Click here to download Manuscript: Barnum&Gleason\_FixedProportionTechnology\_21july20it0k dhere to view linked References

# Measuring Efficiency under Fixed Proportion Technologies

Darold T. Barnum (corresponding author) Departments of Managerial Studies, Information & Decision Sciences, and Pharmacy Administration University of Illinois at Chicago (MC 243) 601 South Morgan Street Chicago, IL 60607-7123 USA (v) 312-996-3073 (f) 312-996-3559 dbarnum@uic.edu

John M. Gleason Department of Information Systems and Technology College of Business Administration Creighton University Omaha, NE 68178 USA (v) 816-272-6088 jgleason@creighton.edu

*Abstract:* Data Envelopment Analysis (DEA) applications frequently involve nonsubstitutable inputs and nonsubstitutable outputs (that is, fixed proportion technologies). However, DEA theory requires substitutability. In this paper, we illustrate the consequences of nonsubstitutability on DEA efficiency estimates, and we develop new efficiency indicators that are similar to those of conventional DEA models except that they require nonsubstitutability. Then, using simulated and real-world datasets that encompass fixed proportion technologies, we compare DEA efficiency estimates with those of the new indicators. The examples demonstrate that DEA efficiency estimates are biased when inputs and outputs are nonsubstitutable. The degree of bias varies considerably among Decision Making Units, resulting in substantial differences in efficiency rankings between DEA and the new measures. And, over 90 percent of the units that DEA identifies as efficient are, in truth, not efficient. We conclude that when inputs and outputs are not substituted for either technological or socio-economic/legal reasons, conventional DEA models should be replaced with models that account for nonsubstitutability.

*Keywords:* Fixed proportion technology, Data Envelopment Analysis, Efficiency measurement, Nonsubstitutable inputs, Nonsubstitutable outputs

JEL classification codes: C14, C61

# **1** Introduction

When inputs are nonsubstitutable, they cannot replace each other in the production of a constant amount of output. Such inputs must be utilized in a fixed proportion to produce their output, and any quantity of an input in excess of the required ratio is wasted. Production systems using nonsubstitutable inputs are well-known in economics, and are often called "Leontief" technologies or "Fixed Factor Proportion Technologies" (Beattie and Taylor 1985). If outputs are nonsubstitutable, then, for a constant amount of input, production of one output cannot be increased by producing less of another. Such production systems are "Fixed Product Proportion Technologies" (Beattie and Taylor 1985). When both inputs (factors of production) and outputs (products) are nonsubstitutable, we call the production systems "Fixed Proportion Technologies."

Sometimes, an industry's inputs (outputs) that are <u>technologically</u> substitutable may, in practice, never be substituted for each other. This could occur because every operation follows a common industry practice to avoid risk of experimentation. Or, if price relationships are relatively uniform across the industry, all organizations may be producing at the same allocatively efficient combinations of inputs and outputs. Other possible causes of nonsubstitutability include custom, norms, regulations and laws. The reasons for nonsubstitutability in the set of data being analyzed are immaterial for its empirical analysis. However, empirical evidence may be required to indentify the presence (or absence) of substitutability if the fixed proportions are not mandated by the technology.

## 1.1 Examples of DEA applications using nonsubstitutable variables

Nonsubstitutable inputs and outputs are employed frequently in applications of Data Envelopment Analysis (DEA), almost always involving the Charnes-Cooper-Rhodes (CCR) and the Banker-Charnes-Cooper (BCC) linear programming models. Examples from industries using nonsubstitutable variables include hospitals, electricity generation, and urban mass transit.

Hollingsworth (2008) reports that DEA studies of hospital efficiency almost always include inputs of staffed beds (as a proxy for capital) and employees (often subdivided into various categories). Although some other proxy for capital might produce a different conclusion, staffed beds and number of staff are technological complements, not substitutes (Barnum et al. 2010a). Indeed, nonsubstitutability among inputs is actually much more widespread in hospitals, as one referee noted:

Within a given hospital department or cost center, capital and labor are non-substitutable, and are in fact complements (think pharmacists and medications in a pharmacy). But it actually goes further than that, because the outputs of various departments/cost centers are themselves (often, but not always) complements, and are used in conjunction with other outputs to produce a more global output (pharmaceutical care and physical rehabilitation; central services and acute care, etc.). Thus, the bias that occurs from using standard DEA to analyze hospital-wide efficiency is much larger than one might expect.

Almost all of the DEA studies involving electricity generation included in Zhou, Ang and Poh's recent survey (2008) used inputs of generating capacity as a proxy for capital, number of employees as a proxy for labor, and amount of fuel as a proxy for energy. These are nonsubstitutable for technological reasons in the short run, and substitution has been minor in the long run as well (Fuss 1977; Fuss and McFadden 1978; Woodruff et al. 2005).

Many DEA studies of transit operations utilize inputs of number of available buses, number of employees and gallons of fuel, and outputs of vehicle miles and passenger miles (Barnum 2009). These variables are nonsubstitutable for technological reasons (Barnum et al. 2008).

