Using Panel Data Analysis to Estimate Confidence Intervals for the DEA Efficiency of Individual Urban Paratransit Agencies

Darold T. Barnum*, University of Illinois at Chicago
John M. Gleason, Creighton University
Brendon Hemily, Brendon Hemily and Associates

* Darold T. Barnum is the corresponding author
Great Cities Institute
College of Urban Planning and Public Affairs
University of Illinois at Chicago

Great Cities Institute Publication Number: GCP-07-10

A Great Cities Institute Working Paper

December 2007

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 (dbarnum@uic.edu) 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.

John M. Gleason, Creighton University (jgleason@creighton.edu) is a Professor in the Department of Information Systems and Technology within in the College of Business Administration at Creighton University in Omaha.

Brendon Hemily, Brendon Hemily and Associates (brendon.hemily@sympatico.ca) is a Principal for Brendon Hemily and Associates located in Toronto, Canada.

Great Cities Institute Publication Number: GCP-07-10

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

Using Panel Data Analysis to Estimate Confidence Intervals for the DEA Efficiency of Individual Urban Paratransit Agencies

Abstract

This paper demonstrates a methodology using Panel Data Analysis to estimate confidence intervals for the Data Envelopment Analysis (DEA) efficiency of individual urban paratransit agencies and the statistical significance of trends in individual agency efficiency. The procedure accounts for stochastic variations of the inputs and outputs of the target agency as well as stochastic variations of the inputs and outputs of its efficient benchmark peers. The procedure is demonstrated using nine years of data from 34 urban paratransit agencies.

UIC Great Cities Institute

Introduction, Literature Review, and Paper Overview

This paper demonstrates the use of Panel Data Analysis (PDA) to estimate confidence intervals and the statistical significance of trends for the technical efficiency of individual Decision Making Units (DMUs), in this case urban paratransit agencies. We use DEA to estimate the technical efficiency (θ) of each DMU in each year, and then use PDA to estimate a confidence interval for each θ in each year and the statistical significance of trends in its value. PDA makes the standard parametric assumptions (residuals are independent and identically distributed (i.i.d.) and Normally distributed). Because DEA scores cannot be assumed to meet these conditions, we test for violations, and, where violations occur, adjust for them.

Developing statistical methodologies to deal with DEA scores has become an important theme in DEA research, as evidenced by several compilations including Grosskopf (1996), and Cooper, Seiford and Zhu (2004). As Chambers and Färe (2004, p. 329) recently noted, "more and more effort has been devoted to determining the statistical properties of the DEA approach."

Bootstrapping of cross-sectional data has been used to estimate the confidence intervals for the efficiencies of fixed input-output sets, with examples of the methodology and applications including Simar and Wilson (2000a; 2000b; 2007), Hollingsworth, Harris and Gospodarevskaya (2002), Gonzalez and Miles (2002), Latruffe, Balcombe, Davidova and Zawalinska (2004), and Zelenyuk and Zheka (2006). This bootstrapping research, however, treats the inputs and outputs of the target DMU as fixed, that is, the effects of stochastic variations in the target DMU's inputs and outputs are not taken into consideration (Simar and Wilson, 2000b, p. 67). Only production frontier variations are reflected in the bootstrapped confidence intervals around the target DMU's efficiency scores.

The need to consider the stochastic inputs and outputs of a target DMU has been recognized by others (Atkinson and Wilson, 1995; Löthgren, 1999; Löthgren and Tambour, 1999a; Löthgren and Tambour, 1999b; Löthgren, 2000). This need can be met only if there are multiple observations of a DMU's inputs and outputs, which, of course, is true for all stochastic variables. Therefore, to construct a confidence interval that reflects the stochastic variations of both the target DMU and the production frontier, it is necessary to make use of panel (longitudinal) data.

