transit industry." *Journal of Economic Behavior & Organization*, 48(3), 291-304.

Nolan, J. F., Ritchie, P. C., and Rowcroft, J. R. (2001). "Measuring efficiency in the public sector using nonparametric frontier estimators: A study of transit agencies in the USA." *Applied Economics*, 33(7), 913-922.

Nolan, J. F. (1996). "Determinants of productive efficiency in urban transit." *Logistics and Transportation Review*, 32(3), 319-342.

Novaes, A. G. N. (2001). "Rapid transit efficiency analysis with the assurance-region DEA method." *Pesquisa Operacional*, 21(2), 179-197.

Nunamaker, T. R. (1985). "Using data envelopment analysis to measure the efficiency of non-profit organizations: A critical evaluation." *Managerial and Decision Economics*, 6(1), 50-58.

Odeck, J. and Alkadi, A. (2001). "Evaluating efficiency in the Norwegian bus industry using data envelopment analysis." *Transportation*, 28(3), 211-232.

Odeck, J. (2006). "Congestion, ownership, region of operation, and scale: Their impact on bus operator performance in Norway." *Socio-Economic Planning Sciences*, 40(1), 52-69.

Petersen, N. C. (1990). "Data envelopment analysis on a relaxed set of assumptions." *Management Science*, 36(3), 305-314.

Pina, V. and Torres, L. (2001). "Analysis of the efficiency of local government services delivery: An application to urban public transport." *Transportation Research Part A-Policy and Practice*, 35(10), 929-944.

Raudenbush, S. W. and Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods,* Sage Publications, Thousand Oaks, Calif.

Ruggiero, J. (2004a). "Performance evaluation when non-discretionary factors correlate with technical efficiency." *European Journal of Operational Research*, 159(1), 250-257.

Ruggiero, J. (2004b). "Performance evaluation in education: Modeling educational production." *Handbook on Data Envelopment Analysis*, W. W. Cooper, L. M. Seiford, and J. Zhu, eds., Kluwer, Boston, MA, 323-348.

Scheel, H. (2000). "EMS: Efficiency Measurement System." <http://www.wiso.uni-dortmund.de/lsfg/or/scheel/ems/> (Dec. 17, 2000).

Sheth, C., Triantis, K., and Teodorovic, D. (2006). "Performance evaluation of bus routes: A provider and passenger perspective." *Transportation Research Part E: Logistics and Transportation Review*, In Press, Corrected Proof

Simar, L. and Wilson, P. W. (2000). "Statistical inference in nonparametric frontier models: The state of the art." *Journal of Productivity Analysis*, 13(1), 49-78.

Simar, L. and Wilson, P. W. (2007). "Estimation and inference in two-stage, semi-parametric models of production processes." *Journal of Econometrics*, 136(1), 31-64.

Skrondal, A. and Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models,* Chapman & Hall/CRC, Boca Raton, FL.

Tandon, S. (2006). *Performance measurement of bus routes using data envelopment analysis,* Thesis, Master of Urban Planning and Policy, College of Urban Planning and Public Affairs, University of Illinois at Chicago, Chicago, IL.

Viton, P. A. (1997). "Technical efficiency in multi-mode bus transit: A production frontier analysis." *Transportation Research Part B: Methodological*, 31(1), 23-39.

Wooldridge, J. M. (2002). *Econometric analysis of cross section and panel data,* MIT Press, Cambridge, MA.

Zelenyuk, V. (2005). "Power of significance test of dummies in Simar-Wilson two-stage efficiency analysis model." <http://www.stat.ucl.ac.be/ISpub/ISdp.html#2005> (Apr. 12, 2006).

Zellner, A. (1962). "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias." *Journal of the American Statistical Association*, 57(298), 348-368.