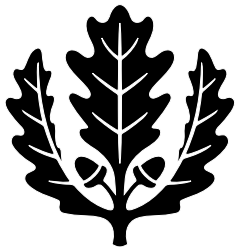

University of Connecticut

Department of Economics Working Paper Series

A Bootstrap-Regression Procedure to Capture Unit Specific Effects in Data Envelopment Analysis

Evangelia Desli
Lloyd's of London

Subhash Ray
University of Connecticut

Working Paper 2004-15

July 2004

341 Mansfield Road, Unit 1063
Storrs, CT 06269–1063
Phone: (860) 486–3022
Fax: (860) 486–4463
http://www.econ.uconn.edu/

Abstract

The Data Envelopment Analysis (DEA) efficiency score obtained for an individual firm is a point estimate without any confidence interval around it. In recent years, researchers have resorted to bootstrapping in order to generate empirical distributions of efficiency scores. This procedure assumes that all firms have the same probability of getting an efficiency score from any specified interval within the [0,1] range. We propose a bootstrap procedure that empirically generates the conditional distribution of efficiency for each individual firm given systematic factors that influence its efficiency. Instead of resampling directly from the pooled DEA scores, we first regress these scores on a set of explanatory variables not included at the DEA stage and bootstrap the residuals from this regression. These pseudo-efficiency scores incorporate the systematic effects of unit-specific factors along with the contribution of the randomly drawn residual. Data from the U.S. airline industry are utilized in an empirical application.

Journal of Economic Literature Classification: C15, C63

Keywords: DEA; Kernel Smoothing; Reflection Method; Empirical Density

A BOOTSTRAP-REGRESSION PROCEDURE TO CAPTURE UNIT SPECIFIC EFFECTS IN DATA ENVELOPMENT ANALYSIS

1. Introduction

One major drawback of Data Envelopment Analysis (DEA) is that it is non-statistical and the efficiency score obtained for an individual firm is a point estimate without any confidence interval around it. In recent years, researchers have resorted to bootstrapping (e.g. Simar (1992, 1996), Simar and Wilson (1998, 2000) among others) in order to generate empirical distributions of efficiency scores from repeated applications of DEA after resampling. The essential procedure is to pool the efficiency measures obtained from the actual data and then randomly sample with replacement from this pool to construct pseudo-data on outputs (or inputs) for the firms. These artificial data on outputs (inputs) are associated with actual input (output) data for another round of DEA. Repeating this procedure a large number of times generates large enough samples of efficiency scores for each firm. Then one can look at the mean and the variance of each of the empirical distributions of efficiency.

While this procedure is quite appealing and is gaining wide acceptance, in a sense, it goes to the other extreme by assuming that all firms have the same probability of getting an efficiency score from any specified interval within the [0,1] range. This reduces efficiency to a purely random variable and there would be little point in talking of the efficiency of one firm relative to the others. In reality, however, some firms are more likely to be rated at a higher efficiency level than other firms. There usually are systematic factors that contribute to differences in efficiency. The existing bootstrapping procedures do not consider the possibility that the distributions of efficiency conditional on unit specific factors may differ across firms. One can argue in favor of including these factors within the scope of the DEA model itself so that the remaining variation in efficiency can be justifiably attributed to purely random factors. However, inclusion of these factors as non-discretionary inputs within the DEA model automatically extends the disposability

property (weak or strong) to such variables. This is not a realistic assumption in many situations. This is one reason why researchers often regress DEA efficiency scores on a number of explanatory variables to adjust for environmental factors and they do not include these factors in the DEA model itself (e.g. Ray (1991), McCarthy and Yaisawarng (1993)).

In this paper we propose an enhanced bootstrap procedure that empirically generates the conditional distribution of efficiency for each individual firm given the systematic factors that influence their efficiency. This new procedure can be characterized as a second stage regression DEA bootstrap. The principal innovation in this study is that instead of resampling directly from the pooled DEA scores, we first regress these scores on a set of explanatory variables not included at the DEA stage and subsequently bootstrap the residuals from this regression. These pseudo-efficiency scores incorporate the systematic effects of unit-specific factors along with the contribution of the randomly drawn residual.

This paper is organized as follows. In section 2 we set up the DEA model, describe the concepts of the bootstrap procedure and how it is currently applied to the DEA model as an one-step bootstrap. Section 3 describes and the regression of the technical efficiency on the unit-specific factors, develops the second stage regression DEA bootstrap procedure and differentiates it from the one-step bootstrap. Section 4 reports the findings from an empirical application using data from the U.S. airline industry. Finally, the last section summarizes.

2. Measurement of Efficiency

In parametric models, one specifies an explicit functional form for the frontier and econometrically estimates the parameters using sample data for inputs and output. Hence the validity of the derived technical efficiency measures depends critically on the appropriateness of the functional form specified.

2.1 Data Envelopment Analysis

The method of DEA introduced by Charnes, Cooper and Rhodes (CCR) (1978) and further generalized by Banker, Charnes, and Cooper (BCC) (1984) provides a nonparametric alternative to parametric frontier production function analysis. In DEA, one makes only a few fairly weak assumptions about the underlying production technology. In particular, no functional specification is necessary. Based on these assumptions a production frontier is empirically constructed using mathematical programming methods from observed input-output data of sample firms. Efficiency of firms is then measured in terms of how far they are from the frontier.

Consider an industry producing a bundle of m outputs, $y=(y_1,y_2,...,y_m)$, from bundles of k inputs, $x=(x_1,x_2,...,x_k)$. Let (x^j, y^j) be the observed input-output bundle of firm j $(j= 1,2,..., n)$. The technology is defined by the production possibility set

$$T=\{(x, y): y \text{ can be produced from } x\}.$$

An input-output combination (x^0, y^0) is feasible if and only if $(x^0, y^0) \in T$. We make the following assumptions about the technology:

- All observed input-output combinations are feasible. Thus, $(x^j, y^j) \in T\ (j = 1,2,...,n)$.
- The production possibility set, T, is convex. Hence, if $(x^1, y^1) \in T$ and $(x^2, y^2) \in T$, then $(\lambda x^1+(1-\lambda)x^2, \lambda y^1+(1-\lambda)y^2) \in T,\ 0 \le \lambda \le 1$.

In other words, weighted averages of feasible input-output combinations are also feasible.

- Inputs are freely disposable. Hence, if $(x^0, y^0) \in T$ and $x^1 \ge x^0$, then $(x^1, y^0) \in T$. This rules out negative marginal productivity of inputs.
- Output is freely disposable. Hence, if $(x^0, y^0) \in T$ and $y^1 \le y^0$, then $(x^0, y^1) \in T$.