Several researchers have used panel data to mitigate the level of noise in DEA scores. Gong and Sickles (1992) reduce noise in target DMU scores by averaging each DMU's DEA scores across all years. Sengupta (1998b), Holland and Lee (2002), and Ruggiero (2004) mitigate noise in both the frontier and the target DMUs by averaging each DMU's inputs and outputs across all years and then doing one DEA on the averages. Sengupta (1998a) uses the longitudinal aspects of the data to filter out noise and then applies DEA to the filtered data. None of these papers develop statistical significance tests or confidence intervals within which the efficiency of each DMU is expected to fall, thereby not taking advantage of much of the value of panel data.

Atkinson and Wilson (1995) were the first to illustrate a procedure for estimating confidence intervals for individual DMU efficiency, using bootstrapping with panel data. They use bootstrapping because the asymptotic properties necessary for valid application of parametric methods should not be assumed for small DEA samples. However, bootstrapping cannot take advantage of the extensive body of knowledge and methodologies available for parametric models.

If a sample is not small, PDA can do so. PDA, sometimes referred to as Longitudinal Modeling or Time-series Cross-section Data Analysis, is well-known and is even the subject of many recent texts, such as Wooldridge (2002), Hsiao (2003), Frees (2004), Skrondal and Rabe-Hesketh (2004), Baltagi (2005) and Baum (2006). PDA has been used in many Stochastic

Frontier Analysis studies, and is examined at length in Kumbhakar and Lovell's (2000) widely-cited text.

However, when the response variables are DEA scores, PDA seldom has been used. In fact, the only publication that we identified was that of Steinmann and Zweifel's (2003) *Applied Economics* paper, which uses PDA to identify environmental influences on DEA scores. PDA never has been used with DEA scores to estimate individual DMU confidence intervals or the statistical significance of trends, or to validate i.i.d. and Normality requirements.

This paper progresses as follows. The inputs and outputs are identified and justified. The DEA linear programming model is presented, followed by the PDA statistical models. The results of comprehensive diagnostic tests on the regression residuals are presented. Finally, for each DMU, the trends and confidence intervals that result from the procedure are examined.

DEA Inputs and Outputs

We use data from 34 Canadian urban paratransit agencies for the nine years 1996-2004 (Canadian Urban Transit Association, 2005). Paratransit generally is defined as public transportation that is demand-responsive rather than operating on a fixed schedule, and usually is door-to-door. Paratransit vehicles typically include cars, vans and small buses. In the case of the Canadian agencies, paratransit service is limited to disabled individuals and their caregivers.

Inputs are (1) number of disabled individuals and caregivers registered for the service, (2) annual operating expenses of paratransit service dedicated to disabled individuals, and (3) annual operating expenses of non-dedicated paratransit service attributable to disabled riders. Outputs are (1) annual number of passenger trips on paratransit service dedicated to disabled individuals, (2) annual number of passenger trips attributable to disabled passengers on non-dedicated paratransit service, (3) annual operating revenue from dedicated paratransit service, and (4) annual operating revenue from non-dedicated paratransit service attributable to disabled passengers.

Two key outputs of Canadian urban paratransit agencies are disabled-passenger trips and operating revenue (mostly from fares). Fare levels vary significantly across systems and years; consequently, the passenger trips and revenue values measure two unique outputs. Agencies obtain different mixes of these two outputs, because increases in one have always led to decreases in the other when fares are changed, holding inputs constant (Litman, 2004).

Both outputs are produced by two distinct organizational subunits: a subunit serving only disabled users (dedicated service), and a subunit serving both disabled and non-disabled users (non-dedicated service). Dedicated service is provided in vehicles exclusively dedicated to the transport of persons with disabilities, which may be operated by the agency itself or subcontracted. Non-dedicated service is provided in vehicles that serve both disabled and non-disabled customers, often taxicabs subcontracted by the paratransit agency to transport its disabled clients.