Some specific DEA applications using inputs or outputs that are likely to exhibit nonsubstitutability, all employing CCR and/or BCC, include an analysis of residential energy efficiency using outputs of living space area and number of residents (Grösche 2009); an analysis involving cigarette production using inputs of quantity of labor and quantity of raw material (Iwasaki and Tremblay 2009); a banking study using inputs of deposits, labor and physical capital (Ray 2007); electricity distribution using inputs of number of employees and length of electric grid (Cullmann and von Hirschhausen 2008); a banking analysis using outputs of loans and interest income (Denizer et al. 2007); and a study of street lighting efficiency using inputs of staff and number of lamps (Lorenzo and Sánchez 2007). Although we think it likely that the aforementioned variables are not substituted for each other, definitive determination would require statistical testing.

Further, there are cases where inputs (outputs) clearly could be substituted for each other, and substitution would be expected to occur. But, at least for the dataset involved, there is no substitution. Consider Olesen and Petersen's (2009) study, in which they compare various DEA models using outputs of student performance on standardized PISA-reading and PISA-other tests, inputs involving measures of annual schooling hours per student and teachers per 100 students, and parental education as an environmental control. Holding inputs and the environmental factor constant, one would expect that reading scores, and an aggregation of science and math scores (called "other scores"), would be traded off as input resources were shifted from one subject area to the other. That is, holding inputs constant, the outputs should show a negative relationship with one output decreasing as the other increases, thereby tracing out a typical piece-wise production possibility curve on the frontier. In fact, when PISA-reading is used as the dependent variable, PISA-other is an independent variable, and inputs and the environmental influence are held constant by including them as control variables, the positive regression coefficient between the two outputs is +1.04 with a t-statistic of 30 and an R-square of 0.996 (Figure 1). (As discussed later in the paper, the positive relationship merely means that the organizations inefficient in producing one output are also inefficient in producing the other. The high correlation between output inefficiencies might have overwhelmed any substitution effects. But, it is still the case that no substitution between outputs can be estimated with these data.)

### -Put Figure 1 about here-

Further, different measures of inputs, outputs and environmental factors might affect these results, as might adjusting for endogeneity and other statistical concerns. Recall again, however, that it is <u>not</u> the true population relationships that we seek to identify, but the <u>actual relationships</u> in the data being used for the DEA. That is, the DEA estimates are based on the extant data's characteristics and relationships, as would be true for any empirical estimates.

### 1.2 Substitutability requirements in DEA theory

Although nonsubstitutable inputs (outputs) have frequently been utilized in DEA applications, DEA mathematical models and the economic theory underlying them

require substitutability among inputs (outputs). Farrell (1957) and Charnes, Cooper and Rhodes (1978) assume substitutability, as do Färe, Grosskopf and Lovell (1994) and more recent theoretical work (Färe and Grosskopf 2004; Cooper et al. 2007b; Fried et al. 2008).

#### 1.3 Addressing the conflict between theory and application

Therefore, the many DEA applications that utilize nonsubstitutable inputs or outputs are in conflict with theory. In this paper, we demonstrate the impact on DEA efficiency scores of nonsubstitutable inputs and outputs.

In Section 2, we provide an illustration of the effects of nonsubstitutability on empirical data. Then, we identify the efficient frontier when a fixed proportion technology is present, and illustrate the differences between efficiency estimates under substitutable and nonsubstitutable relationships. In Section 3, we develop a simple additive efficiency indicator for fixed proportion technologies-Fixed Proportion Additive (FPA) measure—that is directly comparable to Charnes and Cooper's additive measure. The main purpose of this section is to illustrate simply and intuitively the measurement issues arising under fixed proportion technologies. Because additive measures are seldom used in practice, we then move on to the more complex measures needed for ratio indicators. Section 4 first offers a graphical introduction to ratio efficiency measures under fixed proportion technologies. Then it presents the mathematical development of one such measure—Fixed Proportion Ratio (FPR) measure. Finally, it presents the mathematical development of another—Fixed Proportion Charnes Cooper Rhodes (FPCCR) measure. Section 5 compares efficiency estimates of the CCR, FPCCR and FPR measures using a real-world transit dataset and a Monte Carlo simulation. Section 6 examines the practically of applying the FPR and FPCCR measures to real-world data, and identifies methods for adapting DEA measures for fixed proportion technologies. Section 7 concludes. An appendix briefly introduces statistical identification of whether a fixed or variable proportion technology is present in a dataset.

# 2 Characteristics of fixed proportion technologies

Consider a simple illustration of one output and two inputs. Suppose one unit of output is produced using the amounts of each input shown in Figure 2, for the

moment ignoring point E. If the two inputs are substitutable and output is constant, then, for frontier DMUs, as one input decreases, the other must increase, resulting in a negative slope between the two inputs.

#### -Put Figure 2 about here-

Because A, B, C and D each produced one unit of output, and a decline in one input always resulted in an increase in the other, it might appear that substitution is possible in this sample, and these four DMUs estimate the frontier. That is, the frontier observations might provide the necessary evidence to assume that the inputs are substituted when DMUs are efficient.