Varian (1984) pointed out that the smallest set satisfying the above assumptions is:

$$S = \{(x, y) : x \ge \sum_{j=1}^{n} \lambda_j x^j;\ y \le \sum_{j=1}^{n} \lambda_j y^j;\ \sum_{j=1}^{n} \lambda_j = 1;\ \lambda_j \ge 0;\ j = 1,2,...,n\}.$$

Let $\bar{x} = \sum_{j=1}^{n} \lambda_j x^j, \bar{y} = \sum_{j=1}^{n} \lambda_j y^j; \sum_{j=1}^{n} \lambda_j = 1; \lambda_j \ge 0.$ By virtue of convexity, $\{\bar{x}, \bar{y})$ is feasible. Thus, for any $x \ge \bar{x}, (x, \bar{y})$ is feasible. Finally, for any $y \le \bar{y}, (x, y)$ is also feasible.

Under the assumptions listed above, the technical efficiency of any firm producing output y^0 from input x^0 is $1/\varphi^*$, where

$$\varphi^* = \max\ \varphi : (x^0, \varphi y^0) \in S.$$

Consider an industry producing a scalar output y from a vector of k inputs, $x=(x_1, x_2, \ldots, x_k)$. Suppose that the input-output data are observed for n firms. Let the vectors x^i be the input bundle and y_i the output level of the i-th firm. The output-oriented technical efficiency of the j-th firm under variable returns to scale (VRS), also known as the BCC model, can be computed by solving the linear programming (LP) problem:

$$\begin{aligned} & \max \phi_j \\ s.t. \quad & \sum_{i=1}^{n} \lambda_i y_i \geq \phi_j y_j ; \\ & \sum_{i=1}^{n} \lambda_i x^i \leq x_s^j ; \ \text{for s} = 1,2,\ldots,\text{k} ; \\ & \sum_{i=1}^{n} \lambda_i = 1; \\ & \lambda_i \geq 0 \quad \text{for i} = 1,2,\ldots,\text{n}. \end{aligned} \tag{1}$$

The technical efficiency for the j-th firm is the inverse of ϕ_j.

$$\text{TE}_\text{j} = \frac{1}{\phi_\text{j}} \tag{2}$$

When ϕ_j is equal to 1, the technical efficiency is equal to 1, i.e. the firm is 100% efficient. If ϕ_j is greater than 1, the firm is technically inefficient and the efficiency measure is less than 1.

Note that DEA models lead to specific measures of technical efficiency that are point estimates and therefore lack statistical properties. This problem has been addressed with the use of bootstrap methods.

2.2 Bootstrap

The idea of the bootstrap was first introduced by Efron (1979), who proposed the use of computer-based simulations to obtain the sampling properties of random variables. The starting point of any bootstrap procedure is a sample of observed data $X=\{x_1, x_2, \ldots, x_n\}$ drawn randomly

from some population with an unknown probability distribution *f*. The basic assumption behind the bootstrap method is that the random sample actually drawn "mimics" its parent population.

Suppose that a sample of observed data $X=\{x_1, x_2,...,x_n\}$ is drawn randomly from some population with an unknown probability distribution *f*. The sample statistic $\hat{\theta} = \theta(X)$ computed from this state of observed values is merely an estimate of the corresponding population parameter $\theta = \theta(f)$. When it is not possible to analytically derive the sampling distribution of that statistic, one examines its empirical density function. Unfortunately, however, the researcher has access to only one sample rather than multiple samples drawn from the same population. As noted above the basic assumption behind the bootstrap method is that the random sample actually drawn "mimics" its parent population. Therefore, if one draws a random sample with replacement from the observed values in the original sample, it can be treated like a sample drawn from the underlying population itself. Repeated samples with replacement yield different values of the sample statistic under investigation and the associated empirical distribution (over these samples) can provide the sampling distribution of this statistic. For reasons explained later this is known as a naïve bootstrap.

The bootstrap sample $X^*=\{x_1^*, x_2^*,...,x_n^*\}$ is an unordered collection of *n* items drawn randomly from the original sample *X* with replacement, so that any x_i^* *(i=1,2,...,n)* has *1/n* probability of being equal to any x_j *(j=1,2,...,n)*. Some observations from the original sample *X* will not appear in the bootstrap sample at all, while others will appear more than once. Let $\hat{f}$ denote the empirical density function of the observed sample *X* from which *X** was drawn. Then it can take the form:

$$\hat{f}(t) = \begin{cases} 1/n & \text{if } t = x_i^*, i = 1,2,...,n \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

If $\hat{f}$ is a consistent estimator of *f*, then the bootstrap distributions will mimic the original unknown sampling distributions of the estimators that we are interested in. Let $\hat{\theta}^* = \theta(X^*)$ be the estimated parameter from the bootstrap sample *X**. Then the distribution of $\hat{\theta}^*$ around $\hat{\theta}$ in $\hat{f}$ is the same as of $\hat{\theta}$ around θ in *f*. That is:

$$(\hat{\theta}^* - \hat{\theta}) \mid \hat{f} \sim (\hat{\theta} - \theta) \mid f \,. \tag{4}$$

Since every time we replicate the bootstrap sample we get a different sample *X**, we will also get a different estimate of $\hat{\theta}^* = \theta(X^*)$. By selecting a large number, *B*, of bootstrap samples we can extract numerous combinations of x_j *(j=1,2,...,n)* .

The bootstrap algorithm involves the following steps:

i) Compute the statistic $\hat{\theta} = \theta(X)$ from the observed sample *X*.

ii) Select b-th (*b=1,2,...,B*) independent bootstrap sample X_b*, which consists of *n* values drawn with replacement from the observed sample *X*.

iii) Compute the statistic $\hat{\theta}^* = \theta(X_b^*)$ from the b-th bootstrap sample X_b*.

iv) Repeat steps (ii)-(iii) a large number of times (*B* times).

v) Calculate the average of the bootstrap estimates of θ as the arithmetic mean

$$\hat{\theta}^*(\cdot) = \frac{1}{B}\sum_{b=1}^{B}\hat{\theta}_b^* \,. \tag{5}$$

Any individual bootstrap sample will be an imperfect replica of the original sample. As a result, the estimated value of θ obtained from it will differ from what was obtained from the original population. A measure of the accuracy of the estimator $\hat{\theta}$ as an estimate of θ is the bias, which is defined as the difference between the expectation of $\hat{\theta}$ and θ.

$$\text{bias}_{\text{f}} = \text{bias}_{\text{f}}(\hat{\theta}, \theta) = E_f(\hat{\theta}) - \theta . \tag{6}$$