Operating expenses are used as a proxy for physical inputs. Universally, in paratransit operations, a very high percentage of operating expenses is for employee compensation. For example, for U.S. paratransit in 2004, 82 percent of operating expenses was for employee compensation, followed by 6 percent for fuel, 6 percent for insurance, 5 percent for materials and supplies and 1 percent for utilities (Danchenko, 2006). And, average compensation varies substantially among the 34 Canadian agencies, with the 2004 operator top base wage rate ranging from $14.82 to $24.32 per hour. Typically, most other wage rates in the dedicated service subunit are indexed on the top operator base wage rate, which also can serve as a proxy for wage rate variation in the non-dedicated subunit because it reflects local labor market conditions. It is likely that other operating expenses also will show somewhat similar variation due to local cost differences. Therefore, we have divided the operating expense variables by each agency's top operator base wage rate, which should yield a reasonable proxy for physical inputs.

For this analysis, all inputs and outputs are limited to those related to disabled passengers. However, because there are two quite unique methods of providing transportation (dedicated or non-dedicated service), and two different organizational subunits with independent production processes, inputs and outputs from each are entered as separate variables.

The reason that this disaggregation is necessary is to avoid aggregation bias in the reported DEA scores. The four potential sources of aggregation bias are: inter-input aggregation (sum of the quantities of different types of inputs weighted by prices), inter-output aggregation (sum of the quantities of different types of output weighted by prices), intra-input aggregation (sum of the quantities of a given type of input used by multiple subunits or production processes), and intra-output aggregation (sum of the quantities of a given type of output produced by multiple subunits or production processes). Recent research demonstrates that, for inputs and outputs that are allocable, disaggregated data must be used to avoid bias in technical efficiency scores (Tauer, 2001; Färe and Zelenyuk, 2002; Färe, Grosskopf and Zelenyuk, 2004; Färe and Grosskopf, 2004; Barnum and Gleason, 2005; Barnum and Gleason, 2006a; Barnum and Gleason, 2006b; Barnum and Gleason, 2007b).

The first input, "number of disabled individuals and caregivers registered for the service," is not allocable. Registrants requesting transportation are assigned to a dedicated or non-dedicated vehicle after a request is received, with consideration given to the person's disability, vehicle availability at the desired time, and relative cost.

However, the other two inputs are allocable. If operating expenses were to be aggregated across the two subunits, then intra-input aggregation would be present. Intra-output aggregation would be present if disabled passenger trips were to be aggregated across the two subunits, and if operating revenue were to be aggregated across the two subunits. Such aggregation can cause substantial levels of downward bias in an organization's technical efficiency scores, with the amount of bias varying greatly depending on the degree of allocation efficiency (Barnum and Gleason, 2006a; Barnum and Gleason, 2007b).

There is no inter-output aggregation, but using operating expenses as aggregations of physical inputs does subject the analysis to potential inter-input aggregation bias. However, because there is very little substitutability among the various inputs, such as operators, mechanics, fuel, and vehicles, any inter-input aggregation bias should be very minor at most.

In conclusion, the three inputs and four outputs account for variables considered key by this industry, as reflected in the set of published performance indicators that are used to compare agencies (Canadian Urban Transit Association, 2005).

The DEA Model

For each of the nine years of data, DEA scores are computed using a linear programming model (1) that is output oriented and assumes constant returns to scale. The DEAs were conducted with Scheel's EMS software (2000). For each of the *j* DMUs (*j* = 1,…,34), there are data on the *n* = 3 inputs *x* $(x_{11},...,x_{34,3})$, and on the *m* = 4 outputs *y* $(y_{11},...,y_{34,4})$. The DEA score *θ* identifies the technical super-efficiency of the target DMU *k* (Andersen and Petersen, 1993). Because our model is output-oriented, efficient DMUs' scores will be in the interval $(0,1]$, with smaller values indicating increasing efficiency; and, inefficient DMUs' scores will be in the interval $(1,\infty)$, with larger values indicating decreasing efficiency.

$$
\begin{aligned}
&\max_{\lambda} \theta \\
&\text{subject to} \quad \sum_{j=1}^{34} y_{jm}\lambda_j \geq \theta y_{km} && m = 1,2,3,4 \\
&\sum_{j=1}^{34} x_{jn}\lambda_j \leq x_{kn} && n = 1,2,3 \\
&\lambda_k = 0 \\
&\lambda_j \geq 0 && j = 1,2,\ldots,34;\ j \neq k
\end{aligned}
\tag{1}
$$