However, exactly the same pattern would occur if the inputs were <u>not</u> substitutable. A DMU is unlikely to be equally efficient in its use of both inputs, there being no economic or logical reason to assume such equality. Presume for the moment that two inputs are not substitutes. For example, suppose that in truth a fixed ratio of 0.04 units of Input A and 0.05 units of Input B are necessary to efficiently produce one unit of output. And, one DMU uses 0.11 units of A and 0.05 units of B, and another DMU uses 0.04 units of the A and 0.13 of the B. Then, both would be on the alleged frontier and would exhibit a negative relationship between the inputs, but in truth neither are efficient. The same reasoning applies for more than two DMUs with varying degrees of inefficiencies.

Therefore, negative relationships between variables on the frontier provide <u>no</u> evidence concerning their substitutability. Indeed, the linear program followed by DEA in constructing the input frontier <u>always</u> yields a piecewise frontier with an increasingly negative slope when moving west, similar to the one shown in Figure 2, regardless of whether the variables are substitutable or nonsubstitutable.

In fact, Figure 2 is based on the real-world dataset used later in this paper, and the two inputs are *not* substitutable. One input is available driver hours and the other is actual bus hours for a set of transit systems, with each system's inputs being divided by its vehicle miles output so they represent the inputs used to produce one vehicle mile of output. It logically would not be possible to maintain constant vehicle miles by substituting driver hours for vehicle hours, or vice versa.

Also, many transit systems keep extra drivers (called extra board) available. And, because of speed differences, the vehicle hours needed to produce one vehicle mile

will differ among transit systems. Extra board is determined by a system's absenteeism rates, run-cutting protocols and union contracts, while speed is determined primarily by environmental influences. Thus, a transit system's relative level of excess for the two inputs would *not* generally be equal. Therefore, the faux piecewise isoquant in Figure 2 is due to differences in the relative efficiency with which DMUs employ each input, and probably some noise in the data as well.

Note, however, that there appears to be a slightly positive relationship between the two inputs. If output had <u>not</u> been held constant, then one would expect the two inputs to increase together as a result of increases in output, regardless of whether they were substitutable. However, because output is held constant, the positive relationship is not due to increasing output. Instead, it results from the fact that DMUs that are inefficient in the use of one input tend to be inefficient to a somewhat similar degree in the use of the other. Increases to the north reflect increasing inefficiency in the use of Input 2, and increases to the east reflect increasing inefficiency in the use of Input 1. So, increases in a northeasterly direction means that organizational inefficiencies tend to occur to somewhat equal degrees for both inputs. However, the relationship is weak, as shown by an R-square of only +0.13 between the two inputs. Indeed, that fact that the relationship between the input inefficiencies is slight provides further justification for our previous observation that the DMUs on the faux piecewise isoquant reflect differences in the relative efficiency with which each DMU employs the two inputs.

#### 2.1 Best-practice frontiers under fixed proportion technologies

When all inputs and outputs are truly nonsubstitutable, the best-practice frontier consists of a single point in vector space, which we call the "point frontier," whose coordinates are identified by the maximum values of each output/input combination across all DMUs j = 1,...,J. Suppose there are *m* inputs  $(x_m)$ , *n* outputs  $(y_n)$  and j DMUs. The m\*n-dimensional point frontier would be

$$[Max(y_{1j} | x_{1j}), Max(y_{1j} | x_{2j}), ..., Max(y_{1j} | x_{mj}), Max(y_{2j} | x_{1j}), Max(y_{2j} | x_{2j}), ..., Max(y_{nj} | x_{mj})], \forall j$$
(1)

Of course, it is unlikely that any one DMU would be the most efficient for more than a few output/input ratios, at most. Further discussion of this point is provided later when we develop measures for DMU efficiency under fixed proportion technologies.

The simple case of two inputs and a single, constant output is illustrated by the composite DMU at point E in Figure 2. Although not shown on the graph, the frontiers of the reference set increase vertically and horizontally from point E, forming a right-angle or L-shaped reference set frontier, with its boundaries estimated by the maximum output/input ratio in each dimension. Note that the only Pareto-Koopmans efficient subset of the reference set frontier is the point frontier, because only at that point can none of the inputs or outputs be improved without worsening some of the other inputs or outputs.

As can be seen from Figure 2, if the inputs are truly nonsubstitutable then DEA over-estimates the efficiency of DMUs A-D by identifying them as 100 percent efficient when they are not. The four DMUs on the substitutable (piecewise-isoquant) frontier are not efficient if they are measured against E on the nonsubstitutable (point) frontier. Likewise, DMU F is about 90 percent efficient if the inputs are substitutable but less than 50 percent efficient if they are not. Precisely the same problem plagues nonsubstitutable outputs. In short, if inputs and outputs are not substitutable, DEA yields biased efficiency estimates and falsely identifies inefficient DMUs as efficient. And, as demonstrated later in the paper, the degree of bias is not constant across DMUs.