An unbiased estimator will have zero bias, i.e. $E_f(\hat{\theta}) = \theta$. If the bias is positive (negative), then the estimator overestimates (underestimates) the true parameter. The bias-corrected estimator is

$$\hat{\theta}_{bc} = \hat{\theta} - \text{bias}_{\text{f}} . \tag{7}$$

One can approximate the expectation of each bootstrap estimator $\hat{\theta}_b^*$ by the average of the bootstrap estimators $\hat{\theta}^*(\cdot)$ to obtain

$$\text{bias}_{\text{B}} = \hat{\theta}^*(\cdot) - \hat{\theta} . \tag{8}$$

Hence, the bias-corrected estimator of θ is

$$\hat{\theta}_{bc} = \hat{\theta} - \text{bias}_{\text{B}} = 2\hat{\theta} - \hat{\theta}^*(\cdot) . \tag{9}$$

Notice that if $\hat{\theta}^*(\cdot)$ is greater than $\hat{\theta}$, then the bias-corrected estimate $\hat{\theta}_{bc}$ should be less than $\hat{\theta}$. Efron and Tibshirani (1993) point out that bias correction can be problematic in some situations. Even if $\hat{\theta}_{bc}^*$ is less biased than $\hat{\theta}$, it might have substantial greater standard error due to high variability in bias_{B}. The standard error of $\hat{\theta}^*(\cdot)$ is measured as

$$\text{se}_{\text{B}} = \text{se}(\hat{\theta}^*) = \sqrt{\frac{1}{B-1}\sum_{b=1}^{B}(\hat{\theta}_b^* - \hat{\theta}^*(\cdot))^2} . \tag{10}$$

It should be noted, however, that correcting for the bias may result in a larger root mean squared error. If bias_{B} is small compared to the estimated standard error of $\hat{\theta}^*(\cdot)$, then it is safer to use $\hat{\theta}$ than $\hat{\theta}_{bc}$. As a rule of thumb, Efron and Tibshirani (1993) suggest the computation of the ratio of the estimated bootstrap bias to standard error, $\text{bias}_{\text{B}}/\text{se}_{\text{B}}$. If the bias is less than 0.25 standard errors, then it can be ignored.

Finally, we can obtain the bias-corrected estimator from each bootstrap $\hat{\theta}^*_{b,bc}$, *(b=1,2,...B)*. We want the corrected empirical density function of $\hat{\theta}^*_b$, *(b=1,2,...B)* to be centered on $\hat{\theta}_{bc}$, the bias-corrected estimate of θ, i.e. $E(\hat{\theta}^*_{b,bc}) = \hat{\theta}_{\text{bc}}$, *(b=1,2,...B)*. According to this, the bias-corrected estimate from each bootstrap will be

$$\hat{\theta}^*_{b,bc} = \hat{\theta}^*_b - 2\,\text{bias}_B,\ (\text{b} = 1,2,...,\text{B})\,. \tag{11}$$

Once we have the bias-corrected estimates we can use the percentile method to construct the (1-2a)% confidence intervals for θ as

$$(\hat{\theta}^{*(a)}_{bc}, \hat{\theta}^{*(1-a)}_{bc}),\ (\text{b} = 1,2,...,\text{B})\,, \tag{12}$$

where $\hat{\theta}^{*(a)}_{bc}$ is the (100*a$^{\text{th}}$) percentile of the empirical density of $\hat{\theta}^*_{b,bc}$, $(\text{b} = 1,2,...,\text{B})$.

2.2.1 Smooth Bootstrap methodology

One major drawback of the bootstrap procedure outlined is that even when sampling with replacement, a bootstrap sample will not include observations from the parent population that were not drawn in the initial sample in the first place. As a result, the empirical distribution $\hat{f}$ will have jumps at the observed points and look like a collection of boxes of width *h*, a small number, centered at the observations and zero anywhere else. Thus, the bootstrap samples are effectively drawn from a discrete population and they fail to reflect the fact that the underlying population density function *f* is continuous. Hence, the empirical distribution from the bootstrap samples will be an inconsistent estimator of the population density function. This is why it is known as a naïve bootstrap.

One way to overcome this problem is to use kernel estimators as weight functions. The empirical distribution $\hat{f}$ will take the form:

$$\hat{f}(t) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{t - x_i}{h} \right), \tag{13}$$

where h is the window width or smoothing parameters for the density function. $K(.)$ is a kernel function, which satisfies the condition

$$\int_{-\infty}^{\infty} K(x)dx = 1. \tag{14}$$

Usually K is a symmetric probability density function like the normal density function. If we use the standard normal density function as the Kernel density function, then the smoothing is called Gaussian smoothing. The empirical density function then can be written as

$$\hat{f}(t) = \frac{1}{nh} \sum_{i=1}^{n} \phi\left(\frac{t - x_i}{h} \right). \tag{15}$$

Here $\phi(.)$ is the standard density function.

By virtue of the convolution theorem (Efron and Tibshirani, 1993) we can generate the smoothed bootstrap sample $X^{**}=\{x_1^{**}, x_2^{**},\ldots,x_n^{**}\}$ as

$$x_i^{**} = x_i^{*} + h\, \varepsilon_{i,v} \sim f;\ \ i=1,2,\ldots,n, \tag{16}$$

where x_i^{*} is from the naïve bootstrap sample in the previous section.

Sometimes it is the case that the natural domain of the definition of the density function to be estimated is not the whole real line but an interval bounded on one side or both sides. For example we might be interested in obtaining density estimates $\hat{f}$ for which $\hat{f}(x)$ is zero for all negative x. However, the smooth bootstrap could generate points that are outside of the boundaries. One possible solution is to calculate $\hat{f}(x)$ ignoring the boundary restrictions and then to set the empirical density function equal to zero for values of x that are out of the boundary domain. A drawback of this approach is that the estimates of the empirical density function will no longer integrate to unity.

Silverman (1986) suggests the use of the negative reflection technique to handle such problems. Suppose that we are interested in values of x such that $x \geq \alpha$. If the resulting value from the bootstrap is $x_i^{**} < \alpha$, then we will reflect the x_i^{**}, such that $2\alpha - x_i^{**} \geq \alpha$. The empirical density function will be:

$$\hat{f}(t) = \frac{1}{nh} \sum_{i=1}^{n} \left[\phi\left(\frac{t - x_i}{h} \right) + \phi\left(\frac{t - 2\alpha + x_i}{h} \right) \right]. \tag{17}$$

Again by the convolution theorem we can generate the smoothed bootstrap sample $X^{**} = \{x_1^{**}, x_2^{**}, \ldots, x_n^{**}\}$ as

$$x_i^{**} = \begin{cases} x_i^* + h\varepsilon_i & \sim \frac{1}{nh} \sum_{i=1}^{n} \phi\left(\frac{t - x_i}{h} \right) \quad \text{if } x_i^* + h\varepsilon_i \geq \alpha \\ 2\alpha - (x_i^* + h\varepsilon_i) & \sim \frac{1}{nh} \sum_{i=1}^{n} \phi\left(\frac{t - 2\alpha + x_i}{h} \right) \quad \text{otherwise} \end{cases} \tag{19}$$

where x_i^* is from the naïve bootstrap sample in the previous section.