The conventional measure of technical efficiency, with the range $[1,\infty)$, would have been appropriate if our interest had focused on identifying the production frontier, those DMUs defining it, and the distance of inefficient DMUs from it. In that model, efficient DMUs are compared only to themselves. Thus, if an efficient DMU increases its outputs or decreases its inputs, everything else equal, its DEA score of one would remain unchanged in spite of its improved performance. Using super-efficiency, changes in the performance of both efficient and inefficient DMUs will be reflected in their DEA scores. Also, we are interested in always benchmarking the performance of the target DMU on its efficient peers. Super-efficiency scores always benchmark the target DMU on its efficient peers regardless of its own efficiency level.

A second reason for using super-efficiency scores is to avoid a limited-value response variable in second-stage regressions, as discussed by Coelli *et al* (2005). Conventional technical efficiency scores yield a limited-value variable, because an efficient DMU's score of 1 will remain unchanged even if it were to become more efficient by increasing outputs or decreasing inputs. Super-efficiency scores are an observable proxy for latent variable values underlying conventional efficiency scores. Because their range is not limited, this precludes the need to estimate the latent values using sample-selected, truncated or censored regression (as addressed by Breen (1996) in his discussion of limited-value variable methods).

Statistical Models

The initial model was

$$\theta_{jt} = \alpha_j + \beta_j(t-1) + u_{jt} \qquad j = 1,\ldots,34;\ t = 1,\ldots,9 \qquad (2)$$

where θ_{jt} is the super-efficiency score of DMU *j* in year *t*, α_j is the estimated efficiency of DMU *j* in year 1, β_j is the annual change in the efficiency of DMU *j*, and u_{jt} is the random error in the super-efficiency score (θ_{jt}) of DMU *j* in year *t*. The DMUs showed differing trends in efficiency over the nine year period, as reflected by differing estimated values for their β coefficients. Therefore Equation 2 includes a variable adjusting efficiency for the year involved, which permits a heterogeneous trend in each DMU's efficiency over the nine-year period. The fitted value of the efficiency of DMU *j* is α_j in year 1, $\alpha_j + \beta_j$ in year 2, and so on up to $\alpha_j + 8\beta_j$ in year 9. Statistical computations were conducted with Stata 9.2 (StataCorp, 2007).

Residual Diagnostics

In order to use parametric analysis for developing valid confidence intervals from a correctly specified regression model for panel data, the errors (u_{jt}) must be i.i.d. and Normally distributed, or, if they are not, the violations must be addressed in the statistical model.

It never should be assumed that these requirements are met when the dependent variable is a DEA score, because each DMU's score is influenced by the performance of other DMUs. If the same DMUs consistently influence each other, it may cause correlations among their error terms. Such contemporaneous correlation would invalidate the requirement for independent residuals. There may well be other complex and unknown effects on the residuals caused by the interdependence of DEA scores, so the presence of i.i.d. and

Normality should always be confirmed, and model corrections made when they are not present.

Homoskedasticity

For panel data, residual variance should be tested for (1) homoskedasticity across panel members, and (2) relationships between variable size and residual variance. Although homoskedasticity across panel members would not be necessary if we were diagnosing each set of nine residuals individually (which cannot be done because nine are too few observations), it is necessary when diagnosing all 306 residuals at the same time. That is, unless the variances are homoskedastic, the remaining diagnostic tests for i.i.d. and Normality cannot be validly performed.

Equation 2 was solved using data from the 34 paratransit DMUs, and the variances of the nine residuals of each DMU were computed. Subjecting these variances to the Breusch-Pagan/Cook-Weisberg test for heteroskedasticity among the DMUs showed that the heteroskedasticity is statistically significant [χ^2 = 191.25, P(χ^2 (33) > 191.25) < 0.00005].