Of course, both conventional DEA estimation and the preceding fixed-proportion estimation ignore noise. DEA estimates are known to be sensitive to extreme points caused by noise, and point frontier estimates would also be sensitive. This does not present a problem in theoretical comparisons of the models, since both DEA and the point estimation method are based on noise-free assumptions. For the case of two inputs and one output, both DEA and fixed-proportion estimates of a focus DMU's efficiency are based on frontier input and output values from only two DMUs, so the number of frontier DMUs used to estimate efficiency is the same in both cases.

Demonstrating which, if either, of the efficiency estimations would be more sensitive to noise is not a trivial problem, and is beyond the scope of this paper. However, if sensitivity estimates are desired for any of the fixed proportion models, one could employ variations on the same methods used for conventional DEA, such as chance-constrained programming (Cooper et al. 2002; Tser-yieth 2002), use of

means (Ruggiero 2004), fuzzy set estimations (Hougaard 2005), sensitivityrobustness-stability analysis (Neralic and Wendell 2004; Cooper et al. 2004a; Cherchye et al. 2007), bootstrapping (Simar and Wilson 2000), or statistical panel data analysis (Barnum et al. 2010b).

## 3 Additive measure under nonsubstitutability

In order to estimate each DMU's efficiency relative to the point frontier in Figure 2, one simple possibility would be a variation of Charnes' and Cooper's DEA additive model (Cooper et al. 2007a, p. 94), which we abbreviate as **VPA** for *Variable Proportion Additive* model. When the VPA model is used to estimate a DMU's efficiency in producing one output from two inputs, the efficient point for a DMU being assessed is the furthest point on the piecewise isoquant frontier where neither of the DMU's inputs has increased and its output has remained constant. The rectilinear distance between the DMU being assessed and that point is the measure of the DMU's inefficiency.

We call our variation the *Fixed Proportion Additive* (**FPA**) model, because it assumes that the inputs and outputs are not substitutable. Exactly like Charnes' and Cooper's VPA model, the degree of inefficiency is measured by the rectilinear distance between the DMU being assessed and the efficient point. But, for the FPA model, the efficient point is the *point frontier* rather than a point on a piecewise isoquant frontier. In both cases, efficiency <u>decreases</u> as scores increase, so, the higher a DMU's score, the more inefficient it is. The only difference between the models is the location of the point from which efficiency is measured. This can be seen in Figure 3 for two inputs and one constant output. As a result, direct comparisons of the two models' inefficiency scores can be made, because both models' inefficiency metric is exactly the same. Therefore, valid estimates can be made of the bias in DEA scores when inputs and outputs are nonsubstitutable.

#### -put Figure 3 about here-

As can be seen in Figure 3, the VPA model reports DMUs A, B, C and D to be efficient, while the FPA model reports that none of these DMUs is efficient, as would be expected based on the VPA- and FPA-estimated frontiers. For the DMU being assessed, the VPA model's inefficiency score is 0.04 and the FPA model's

inefficiency score is 0.09, more than twice as high. The FPA inefficiency score  $\left(FPA_{k}^{I}\right)$  for DMU<sub>k</sub>, where the superscript I indicates an input-oriented measure, can be obtained mathematically from a set of j = 1, 2, ..., J DMUs (Equation 2).

$$FPA_{k}^{I} = \sum_{m=1}^{M} \left[ \left| \left( x_{km} / y_{k1} \right) - M_{j} in(x_{jm} / y_{j1}) \right| \right]$$
(2)

To insure that the scores of the VPA model will be directly comparable to those of the FPA model, each DMU's inputs are divided by its sole output. Therefore, the input and output values used to compute the  $(FPA_k^I)$  score (Equation 2) and the  $VPA_k$  model (Linear Program 3.0 – 3.2) are identical, so the resulting sums of the slacks for the assessed DMU *k* are directly comparable.

$$VPA_{k} = Max \sum_{m=1}^{M} s_{m}^{-}$$
3.0

Subject to 
$$\sum_{j=1}^{J} \lambda_j (x_{jm} / y_{j1}) + s_m^- = x_{km} / y_{k1}$$
  $m = 1, 2, ..., M$  3.1

$$\lambda_j \ge 0 \qquad \qquad j = 1, 2, \dots J \qquad \qquad 3.2$$

If there are multiple outputs and one input, then the FPA model in Equation 4 can be used, with the FPA score  $\left(FPA_{k}^{O}\right)$  measuring the inefficiency of DMU<sub>k</sub>. The comparable VPA measure is shown in the linear program 5.0 - 5.2.

$$FPA_{k}^{O} = \sum_{n=1}^{N} \left[ \left| \left( y_{kn} / x_{k1} \right) - Max_{j} \left( y_{jn} / x_{j1} \right) \right| \right]$$
(4)

$$VPA_{\lambda_j} = Max \sum_{n=1}^{N} s_n^+$$
5.0

Subject to 
$$\sum_{j=1}^{J} \lambda_j (y_{jn} / x_{j1}) + s_n^+ = y_{kn} / x_{k1}$$
  $n = 1, 2, ..., N$  5.1

 $\lambda_j \ge 0 \qquad \qquad j = 1, 2, \dots J \qquad 5.2$ 

All four of these additive models yield values in non-negative dimensional units as measures of inefficiency, with the higher the value, the higher the inefficiency.