The choice of the smoothing parameter (h) is crucial to the estimated empirical density function. Following Silverman (1986) we can select the value of the window width that minimizes the approximate mean integrated square error. This leads to

$$h = 0.9\, A\, n^{-1/5},$$

where $A = min$ (standard deviation of X, inter-quartile range of $X/1.34$).

The bootstrap algorithm can be re-written as follows:

i) Compute the statistic $\hat{\theta} = \theta(X)$ from the observed sample X.

ii) Select b-th $(b=1,2,\ldots,B)$ independent naive bootstrap sample $X_b^* = \{x_{1,b}^*, x_{2,b}^*, \ldots, x_{n,b}^*\}$, which consists of n data values drawn with replacement from the observed sample X.

iii) Construct the smoothed bootstrap sample $X_b^{**} = \{x_{1,b}^{**}, x_{2,b}^{**}, \ldots, x_{n,b}^{**}\}$, from the naïve bootstrap sample as described in (19).

iv) Compute the statistic $\hat{\theta}^* = \theta(X_b^{**})$ from the b-th bootstrap sample X_b^{**}.

v) Repeat steps (ii)-(iii) a large number of times (say B times).

vi) Calculate the average of the bootstrap estimates of θ as the arithmetic mean

$$\hat{\theta}^*(\cdot) = \frac{1}{B}\sum_{b=1}^{B}\hat{\theta}_b^* \ . \tag{19}$$

If desired, we can calculate the bias, bias-corrected estimates and construct confidence intervals following the steps described above.

2.3 DEA and Bootstrap

Recently Simar (1992, 1996), Simar and Wilson (1998, 2000) set the foundation for the consistent use of bootstrap techniques to generate empirical distributions of efficiency scores and have developed tests of hypotheses relating to returns to scale of bootstrapping. Following Simar and Wilson (1997a) we can describe the existing bootstrap techniques for the output-oriented technical efficiency measure given in (1) with the following algorithm:

i) Solve the DEA problem to obtain $\hat{\phi}_j$ for each firm *j=1,2,…,n*.

ii) Select the b-th *(b=1,2,…,B)* independent naive bootstrap sample $\{\phi_{1,b}^*, \phi_{2,b}^*, \ldots, \phi_{n,b}^*\}$, which consists of n data values drawn with replacement from the estimated values $\hat{\phi}_j$ s.

iii) Construct the smoothed bootstrap sample $\{\phi_{1,b}^{**}, \phi_{2,b}^{**}, \ldots, \phi_{n,b}^{**}\}$, from the naïve bootstrap sample. Notice that all the ϕ_j s are greater than or equal to 1. Therefore, the smoothed bootstrap sample should be appropriately bounded. It will be computed according to:

$$\phi_{j,b}^{**} = \begin{cases} \phi_j^* + h\varepsilon_j & \text{if } \phi_j^* + h\varepsilon_j \geq 1 \\ 2 - (\phi_j^* + h\varepsilon_j) & \text{otherwise} \end{cases}; \ \text{for j} = 1,2,...,\text{n} \ . \tag{20}$$

As before, h is the optimal width that minimizes the approximate mean integrated square error of $\hat{\phi}_j$ s distribution:

$h = 0.9\, A\, n^{-1/5}$, where $A = min$ (standard deviation of ϕ, inter-quartile range of $\phi/1.34$).

iv) Create the b-th pseudo-data set as {(xj*, yj*= $y_j \hat{\phi}_j / \phi_j^{**}$); j=1,2,…,n}.

v) Use the pseudo-data set to compute new $\hat{\varphi}_j^*$ s from the linear program described in (1).

vi) Repeat steps (ii)-(iv) B-times to obtain $\{\hat{\varphi}_{j,b}^*; b=1,2,\ldots,B\}$ for each firm j, j=1,2,…,n.

vii) Calculate the average of the bootstrap estimates of ϕ s, the bias and the confidence intervals as they are described in the previous section.

It should be noted here, that an interpretation of the results obtained from the bootstrap procedure is not always clear. For example, in the b^{th} replication using the pseudo-data consisting of the actual input bundles coupled with the fictitious output levels of firms, the optimal solution φ^* shows the scalar expansion factor for the fictitious output quantity and its inverse is *not a measure of the efficiency of the actual input output bundle.* One may, of course, use the optimal solutions from the (bootstrap) DEA problems to construct measures of the *frontier output level* producible from the fixed input bundle of a firm. Thus, it is more meaningful to construct a 95% confidence interval of the maximum output with lower and upper bounds $[y_L^*, y_U^*]$. In principle, the upper bound (y_U^*) may be used to derive a probabilistic measure of the technical efficiency of an observed input-output bundle. It should be noted that the actually observed output from a given input bundle may exceed its corresponding upper bound.

3. A Bootstrap-Regression Procedure to Capture Unit Specific Effects in DEA

3.1 Combining DEA and Regression

A problem with bootstrapping the technical efficiency measures is the assumption that all firms have the same probability of getting an efficiency score from any specified interval within the (0-1) range. However, there usually are systematic factors that contribute to differences in efficiency and can lead to different technical efficiency scores. For example, for an inter-country

analysis of manufacturing production it is not sensible to conceptualize a data generating process where Germany and Ethiopia have the same probability of getting efficiency scores in excess of 0.975. The existing bootstrapping procedures do not consider the possibility that the distributions of efficiency conditional on unit specific factors may differ across firms. One can argue in favor of including these factors within the scope of the DEA model itself so that the remaining variation in efficiency can be justifiably attributed to purely random factors. However, inclusion of these factors as non-discretionary inputs within the DEA model automatically extends the disposability property (weak or strong) as well as the convexity assumptions to such variables. This is not a realistic assumption in many situations. For example, in the context of measuring the efficiency of public schools, one recognizes a pupil's family income and the level of parental education as socioeconomic conditions in the home-life of the student. These variables do influence the student's performance in school and thereby affect the efficiency level of the school, but a researcher cannot assume that free disposability is applicable. This is one reason why researchers often regress DEA efficiency scores on a number of explanatory variables to adjust for environmental factors; they do not include these factors in the DEA model itself.

A. (as discussed; this argument relates to whether systematic factors should be included in the estimation of the TE model) the systematic factors might violate the weak disposability assumption. More generally they might not be technically inputs if they are not under the direct control of the DMU even if they can be increased or reduced by some other central institution (example government regulations).