Because the estimated residuals of each DMU are known, we divide each dependent variable by the standard deviation of its DMU's nine residuals. Thus, $w_j = 1/\hat{\sigma}_{u_j}$, where w_j is the weight assigned to the scores of DMU *j*, and $\hat{\sigma}_{u_j}$ is the standard error of the regression equation residuals for DMU *j*. Replacing Equation (2) with (3):

$$w_j\theta_{jt} = \alpha_j + \beta_j(t-1) + u_{jt} \qquad j = 1,\ldots,34;\ t = 1,\ldots,9 \qquad (3)$$

Subjecting the residuals from Equation 3 to the Breusch-Pagan/Cook-Weisberg test for heteroscedasticity among the DMUs yields no statistically significant difference [χ^2 < 0.005, P(χ^2 (33) > 0.005) = 1.0000]. The Breusch-Pagan/Cook-Weisberg test for heteroskedasticity of error terms yields no statistically significant relationships between variance of the residuals and (1) fitted scores, u_{jt} on E$(w_j\theta_{jt})$, [χ^2 < 0.005, P($\chi^2(1)$ > 0.005) = 0.9888], (2) time, u_{jt} on t, [χ^2 = 0.93, P($\chi^2(1)$ > 0.93) = 0.3357)], and (3) weighted scores, u_{jt} on $w_j\theta_{jt}$, [χ^2 = 0.02, P($\chi^2(1)$ > 0.02) = 0.8825]. In sum, homoskedastic variance is present when the weighted dependent variable is utilized. Therefore, the rest of this paper is based on the model in Equation 3.

Independence of Error Terms

There are two issues of concern. The first issue is the potential for correlation of errors over time, that is, serial correlation. The second issue is the potential for cross-sectional correlation of the error terms of the DMUs, also called contemporaneous correlation. To test for first-order serial correlation, we regressed each residual on the prior year's residual, eliminating the first year from the regression since no prior residual was available (Equation 4). We followed the advice of Wooldridge (2002) and ran a pooled OLS regression with a fully robust standard error of

$$\hat{u}_{jt} \text{ on } \hat{u}_{j,t-1} \qquad t = 2,\ldots,9;\ j = 1,\ldots,34 \qquad (4)$$

There was no significant relationship [R^2 = 0.0024, F = 0.72, P(F (1, 270) > 0.72) = 0.3966].

The second potential influence on the independence of error terms involves the effect of interdependence between DEA scores of two or more DMUs as described by Xue and Harker (1999) and Simar and Wilson (2007). Tests for contemporaneous correlation cannot be conducted if the data are treated as separate cross-sections, or if each DMU's data are treated as a separate time-series. Such tests are available if the data are treated as a panel and analyzed with PDA.

Contemporaneous correlation does *not* bias the expected value of estimated efficiency levels, but variance estimates can be more precise if the correlation is taken into account. That is, if the residuals are contemporaneously correlated to a statistically significant degree, one could use such methods as Generalized Least Squares to decrease the estimated standard errors.

If the number of DMUs exceeds the number of time periods, as is true for our data, then tests for cross-sectional independence of residuals include Freedman's R_{AVE} and Frees' R_{AVE}^2 evaluated with his Q-distribution (Frees, 1995; Frees, 2004), Pesaran's CD cross-sectional dependence test (Baum, 2006, p. 222), and a test for pairwise correlation of residuals. The results are CD = 0.915, P(CD > |0.915|) = 0.3602; and R_{AVE} = 11.969, P(R_{AVE} > 11.969) = 0.9997. None of the pair-wise correlations were statistically significant at the 0.05 level, for either Bonferroni or Sidak-adjusted significance tests. However, Frees' R_{AVE}^2 = 1.127 is statistically significant at the 0.05 level.

Because only three of the four tests reported no statistically significant contemporaneous correlation, the evidence is not unanimous. The conservative decision, which we adopt, is to assume that there is no true cross-correlation. This is conservative because models accounting for cross-correlation will estimate narrower variances than i.i.d. variances, whether or not cross-correlation is truly present.