Now, consider an illustrative example. Figure 4 shows two outputs commonly used in transit DEA studies, Passenger Miles and Vehicle Miles, each of which have been divided by the input Operating Expense (adjusted for differences across time and location), so the data represent the passenger miles per dollar of operating expense and vehicle miles per dollar of operating expense.

#### -Put Figure 4 about here-

Given that input is identical for all DMUs, we construct the piecewise variable product proportion frontier H-B-Q-R. The point fixed product proportion frontier is P, the coordinates of which are a composite of DMU Q, which yields the most vehicle miles per dollar of operating expense, and DMU B, which yields the most passenger miles per dollar of operating expense.

We know that the two outputs are not substitutes because increases in vehicle miles result in increases in passenger miles, with typical elasticities ranging from +0.6 to +1.0 (Litman 2004). If input is held constant, as it is in Figure 4, then one would expect a DMU's inefficiencies in the production of one of the outputs would be somewhat mirrored by its inefficiencies in the production of the other. This is precisely what is shown in Figure 4, with joint inefficiencies increasing in the southwest direction. The relationship between the output inefficiencies is weak, having an R-square value of only +0.09.

Using the  $VPA_k$  model from Linear Program 5.0 – 5.2, and the  $FPA_k^o$  model from Equation 4, we illustrate in Table 1 selected results (DMUs labeled A-G in Figure 4) for both measures (in dimensional units).

#### -Put Table 1 about here-

Both DMUs (B and Q) found efficient by the VPA model are <u>not</u> efficient. Because we know substitution is impossible, B is inefficient in the production of vehicle miles (its FPA score is 1.07), and Q is inefficient in the production of passenger miles (its FPA score is 0.02).

Note that all of the DMUs have lower efficiency under the FPA measure than under the VPA measure, which are real differences since both models are using the same metric. Also, the efficiency rankings by the FPA and VPA models are inconsistent, with FPA rankings sometimes being higher and other times lower than the VPA rankings. Finally, sometimes the efficiency rankings differ substantially; F and G (for example) are respectively ranked 46 and 126 ranks more efficient by VPA then by FPA (out of 216 possible ranks). For this simple illustration of the problem, the DMUs analyzed were chosen to focus on the extent of rank differences that can occur, and most differences are not so great. But, because DEA scores are often used to compare individual DMUs, the DEA model produces efficiency estimates for some DMUs that have little to do with reality.

These are total differences in efficiency and do not identify or adjust for causes of these differences. For example, the level of passenger miles per vehicle mile will be affected by such factors as fare level, proportion of the population served who don't have cars, population density, and income levels, to name the most important factors. What we have measured is indeed overall inefficiency, but proportions of inefficiency that are caused by management and by the environment are not identified.

Comparing the VPA and FPA additive models is our effort to make the efficiency differences in the presence on nonsubstitutability intuitively clear. Indeed, in the interests of intuitive understanding, our FPA models require the use of either one input or one output, so are not amenable to the case of multiple inputs and multiple outputs. Further, additive models are almost never used in practice, so comparisons tell us little about the extent of bias for the ubiquitous ratio models, and developing a multiple-input, multiple-output additive model would not be a productive exercise. More useful efficiency measures for nonsubstitutable inputs and outputs would be ones whose results could be compared to ratio measures such as CCR. In the next three sections, therefore, we develop ratio measures to deal with nonsubstitutability. The first builds on the FPA measure, and the second on the conventional CCR measure that is adapted for nonsubstitutability.

# **4** Ratio efficiency measures

 In this introductory section we graphically compare the CCR measure and the two new ratio measures that assume nonsubstitutability. The two sections following this one present the mathematical development for each of the two new measures. Figure 5 shows the output/input ratios for two outputs and a single input, for three DMUs (A,B,C). "A" is the focus DMU for which we wish to compute efficiency.

--put Figure 5 about here--

CCR is a ratio measure, because it is estimated by the ratio of aggregated outputs to inputs of the focus DMU divided by the aggregated outputs to inputs ratio of a composite DMU on the efficiency frontier. Graphically, the maximum efficiency attainable by the focus DMU occurs at the point on the frontier intersected by a ray beginning at the origin and passing through the focus DMU. For the situation at hand, this point occurs for a composite frontier DMU with an output/input ratio of 7.29/1 for Output 1 and 5.83/1 for Output 2. Thus, on the frontier, the total output produced by one unit of input is the aggregated value of the outputs, 7.29 + 5.83 = 13.125. The total output produced by the focus DMU for one unit of input is 5 + 4 = 9. Therefore the focus DMU's DEA efficiency is 9/13.125 = 0.6857. This is exactly the same efficiency obtained by the ratio of the length of the ray from the origin to the focus DMU (6.0431) to the length of the ray from the origin to the efficiency frontier (9.3379).