B. (this relates to the violation of the 'identical-independent distribution' of the TE scores that the Bootstrap requires) The systematic factors influence the variability of the TE scores and as such even if the TE scores come from similar distributions that are in the same (0,1) interval they have different variance. Thus the distribution of the DMU specific TE scores are not identical.

Additionally, if groups of units have their systematic factors influenced by the same external circumstances then a given external change will result to the systematic factor values to move to the same direction for all the DMUs within the same group and thus the group's TE score will move together as a cluster. Hence, their corresponding Technical Efficiency scores are not independent.

Consider the alternative specifications of the frontier production function

$$y^* = f(x,z) \tag{21a}$$

and

$$y^* = g(x)h(z). \tag{21b}$$

In performing a one-step DEA we assume that $f(x,z)$ is a concave function. An alternative is to assume only that $g(x)$ is a concave function. As shown by Ray (1988), the DEA score obtained from a model incorporating only *x* and *y* captures the factor $g(x)$ of the frontier production function.

Let the vector z^i represent such characteristics of the i-th firm. A regression permits us to determine the part of the technical efficiency that is due to these characteristics and the proportion that is due to random error:

$$\phi_i = \alpha + z^i \gamma + u_i, \tag{22}$$

where $(\alpha+z^i\gamma)$ is the component of technical efficiency that varies systematically with the firm characteristics and u_i is a random error. We estimate the above regression by Ordinary Least Squares to get the estimated $\hat{\phi}(=\hat{\alpha}+z^i\hat{\gamma})$s for the individual observations. In a bootstrap regression analysis one pools the OLS residuals $e_i = \phi_i - (\hat{\alpha}+z^i\hat{\gamma})$ and draws an appropriately smoothed bootstrap sample $e^* = \{e_1^*, e_2^*, ..., e_n^*\}$. These bootstrap residuals can then be used to

construct the pseudo-data $\widetilde{\phi}_i = \hat{\alpha} + z^i\hat{\gamma} + e_i^*$. The corresponding pseudo-value of output would be $\widetilde{y}_i = y_i\widetilde{\phi}_i$.

There is, however, a potential problem. In a bootstrap sample whenever $e_i^* < 1-(\hat{\alpha} + z^i\hat{\gamma}),$ the value of $\widetilde{\phi}_i$ will be less than 1, or equivalently $(\widetilde{\phi}_i - 1) < 0$, which violates the natural restriction on efficiency. To address this problem we apply a version of the reflection method described earlier. Suppose that $\widetilde{\phi}_i = \hat{\alpha} + z^i\hat{\gamma} + e_i^* = 1-\delta_i; (\delta_i > 0).$ We would then replace e_i^* by $e_i^{**} = e_i^* + 2\delta_i$ so that the pseudo value of $\widetilde{\phi}_i$ becomes $1+\delta_i; (\delta_i > 0)$.

3.2 A Bootstrap Regression Procedure

The bootstrap algorithm that generates the distribution of efficiency for each individual firm, conditional on unit specific factors, can be described as follows:

i) For each firm *i* compute ϕ from the DEA model in (1), for *i=1,2,...,n*.

ii) Regress ϕ_i on the firm characteristics z^i.

iii) Calculate the residuals $e_i = \phi_i - \hat{\phi}_i$. for each *i=1,2,...,n*.

iv) Select the b-th (*b=1,2,...,B*) bootstrap sample $e_b^* = \{e_{1b}^*, e_{2b}^*, ..., e_{nb}^*\}$, which consists of *n* pseudo data values drawn with replacement from the observed sample $e = \{e_1, e_2, ..., e_n\}$.

v) Generate the smoothed bootstrap sample

$e_{ib}^{**} = e_{ib}^* + h\varepsilon_i; \quad \varepsilon_i \sim N(0,1)$ for $i = 1,2,...,n$,

where h is the smoothing parameter.

vi) Create the b-th pseudo sample (x^i, y_b^{i*}) *i=1,2,...,n*, where

$y_b^{i*} = y^i * \phi_{i,b}^*$ and

$\phi^*_{i,b} = \hat{\alpha} + z^i\hat{\gamma} + e^{**}_{ib}$ for $i = 1,2,...,n$ if $\hat{\alpha} + z^i\hat{\gamma} + e^{**}_{ib} \geq 1$,

otherwise, $\phi^*_{ib} = \hat{\alpha} + z^i\hat{\gamma}e^{**}_{ib} + 2\delta_i$ where $\delta_i = 1 - (\hat{\alpha} + z^i\hat{\gamma} + e^{**}_{ib}) > 0$.

vii) Use the pseudo-data set to compute new $\hat{\varphi}^*_{i,b}$ s from the linear program described in (1).

viii) Repeat steps (iv)-(vii) B-times to obtain the maximum producible output for each firm *i*, (*i=1,2,...,n*) :

$\hat{y}^f_{i,b} = y^{i*}_b \hat{\phi}^*_{i,b}$; $b = 1,2,...,B$.

ix) Calculate the average of the bootstrap estimates of y^f s, the bias and the confidence intervals.

4. A Study of U.S. Airlines

In this section we present an application of the bootstrap-regression in DEA procedure proposed in this paper to a data set for a number of U.S. airlines from the year 1984. A single output, five-input technology is considered at the DEA stage. The data form a subset of a larger data set constructed by Caves, Christensen, and Trethaway (1984). The output is a quantity index (QYI) constructed from the numbers revenue passenger miles flown, ton-kilometers of cargo flown, and ton-kilometers of mail flown. The inputs are quantity indexes of labor (QLI), fuel (QFI), materials (QMI), flight equipment (QFLI), and ground equipment (QGRI). For the second stage regression we consider the stage length defined by the average distance flown between take off and landing (STAGE), passenger load factor (LOAD), and the number of points served (POINTS) as explanatory variables. The data used for the study are reported in Table 1.

The Output-oriented BCC DEA results using only the output and input quantities are shown in Table 2. Of the 21 firms in the sample, 10 were found to have efficiency equal to 1. RHA and USAir have the lowest levels of efficiency (highest levels of PHI) followed by Ozark, Piedmont, and Air Canada. Table 3 shows the results from a regression of PHI (obtained in Table

2) on STAGE and POINTS. The third explanatory variable LOAD was not statistically significant and was not included in the selected model. As expected an increase in the average length of flights between take of and landing improves efficiency lowering PHI. On the other hand, an increase in the number of POINTS served reduces efficiency. This is consistent with findings elsewhere. The R^2 value of 0.50 shows a moderately good fit. In Table 4 we report the DEA results with the attribute variables (STAGE and POINTS) included in the DEA model. Now RHA becomes 100% efficient while efficiency levels of Piedmont, USAir, and (to a considerable extent) Eastern Airlines improve drastically (i.e., PHI declines noticeably).