Normality of Residuals

The hypothesis that the residuals are Normally distributed could not be rejected according to the Shapiro-Wilk W [W = 0.994, V = 1.259, z = 0.538, P(z > 0.5380) = 0.295], the Shapiro-Francia W' [W' = 0.006, V' = 1.004, z = 0.008, P(z > 0.008) = 0.497], and a skewness/kurtosis test [P(skewness) = 0.597, P(kurtosis) = 0.055, joint adj. χ^2 (2) = 3.98, P(χ^2 > 3.98) = 0.136].

Summary

The hypotheses that the residuals from Equation 3 are i.i.d. and Normally distributed cannot be rejected. This demonstrates that violations of the standard parametric assumptions by DEA residuals are not inevitable. It may be worthwhile to recall that <u>any</u> disturbance term encapsulates complicated, unidentified influences, so it is never truly random. But, a residual can be treated as random if it meets appropriate statistical tests for randomness (Neter and Wasserman, 1974; Frees, 2004). So, we make the standard asymptotic assumptions in developing confidence intervals.

Results

Because of the weighting of the dependent variable, the individual DMU parameter estimations are not of interest in their own right, and are only needed in order to predict the range within which a DMU's efficiency will occur. It is worth noting that 97 percent of the variance in efficiency scores is explained by differences between the DMUs [R^2 = 0.9675, F = 105.64, P(F(67, 238) >105.64) < 0.00005].

Because many DMUs exhibit linear trends in efficiency over the nine year period, it is necessary to include a variable adjusting efficiency for the year involved, as is done in equation 3. If the yearly trend variables are not included, R-square decreases from 0.97 to 0.85, with the difference between the full and reduced models being statistically significant at the 0.05 level. Thus, the expected efficiency of a DMU depends on the year of interest in many of the cases.

Table 1 reports the estimates for 2004. The range is based on the standard error of prediction of the mean value, at the 0.90 level of confidence. The point estimates for 17 DMUs showed them to be inefficient, but 4 of these were not inefficient to a statistically significant degree. Of the 17 DMUs with efficient point estimates, 7 were not efficient to a statistically significant degree. Therefore, at least for Canadian paratransit properties, the point estimates of efficiency for a given year should not be used to identify their levels of efficiency, because the range within which their efficiency could occur is often going to overlap the inefficient and efficient ranges. Thus, whether 11 of the 34 systems were or were not efficient cannot be reported with statistical confidence.

Whether trends in efficiency are present to a statistically significant degree also is likely to be of interest. These are shown in the last column of Table 1. Because the DEA model is output oriented, higher scores mean lower efficiency. Therefore, a negative coefficient means efficiency is increasing, and a positive coefficient that efficiency is decreasing. For the DMUs, 21 report increasing efficiency over the nine years involved, 15 of which show statistically significant improvements. And, 13 report decreasing efficiency over the nine year period, 8 of which are worse to a statistically significant degree.

Conclusions

As exhibited in this paper, Panel Data Analysis offers a useful methodology for estimating confidence intervals and the statistical significance of trends for the DEA efficiency of individual DMUs. It addresses stochastic variations in the data of the target DMU and the production frontier. It permits testing of residuals for conformity to the standard parametric assumptions, and provides methods for correcting violations. Although not addressed herein, PDA provides more valid and powerful procedures than cross-sectional analysis for estimating the effects of the environment on efficiency. However, as Barnum and Gleason (2007a) have demonstrated, using DEA scores as the response variable in such regressions can result in inefficient and inconsistent estimates.