This equality is not surprising. Geometrically, if we label the sides of the right triangle ODE as a1 for its opposite side, b1 for its adjacent side and c1 for its hypotenuse, and similarly label the sides of the right triangle OAF as a2 for its opposite side, b2 for its adjacent side and c2 for its hypotenuse, then c2/c1 = (a2+b2)/(a1+b1) = focus ray/composite ray = sum of focus outputs/sum of composite outputs, because a2/b2 = a1/b1.

Our *JPA* Associate Editor demonstrated the same conclusion algebraically, assuming constant returns to scale and no slacks (which are implicitly present geometrically): Let y1 and y2 be the coordinates of the observed outputs, and y1\* and y2\* be the coordinates of the target outputs (which are on the frontier). Because the observed and target coordinates are both on a ray through the origin,  $y1/y1* = y2/y2* = \theta$ . Therefore  $(y1+y2)/(y1*+y2*) = (\theta y1*+\theta y2*)/(y1*+y2*) = \theta(y1*+y2*)/(y1*+y2*) = \theta$ 

In short, CCR efficiency can be solved as ratios of aggregated outputs to inputs. It does so by identifying the efficient point on the frontier that will maximize the focus DMU's efficiency score, in the present case the point that maximizes the sum of the outputs. It assumes that the DMUs used to construct the composite DMU are efficient and the two outputs are substitutable, so any point on the line between them (producing less of one output and more of the other) are also efficient.

However, what if the two outputs are not substitutable? Then, if only B and C are being compared, we know that B is producing 1 unit too little of output 1, and C is producing 4 units too little of output 2. Maybe worse, the composite DMU located at D is inefficient in the production of both outputs! Clearly, if outputs are not substitutable and everything else is equal, the composite DMU at D is the wrong point with which to compare the focus DMU A.

How, then, can we estimate the right point, with the most efficient DMU having a score of 1? For our models below, we suggest two different ways. Both assume that substitution is not possible, so convex combinations of multiple DMUs are invalid.

The first estimation method is to follow the DEA protocol of using the ratio of the aggregated outputs to the aggregated inputs. For the present case, A's score would be (5+4)/(1) = 9, B would score (7+7)/(1) = 14, and C would report (8+3)/(1) = 11. The resulting efficiency scores are A = 9/14 = 0.64, B = 14/14 = 1, and C = 11/14 = 0.79.

The second estimation method is to use the average of the individual output/input efficiency ratios, with each individual output/ input efficiency being the focus DMU's output/input ratio for the output and input in question divided by the maximum output/input ratio for the same output and input. Recall that this criterion for choosing efficient points was used to develop the earlier FPA measures. As shown in Figure 5, C has the highest output ratio for output 1 (8) and B has the highest output ratio for output 2 (7). So, A's non-normalized efficiency would be (5/8 + 4/7)/2 = 0.60, B's score would be (8/8 + 3/7) = 0.71, and C's would be (7/8 + 7/7)/2 = 0.94. Normalizing these scores so they follow the DEA convention of efficiency equal to 1, A = 0.60/0.94 = 0.64, B = 0.71/0.94 = 0.76, and C = 0.94/0.94 = 1.

If the data contain noise, then both CCR and the two new measures are subject to bias caused by outliers, with the likelihood of bias increasing as the number of inputs and outputs decreases. For example, for the two-output, one-input case (Figure 5), all three indicators are estimated from only two outputs and two inputs – if any one of these four is extreme, the estimates will be biased. It is beyond the scope of this paper to identify differences in the three indicators' reliability, but, as discussed

earlier, a variety of statistical approaches can be used to adjust for reliability if desired. Discussion of the relative strengths and weaknesses of the new ratio measures is in the next two sections, which cover their mathematical development.

In conclusion, please recall that CCR and the two new measures are all based on the ratio of outputs to inputs. The major difference between them is their identification of efficient points in production hyperspace, with CCR choosing its points based on the assumption of substitutability, and the remaining two choosing their points based on the assumption of nonsubstitutability.

## 4.1 Average of individual output/input ratios

In order to develop a measure of a DMU's relative degree of inefficiency in the use of an input to produce an output, we start with an FPA measure in which one input is used to produce one output, for DMUs j = 1, 2, ..., J:

$$FPA_{k}^{O} = \left| (y_{kn} / x_{km}) - Max_{j}(y_{jn} / x_{jm}) \right| \qquad m=1, n=1 \qquad (6)$$

Note that the point from which DMU k's efficiency is measured is the point frontier. In order to transform this equation from yielding dimensional units to dimensionless proportions, we simply divide rather than subtract:

$$eff_{kmn} = (y_{kn} / x_{km}) / Max_{j} (y_{jn} / x_{jm}) \qquad m=1, n=1$$
(7)

We estimate the efficiency of DMU k in producing output n from input m by dividing the assessed DMU's output/input ratio by that of the DMU j that is the most efficient for that particular output/input ratio. This equation differs from the initial FPA equation only in that its measure is a dimensionless proportion rather than a dimensional difference. So, the range of efficiencies for the output/input pair is [0,1],

with at least one DMU  $\left( DMU_{k=Max}_{j} \right)$  achieving efficiency of 1.