It should be noted that the values of PHI reported for the individual firms in either Table 2 (based only on the firm inputs) or Table 4 (based on the inputs and other attributes) are point estimates obtained from a single random sample. To overcome this limitation we performed three sets of bootstraps and obtained the empirical distribution of the frontier output levels from the individual input bundles in the sample.

Table 5 shows the bootstrap average of the frontier output (y^*) producible from the observed input bundle (x) computed from the DEA runs –presented in table 2- that do not include the attributes (z). Note that the y^* obtained from the DEA using the actual (x, y) data is lower than the corresponding bootstrap average value in 15 out of 21 cases. In several cases it is lower than the lower 5-percentile of the bootstrap distribution. This is true of AM, MI, MU, NW, PA, PE, PS, SW, TWA, UN, and WE.

Table 6 presents the results of the bootstrap-regression procedure that captures the impact of the firm specific attributes. This Table shows the bootstrap average and the confidence interval for y^* based on the predicted PHI from the regression model reported in Table 3. What is noticeable about this Table is that the standard deviation of the (predicted) y^* for every single airline shown here is uniformly smaller than what is reported in Table 5. This is not surprising because the y^* in Table 6 is conditional on the attributes (z) while this is not the case in Table 5.

Table 7 reports the mean and confidence limits for the frontier output y^* bootstrap DEA runs incorporating the attributes *(z)* along with the inputs and outputs *(x, y)* presented in Table 1. The average values of the frontier output shown in Table 7 are uniformly lower than the corresponding values reported for the individual airlines in table 6. The fundamental difference lies in the fact that the DEA models underlying Table 7 assume that the output y is a concave function of the variables x and z as in (21a). One the other hand, in deriving Table 6, the underlying function is of the form (21b).

For a better understanding of the differences between the summary results shown in Table 5-7, we report in Table 8 the underlying bootstrap results on the PHIs. The columns labeled $\overline{\phi}^1, \phi^1_{low}$, and ϕ^1_{high} show the average, 5-percentile, and the 95-percentile of the bootstrap PHI from the DEA with only the inputs and output. The next three columns $\overline{\phi}^2, \phi^2_{low}$, and ϕ^2_{high} are the average and the confidence limits of PHI obtained from the regression-bootstrap. Finally, $\overline{\phi}^3, \phi^3_{low}$, and ϕ^3_{high} are the average and confidence limits from the one stage DEA with inputs, attributes, and output. Note that the bootstrap average values $\overline{\phi}^1$ and the corresponding confidence intervals are virtually the same for all firms (with little or no difference in the first two decimal points). This is due to the underlying assumption that the PHI scores are drawn from identical distributions and does not take into account the systematic factors that influence their variation. This is in contrast to the variation of the bootstrap average values $\overline{\phi}^2$ and the corresponding confidence intervals $(\phi^2_{low}, \phi^2_{high})$ are uniformly narrower than what we obtain from the DEA bootstrap with inputs and output. The bootstrap averages $\overline{\phi}^3$ and the associated confidence limits are, again, virtually the same for all firms (without any difference in the first two decimal points), are smaller in magnitude, and with even less variation than what one observes for the DEA bootstrap with inputs and output alone. This is mainly due to the lower values of PHI obtained from the one stage DEA model with the attributes. The regression

bootstrap values of PHI allow explicitly for the variability of the attributes. First this variability is removed from the sample that it is bootstrapped and then it is added back for the creation of the original pseudo- input-output bundles.

4. Summary

A smoothed bootstrap of the DEA scores can generate the empirical density function of the frontier output producible from any specific input bundle. But in many situations there are factors other than the inputs used by a firm that determine the maximum output producible, Bootstrapping from a DEA model that exclude these attributes can lead to misleading results. On the other hand, including these attributes within the DEA itself has its own problems. The regression bootstrap procedure proposed here offers an alternative that generates the distribution of efficiency conditional on the attributes and adds back the systematic influence of these attributes to obtain the distribution of the frontier output.

References

1. Banker, R. D., A. Charnes, and W. W. Cooper (1984) Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis. *Management Science*, vol. 30, no. 9, pp.1078-1092.

2. Caves, D.W., L.R. Christensen, and M. Trethaway. (1984) "Economies of Density versus Economies of Scale: Why Trunk and Local Airline Service Costs Differ"; *RAND Journal of Economics,* Vol 50, No. 6, 1393-1414.

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

4. Efron, B. (1979) Bootstrap Methods: Another Look at the Jackknife" *Annals of Statistics*, **7**, 1-26.

5. Efron, B., and R. J. Tibshirani. (1993) *An Introduction to the Bootstrap.* New York: Chapman Hall, Inc.

6. McCarthy T. A and S. Yaisawarng (1993) Technical Efficiency in New Jersey Scholl Districts. In *The Measurement of Productive Efficiency, Techniques and Applications,* ed. H. O. Fried, C. A. K. Lovell and S. S. Schmidt New York: Oxford University Press, 271-287.

7. Ray, S. C. (1991) Resource-use Efficiency in Public Schools: A Study of Connecticut Data. *Management Science,* vol. 37, no. 12, 1620-1628.

8. Silverman B. W. (1986) *Density Estimation in Statistics and Data Analysis.* New York: Chapman Hall, Inc.

9. Simar, L. (1992) Estimating Efficiencies from Frontier Models with Panel Data: A Comparison of Parametric, Non-Parametric and Semi-Parametric Methods with Bootstrapping. *Journal of Productivity Analysis,* vol. 3, no 1/2, 167-203.

10. Simar, L. (1996) Aspects of Statistical Analysis in DEA-Type Frontier Models. *Journal of Productivity Analysis,* 7, 177-185.

11. Simar, L. and P. Wilson (1998) Sensitivity of efficiency scores: How to bootstrap in Nonparametric frontier models, *Management Sciences*, 44, 1, 49-61.