Table 1. 2004 Superefficiency Scores, 34 Canadian Paratransit Properties

DMU_j	2004 Superefficiency Score (0.90 Confidence Interval)				Mean Annual Change
	$E(\theta_{j9})$	Lower limit	Upper limit	Conclusion	$(\hat{\beta}_j)$
1	0.67	0.59	0.76	Efficient *	-0.138*
2	1.25	1.12	1.37	Inefficient *	0.031
3	1.57	1.40	1.73	Inefficient *	0.089*
4	0.72	0.59	0.84	Efficient *	0.078*
5	0.85	0.46	1.24	Efficient	-0.111*
6	0.63	0.51	0.76	Efficient *	-0.068*
7	0.80	0.64	0.96	Efficient *	-0.020
8	1.20	1.04	1.36	Inefficient *	0.047*
9	0.95	0.86	1.04	Efficient	0.013
10	0.74	0.51	0.97	Efficient *	-0.040*
11	1.19	0.95	1.42	Inefficient	-0.030*
12	0.63	0.46	0.80	Efficient *	-0.031*
13	1.00	0.90	1.11	Efficient	-0.039
14	0.98	0.92	1.03	Efficient	-0.103*
15	1.18	0.99	1.37	Inefficient	-0.025*
16	1.02	0.96	1.09	Inefficient	-0.086*
17	0.90	0.62	1.19	Efficient	-0.010
18	1.48	1.32	1.64	Inefficient *	0.044*
19	1.41	1.25	1.57	Inefficient *	0.044*
20	0.69	0.60	0.79	Efficient *	0.036
21	0.85	0.79	0.91	Efficient *	-0.002
22	1.15	1.05	1.25	Inefficient *	-0.081*
23	1.32	1.21	1.43	Inefficient *	0.066*
24	0.87	0.80	0.95	Efficient *	-0.045*
25	1.11	1.01	1.21	Inefficient *	-0.161*
26	0.48	0.32	0.64	Efficient *	-0.065*
27	2.09	1.91	2.27	Inefficient *	0.018
28	0.87	0.68	1.05	Efficient	-0.035
29	1.43	1.33	1.53	Inefficient *	-0.051*
30	1.68	1.58	1.77	Inefficient *	0.213*
31	1.23	1.17	1.29	Inefficient *	-0.048
32	0.94	0.72	1.15	Efficient	0.016
33	1.01	0.93	1.09	Inefficient	-0.134*
34	1.25	1.10	1.41	Inefficient *	0.053*

* Statistically significant at the 0.9 two-tailed level.

References

Andersen, P. and Petersen, N. C. (1993) A procedure for ranking efficient units in data envelopment analysis, *Management Science, 39*(10), 1261-1265.

Atkinson, S. E. and Wilson, P. W. (1995) Comparing mean efficiency and productivity scores from small samples - a bootstrap methodology, *Journal of Productivity Analysis, 6*(2), 137-152.

Baltagi, B. H. (2005) *Econometric analysis of panel data*, West Sussex, England, John Wiley & Sons, Ltd.

Barnum, D. T. and Gleason, J. M. (2005) Technical efficiency bias caused by intra-input aggregation in data envelopment analysis, *Applied Economics Letters, 12* (13), 785-788.

Barnum, D. T. and Gleason, J. M. (2006a) 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. and Gleason, J. M. (2006b) Measuring efficiency in allocating inputs among outputs with DEA, *Applied Economics Letters, 13* (6), 333-336.

Barnum, D. T. and Gleason, J. M. (2007a) Bias and precision in the DEA two-stage method. Forthcoming in *Applied Economics*.

Barnum, D. T. and Gleason, J. M. (2007b) Technical efficiency bias in data envelopment analysis caused by intra-output aggregation. Forthcoming in *Applied Economics Letters*.

Baum, C. F. (2006) *An introduction to modern econometrics using Stata*, College Station, Texas, Stata Press.

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

Canadian Urban Transit Association (2005) *Specialized transit services fact books 1996-2004 operating data*, Toronto, Canada, Canadian Urban Transit Association.

Chambers, R. G. and Fare, R. (2004) Additive decomposition of profit efficiency, *Economics Letters, 84* (3), 329-334.

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

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

Danchenko, D. (2006) *Public Transportation Fact Book*, http://www.apta.com/research/stats/factbook/documents/2006factbook.pdf.

Färe, R., Grosskopf, S. and Zelenyuk, V. (2004) Aggregation bias and its bounds in measuring technical efficiency, *Applied Economics Letters, 11* (10), 657-660.

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

Färe, R. and Zelenyuk, V. (2002) Input aggregation and technical efficiency, *Applied Economics Letters*, 9 (10), 635-636.