As is done with conventional DEA, we allow the most efficient ratios to envelop the data. That is, for <u>each</u> output/input ratio, the DMU with the maximum value is

assumed to be efficient and that value is used as the basis of comparison for the same ratio of all remaining DMUs.

Because an assessed DMU's relative efficiency normally would vary for each output/input combination, the DMU's average efficiency can be computed as the mean of its individual efficiencies. Note that a similar approach is taken by both the VPA and the FPA measures, except that with the additive measures it is the efficiency levels of each output-input combination in dimensional units that are aggregated, so it is total units of inefficiency rather than the mean proportional efficiency that is estimated, but the principle is identical in both cases.

Thus, for each of DMU *k*'s output/input ratios, Equation 7 is used to compute the normalized efficiency measure for that ratio. If there are *m* inputs and *n* outputs, then there will be  $m \times n$  efficiency measures of the form  $eff_{kmn}$ . So, for each DMU *k* in a set of J DMUs, the mean of its  $m \times n$  efficiency measures is computed, which yields a partially normalized efficiency measure for that DMU. Then, each DMU *k*'s partially normalized efficiency measure is divided by the maximum partially normalized efficiency measure, which yields a normalized efficiency measure in [0, 1]. We call this the *Fixed Proportion Ratio* (FPR) measure (Equation 8):

$$FPR_{k} = \frac{(1/mn)\sum_{m=1}^{M}\sum_{n=1}^{N} eff_{kmn}}{Max_{j}\left((1/mn)\sum_{m=1}^{M}\sum_{n=1}^{N} eff_{jmn}\right)}$$
(8)

The last normalization does not change the relative efficiencies, of course, but it provides a measure that follows the DEA convention of using an efficiency index in [0, 1], so the relatively most efficient DMU(s) will have a score of 100 percent. The reader may recall that measures such as CCR are also normalized in [0, 1] by changing the base of comparison for an assessed DMU from peer DMUs to the assessed DMU itself when a score of 1 is reached, thereby assuring that that a score of 1 will never be exceeded. Although the normalization procedure differs, the reason for normalizing is the same: efficiency is not an absolute value but is measured relative to the most efficient DMU(s) in the dataset. Therefore, the most relatively efficient DMU(s) must be assigned a score of 100 percent (or 1).

The key characteristics of the FPR measure include the following. Technical and scale efficiencies are included in the FPR, and the measure is monotone decreasing in each input excess and output shortfall. Further, FPR envelops the data. Each individual ratio of a DMU being assessed is compared to the maximum ratio for that particular output/input pair, which yields its efficiency relative to the most efficient pair. Then, the mean of each DMU's individual ratio is compared to the maximum mean ratio, which yields its mean efficiency relative to the maximum mean efficiency. As a result, the formula complies with the DEA convention of basing efficiency on the most efficient observations in the set of units being analyzed.

Because of the nature of the formula's construction, it is easy to determine which of the assessed DMU's output/input pairs are the sources of inefficiency. Further, if an analyst wished to weight some pairs more heavily than others, the formula could easily be modified to accommodate this.

The only cause of different efficiency rankings between the FPA and the FPR would be that one is based on aggregated dimensional unit values and the other on aggregated dimensionless ratios, but both use the same efficient points from which to determine an assessed DMU's efficiency. Of course, the FPR score is units-invariant so it will remain consistent even if the measurement units for inputs or outputs are changed, which is not true for the FPA measure.

## 4.2 Ratio of aggregated outputs to aggregated inputs

The FPR index measures efficiency as the average of the individual output/input ratios. Another measure would be the ratio of the aggregated outputs to aggregated inputs, which is closer to conventional DEA ratio indicators.

To change the conventional substitutable-variable CCR to a fixed proportion CCR indicator (FPCCR), we begin with the simple ratio form of the conventional CCR measure, which is based on maximizing the ratio of the aggregated weighted outputs to the aggregated weighted inputs by varying the output and input weights (Cooper et al. 2004b). Program 9.0 - 9.2 maximizes the efficiency score of DMU *k* by determining output weights for *R* outputs ( $u_r$ ) and input weights for *I* inputs ( $v_i$ ) that results in DMU *k* having the highest score possible compared to the composite DMU that is the first to reach 1 with that particular set of weights applied to forming its

weighted aggregated output to weighted aggregated input. If this composite DMU  $(DMU^*)$  is the DMU *k*, then DMU *k* is efficient; otherwise, DMU *k*'s efficiency is reported as a proportion of the composite DMU\*.

$$M_{u,v} \theta_{k} = \sum_{r=1}^{R} u_{r} y_{rk} / \sum_{i}^{I} v_{i} x_{ik}$$
(9.0)

s.t. 
$$\sum_{r=1}^{R} u_r y_{rj} / \sum_{i}^{I} v_i x_{ij} \le 1$$
  $j = 1, 2, ..., J$  (9.1)

$$u_r, v_i \ge 0$$
  $r = 1,...R; i = 1,...I$  (9.2)

This model adjusts for declining rates of technical substitution among inputs, and declining rates of technical substitution among outputs. For the two-input, one-output case, for example, it is the declining rate of technical substitution that results in the conventional bowl-shaped isoquant. Total input usage will be at a minimum when the slope of the isoquant is -1, and the total amount of input will increase to the right or the left of that point. For DEA to report 100 percent efficiency for a DMU anyplace on the isoquant, it is necessary to assign higher weights to inputs that are relatively unused and lower weights to inputs that are relatively heavily used. So we would expect any DMU on a 45 degree slope portion of the isoquant to have equal weights for each input, a DMU to the left to have heavier weight on the horizontal input, and a DMU to the right to have a heavier weight on the vertical input, so all three would yield the same weighted output/input ratio and therefore be judged equally efficient even though they use different amounts of total input.