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

13. Varian, Hal R. (1984) *Microeconomic Theory*, Norton, New York, 1984.

Table 1. Data used in the analysis

	1st Stage						2nd Stage		
	Output	Inputs							
Airline	YI	QFI	QGRI	QLI	QFLI	QMI	STAGE	LOAD	POINTS
Air Canada (AC)	0.0816	0.0702	0.0784	0.0743	0.1083	0.1358	349.5	0.56432	12
American(AM)	1.9365	1.3036	2.1644	1.2637	1.5932	2.209	840.2	0.6256	128
Continental(COT)	0.5455	0.3906	0.4303	0.3078	0.4916	0.6076	813.2	0.62726	95
Delta(DE)	1.3897	1.123	1.7945	1.2116	1.2238	1.7274	569.1	0.52877	94
Eastern(EA)	1.5157	1.1765	1.444	1.2891	1.6191	1.9574	604.5	0.5694	133
Frontier(FR)	0.2133	0.1524	0.1961	0.1723	0.2069	0.3031	445.3	0.63555	91
Midwest(MI)	0.037	0.0456	0.0233	0.0426	0.0788	0.0992	487.1	0.50367	13
Muse(MU)	0.0439	0.0395	0.0323	0.0288	0.048	0.0523	380.6	0.47236	11
Northwest(NW)	1.2485	0.7906	0.6194	0.4998	1.2125	1.224	851.5	0.6093	89
NewYork Air(NYA)	0.0458	0.0459	0.0339	0.0498	0.0673	0.1235	321.2	0.56014	19
Ozark(OZ)	0.1387	0.1236	0.1266	0.1347	0.1826	0.226	420.2	0.55198	59
PanAmerican(PA)	1.5685	0.9764	1.2589	0.9195	1.4026	1.7506	1149.8	0.64368	122
Peoples(PE)	0.3277	0.2154	0.2064	0.1438	0.2924	0.3438	531	0.69819	31
Piedmont(PI)	0.304	0.3004	0.2591	0.325	0.3228	0.4835	347.2	0.52578	66
PacificSouth(PS)	0.155	0.1168	0.2274	0.124	0.1986	0.2176	359.6	0.53781	22
Republic-HughesAir(RHA)	0.4332	0.4369	0.3107	0.4685	0.6375	0.6832	396.2	0.50219	128
SouthWest(SW)	0.1997	0.1806	0.1587	0.131	0.1796	0.1987	320.9	0.58517	22
TWA	1.5134	0.9349	1.5457	0.8959	1.3134	1.7681	968.1	0.62178	90
United(UN)	2.4424	1.5965	2.7084	1.484	2.1049	2.4479	786.1	0.60482	151
USAir(USA)	0.4214	0.374	0.4883	0.4134	0.5468	0.6453	374.3	0.58552	71
Western(WE)	0.4933	0.3547	0.3141	0.3509	0.4229	0.5251	623	0.57708	76

Table 2. DEA results: Output Oriented Technical Efficiency (TE) using only Input-Output Quantities		
Airline	**PHI (x,y)**	**TE (x,y)**
AC	1.15285	0.86741
AM	1	1
COT	1.02996	0.97091
DE	1.06789	0.93642
EA	1.13464	0.88134
FR	1.0695	0.93502
MI	1	1
MU	1	1
NW	1	1
NYA	1.0926	0.91525
OZ	1.29826	0.77026
PA	1	1
PE	1	1
PI	1.23075	0.81251
PS	1.10169	0.9077
RHA	1.43376	0.69747
SW	1	1
TWA	1	1
UN	1	1
USA	1.39496	0.71687
WE	1	1

Table 3: Regression Results of PHI (x,y) on attributes STAGE and POINT					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t**	**Pr > |t|**
Intercept	1	1.24144	0.05731	21.66	<.0001
STAGE	1	-0.00051	0.00012	-4.22	0.0005
POINTS	1	0.00197	0.000643	3.06	0.0067
R-Square	0.5003				
Adjusted R-Square	0.4447				

Table 4. DEA results: Output Oriented Technical Efficiency (TE) using Input-Output Quantities and Attributes (STAGE, POiNTS)			
Airline	**PHI (x,y,z)**	**TE (x,y,z)**	**Difference of PHI (x,y,z) from PHI (x,y)**
AC	1	1	0.15285
AM	1	1	0
COT	1.02996	0.97091	0
DE	1.00472	0.9953	0.06317
EA	1	1	0.13464
FR	1.05554	0.94739	0.01396
MI	1	1	0
MU	1	1	0
NW	1	1	0
NYA	1	1	0.0926
OZ	1.28427	0.77865	0.01399
PA	1	1	0
PE	1	1	0
PI	1.02049	0.97992	0.21026
PS	1.01345	0.98673	0.08824
RHA	1	1	0.43376
SW	1	1	0
TWA	1	1	0
UN	1	1	0
USA	1.08481	0.92182	0.31015
WE	1	1	0

Table 5. Bootstrap Results: One-step Bootstrap of PHI (x,y)						
			Bootstrap		95% confidence interval	
Airline	Observed Output (YI)	Frontier Output (Y^*=YI∗PHI(x,y))	Average Output	Standard Deviation	Lower Limit	Upper Limit
AC	0.0816	0.0941	0.0912	0.0103	0.0818	0.1180
AM	1.9365	1.9365	2.1692	0.2460	1.9413	2.8087
COT	0.5455	0.5618	0.6085	0.0679	0.5474	0.7884
DE	1.3897	1.4840	1.5451	0.1667	1.3933	2.0073
EA	1.5157	1.7198	1.6965	0.1944	1.5196	2.1986
FR	0.2133	0.2281	0.2386	0.0273	0.2140	0.3105
MI	0.037	0.0370	0.0413	0.0046	0.0371	0.0535
MU	0.0439	0.0439	0.0489	0.0053	0.0440	0.0632
NW	1.2485	1.2485	1.3907	0.1530	1.2525	1.7970
NYA	0.0458	0.0500	0.0510	0.0056	0.0459	0.0659
OZ	0.1387	0.1801	0.1555	0.0177	0.1391	0.2014
PA	1.5685	1.5685	1.7469	0.1935	1.5726	2.2710
PE	0.3277	0.3277	0.3656	0.0412	0.3285	0.4745
PI	0.304	0.3741	0.3384	0.0372	0.3049	0.4375
PS	0.155	0.1708	0.1728	0.0188	0.1554	0.2233
RHA	0.4332	0.6211	0.4837	0.0538	0.4344	0.6257
SW	0.1997	0.1997	0.2230	0.0250	0.2002	0.2888
TWA	1.5134	1.5134	1.6879	0.1888	1.5179	2.2133
UN	2.4424	2.4424	2.7307	0.3105	2.4502	3.5477
USA	0.4214	0.5878	0.4703	0.0522	0.4225	0.6067
WE	0.4933	0.4933	0.5486	0.0596	0.4946	0.7088