Frees, E. W. (1995) Assessing cross-sectional correlation in panel data, *Journal of Econometrics, 69* (2), 393-414.

Frees, E. W. (2004) *Longitudinal and panel data: Analysis and applications in the social sciences*, Cambridge, U.K., Cambridge University Press.

Gong, B. H. and Sickles, R. C. (1992) Finite-sample evidence on the performance of stochastic frontiers and data envelopment analysis using panel data, *Journal of Econometrics, 51* (1-2), 259-284.

Gonzalez, X. M. and Miles, D. (2002) Statistical precision of DEA: A bootstrap application to Spanish public services, *Applied Economics Letters, 9* (2), 127-132.

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

Holland, D. S. and Lee, S. T. (2002) Impacts of random noise and specification on estimates of capacity derived from data envelopment analysis, *European Journal of Operational Research, 137* (1), 10-21.

Hollingsworth, B., Harris, A. and Gospodarevskaya, E. (2002) The efficiency of immunization of infants by local government, *Applied Economics, 34* (18), 2341-2345.

Hsiao, C. (2003) *Analysis of panel data*, Cambridge, Cambridge University Press.

Kumbhakar, S. and Lovell, C. A. K. (2000) *Stochastic frontier analysis*, Cambridge England, Cambridge University Press.

Latruffe, L., Balcombe, K., Davidova, S. and Zawalinska, K. (2004) Determinants of technical efficiency of crop and livestock farms in Poland, *Applied Economics, 36* (12), 1255-1263.

Litman, T. (2004) Transit price elasticities and cross-elasticities, *Journal of Public Transportation, 7* (2), 37-58.

Löthgren, M. (1999) Bootstrapping the Malmquist productivity index - a simulation study, *Applied Economics Letters, 6* (11), 707-710.

Löthgren, M. (2000) On the consistency of the DEA-based average efficiency bootstrap, *Applied Economics Letters, 7* (1), 53-57.

Löthgren, M. and Tambour, M. (1999a) Bootstrapping the data envelopment analysis Malmquist productivity index, *Applied Economics, 31* (4), 417-425.

Löthgren, M. and Tambour, M. (1999b) Testing scale efficiency in DEA models: a bootstrapping approach, *Applied Economics, 31* (10), 1231-1237.

Neter, J. and Wasserman, W. (1974) *Applied linear statistical models: Regression, analysis of variance, and experimental designs*, Homewood, IL, Irwin.

Ruggiero, J. (2004) Data envelopment analysis with stochastic data, *Journal of the Operational Research Society, 55* (9), 1008-1012.

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

Sengupta, J. K. (1998a) Stochastic data envelopment analysis: A new approach, *Applied Economics Letters, 5* (5), 287-290.

Sengupta, J. K. (1998b) Testing allocative efficiency by data envelopment analysis, Applied *Economics Letters, 5* (11), 689-692.

Simar, L. and Wilson, P. W. (2000a) A general methodology for bootstrapping in non-parametric frontier models, *Journal of Applied Statistics, 27* (6), 779-802.

Simar, L. and Wilson, P. W. (2000b) 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*, Boca Raton, FL, Chapman & Hall/CRC.

StataCorp, *Stata Statistical Software*, Release 9.2, 2007.

Steinmann, L. and Zweifel, P. (2003) On the (in)efficiency of Swiss hospitals, *Applied Economics, 35* (3), 361-370.

Tauer, L. W. (2001) Input aggregation and computed technical efficiency, *Applied Economics Letters, 8* (5), 295-297.

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

Xue, M. and Harker, P. T. (1999) *Overcoming the inherent dependency of DEA efficiency scores: A bootstrap approach*, 99-17, Financial Institutions Center, Wharton School, Univ. of Pennsylvania, Philadelphia, PA, http://fic.wharton.upenn.edu/fic/papers/99/9917.pdf.

Zelenyuk, V. and Zheka, V. (2006) Corporate governance and firm's efficiency: The case of a transitional country, Ukraine, *Journal of Productivity Analysis, 25* (1), 143-157.