Table 6. Bootstrap Results: Bootstrap-Regression to capture Unit Specific Effects (based on Regression in Table3)						
			Bootstrap		95% confidence interval	
Airline	Observed Output (YI)	Frontier Output (Y^*=YI∗PHI(x,y))	Average Output	Standard Deviation	Lower Limit	Upper Limit
AC	0.0816	0.0941	0.0902	0.0068	0.0820	0.1065
AM	1.9365	1.9365	2.1243	0.1522	1.9441	2.4881
COT	0.5455	0.5618	0.5910	0.0351	0.5476	0.6747
DE	1.3897	1.4840	1.5810	0.1323	1.3981	1.8836
EA	1.5157	1.7198	1.8182	0.1578	1.5496	2.1486
FR	0.2133	0.2281	0.2549	0.0225	0.2179	0.3027
MI	0.037	0.0370	0.0401	0.0023	0.0371	0.0455
MU	0.0439	0.0439	0.0481	0.0034	0.0441	0.0562
NW	1.2485	1.2485	1.3574	0.0745	1.2519	1.5148
NYA	0.0458	0.0500	0.0515	0.0042	0.0460	0.0612
OZ	0.1387	0.1801	0.1597	0.0139	0.1397	0.1893
PA	1.5685	1.5685	1.7663	0.1182	1.5801	2.0055
PE	0.3277	0.3277	0.3559	0.0223	0.3288	0.4096
PI	0.304	0.3741	0.3630	0.0311	0.3109	0.4279
PS	0.155	0.1708	0.1728	0.0135	0.1556	0.2031
RHA	0.4332	0.6211	0.5601	0.0462	0.4804	0.6551
SW	0.1997	0.1997	0.2255	0.0189	0.2007	0.2672
TWA	1.5134	1.5134	1.6730	0.1037	1.5206	1.8898
UN	2.4424	2.4424	2.7995	0.2350	2.4641	3.3152
USA	0.4214	0.5878	0.5022	0.0438	0.4307	0.5941
WE	0.4933	0.4933	0.5417	0.0387	0.4948	0.6344

Table 7. Bootstrap Results: One-step Bootstrap of PHI (x,y,z)						
			Bootstrap		95% confidence interval	
Airline	Observed Output (YI)	Frontier Output (Y^*=YI∗PHI(x,y,z))	Average Output	Standard Deviation	Lower Limit	Upper Limit
AC	0.0816	0.0816	0.0837	0.0049	0.0816	0.1048
AM	1.9365	1.9365	1.9919	0.1274	1.9370	2.4884
COT	0.5455	0.5618	0.5606	0.0351	0.5456	0.7007
DE	1.3897	1.3963	1.4232	0.0798	1.3900	1.7828
EA	1.5157	1.5157	1.5546	0.0912	1.5161	1.9464
FR	0.2133	0.2251	0.2196	0.0143	0.2134	0.2740
MI	0.037	0.0370	0.0380	0.0023	0.0370	0.0475
MU	0.0439	0.0439	0.0450	0.0026	0.0439	0.0564
NW	1.2485	1.2485	1.2813	0.0762	1.2488	1.6028
NYA	0.0458	0.0458	0.0470	0.0029	0.0458	0.0589
OZ	0.1387	0.1781	0.1426	0.0092	0.1387	0.1783
PA	1.5685	1.5685	1.6076	0.0938	1.5688	2.0154
PE	0.3277	0.3277	0.3359	0.0191	0.3278	0.4205
PI	0.304	0.3102	0.3115	0.0177	0.3041	0.3904
PS	0.155	0.1571	0.1591	0.0097	0.1550	0.1990
RHA	0.4332	0.4332	0.4453	0.0279	0.4333	0.5562
SW	0.1997	0.1997	0.2053	0.0130	0.1997	0.2566
TWA	1.5134	1.5134	1.5527	0.0908	1.5137	1.9427
UN	2.4424	2.4424	2.5028	0.1447	2.4429	3.1353
USA	0.4214	0.4571	0.4329	0.0262	0.4215	0.5410
WE	0.4933	0.4933	0.5057	0.0290	0.4934	0.6324

Table 8. Mean and Confidence Intervals for ϕ from Alternative Bootstraps									
	One-Step Bootstrap of PHI(x,y)			Bootstrap-Regression of PHI(x,y) on z			One-Step Bootstrap of PHI(x,y,z)		
Airline	$\bar{\phi}^1$	ϕ^1_{low}	ϕ^1_{high}	$\bar{\phi}^2$	ϕ^2_{low}	ϕ^2_{high}	$\bar{\phi}^3$	ϕ^3_{low}	ϕ^3_{high}
AC	1.11765	1.00245	1.44608	1.10539	1.0049	1.30515	1.02574	1	1.28431
AM	1.12017	1.00248	1.4504	1.09698	1.00392	1.28484	1.02861	1.00026	1.285
COT	1.11549	1.00348	1.44528	1.08341	1.00385	1.23685	1.02768	1.00018	1.28451
DE	1.11182	1.00259	1.44441	1.13766	1.00604	1.3554	1.02411	1.00022	1.28287
EA	1.11928	1.00257	1.45055	1.19958	1.02237	1.41756	1.02566	1.00026	1.28416
FR	1.11861	1.00328	1.4557	1.19503	1.02157	1.41913	1.02954	1.00047	1.28458
MI	1.11622	1.0027	1.44595	1.08378	1.0027	1.22973	1.02703	1	1.28378
MU	1.1139	1.00228	1.43964	1.09567	1.00456	1.28018	1.02506	1	1.28474
NW	1.1139	1.0032	1.43933	1.08722	1.00272	1.2133	1.02627	1.00024	1.28378
NYA	1.11354	1.00218	1.43886	1.12445	1.00437	1.33624	1.0262	1	1.28603
OZ	1.12112	1.00288	1.45205	1.15141	1.00721	1.36482	1.02812	1	1.28551
PA	1.11374	1.00261	1.44788	1.12611	1.0074	1.27861	1.02493	1.00019	1.28492
PE	1.11565	1.00244	1.44797	1.08605	1.00336	1.24992	1.02502	1.00031	1.28319
PI	1.11316	1.00296	1.43914	1.19408	1.0227	1.40757	1.02467	1.00033	1.28421
PS	1.11484	1.00258	1.44065	1.11484	1.00387	1.31032	1.02645	1	1.28387
RHA	1.11657	1.00277	1.44437	1.29294	1.10896	1.51223	1.02793	1.00023	1.28393
SW	1.11668	1.0025	1.44617	1.12919	1.00501	1.33801	1.02804	1	1.28493
TWA	1.1153	1.00297	1.46247	1.10546	1.00476	1.24871	1.02597	1.0002	1.28367
UN	1.11804	1.00319	1.45255	1.14621	1.00888	1.35735	1.02473	1.0002	1.2837
USA	1.11604	1.00261	1.43972	1.19174	1.02207	1.40982	1.02729	1.00024	1.28382
WE	1.1121	1.00264	1.43685	1.09811	1.00304	1.28603	1.02514	1.0002	1.28198