If, however, the inputs cannot be substituted, then there is no isoquant and the justification for differing weights disappears. Only total input usage matters, with total being at a minimum when the minimum amount of unweighted aggregated input for a given amount of output is used. Likewise, the highest total amount of output will be produced when the maximum amount of unweighted aggregated output is produced for a given amount of input.

As a result, if all variables are nonsubstitutable, maximum efficiency can be estimated by the DMU that has the highest aggregated-output to aggregated-input ratio, using equal weights for all outputs and equal weights for all inputs. The following changes can be made in the CCR model in order for it to be transformed into a model permitting no substitution. Because weights cannot vary by input or by output, a common u and a common v are required:

$$\underset{u,v}{Max}\theta_{k} = \sum_{r=1}^{R} uy_{rk} / \sum_{i}^{I} vx_{ik}$$
(10.0)

s.t. 
$$\sum_{r=1}^{R} u y_{rj} / \sum_{i}^{I} v x_{ij} \le 1$$
  $j = 1, 2, ..., J$  (10.1)

$$u, v \ge 0 \tag{10.2}$$

Then, we can place the weights outside the summations.

$$\underbrace{Max}_{u,v} \theta_k = (u / v) \left( \sum_{r=1}^R y_{rk} / \sum_{i=1}^I x_{ik} \right)$$
(11.0)

$$(u / v) \left( \sum_{r=1}^{R} y_{rj} / \sum_{i=1}^{I} x_{ij} \right) \le 1 \qquad j = 1, 2, ..., J$$
 (11.1)  
$$u, v \ge 0 \qquad (11.2)$$

The u/v ratio can be increased until the aggregated output/input ratio of one or more of the j DMUs equals 1. Since we will not need mathematical programming, the equation set becomes:

$$\underbrace{Max}_{u,v} \theta_k = (u / v) \left( \sum_{r=1}^R y_{rk} / \sum_i^I x_{ik} \right)$$
(12.0)

$$u / v = \frac{1}{\left(\sum_{r=1}^{R} y_{rj} / \sum_{i}^{I} x_{ij}\right)^{*}} \qquad j = 1, 2, ..., J \qquad (12.1)$$

 $u, v \ge 0 \tag{12.2}$ 

65

.

where the u/v ratio equals that of the DMU(s) with the highest aggregated output/input ratio, with that ratio starred\* above. So, for DMU k, the maximum efficiency attainable ( $\theta_k$ ) is

$$\theta_{k} = \left(\sum_{r=1}^{R} y_{rk} / \sum_{i}^{I} x_{ik}\right) / \left(\sum_{r=1}^{R} y_{rj} / \sum_{i}^{I} x_{ij}\right)^{*}$$
(13.0)

In short, the FPCCR program can be solved for DMU k's efficiency score with the ratio of DMU k's output/input to the highest output/input ratio of the DMUs in the dataset. Again note that the only difference between the original CCR program and the FPCCR equation is the assumption that weights must be equal because no substitution is possible.

One more step is necessary, however, in order to make the FPCCR measure unitsinvariant. The CCR measure deals with differing units by adjusting the weights assigned to each input and output, which cannot be done here. However, an input or output can be made dimensionless and therefore units-invariant by replacing its values with the ratio of each individual value to the maximum value for that variable.

In this way, all variables are rendered dimensionless, and their values range from 0 to 1. We chose to form the ratio of each variable value to the maximum value for its variable so the range would be 0 to 1, because values in the range [-1,+1] are optimal for numerical stability (Golub and Van Loan 1989). Changing from dimensional to dimensionless values in no way affects DEA computations and indeed commercial DEA programs automatically change the units of the variables in order to improve the numerical stability of the computations. So

$$\hat{x}_{ij} = x_{ij} / Max(x_{ij}) \qquad i = 1, 2, \dots, I; j = 1, 2, \dots J \qquad (14.0)$$

$$\hat{y}_{rj} = y_{rj} / Max(y_{rj})$$
  $r = 1, 2, ..., R; j = 1, 2, ..., J$  (14.1)

And the revised equation with dimensionless variables is:

$$\theta_{k} = \left(\sum_{r=1}^{R} \hat{y}_{rk} / \sum_{i=1}^{I} \hat{x}_{ik}\right) / \left(\sum_{r=1}^{R} \hat{y}_{rj} / \sum_{i=1}^{I} \hat{x}_{ij}\right)^{*}$$
(15.0)

Several characteristics of the FPCCR measure are of special interest. First,  $\theta_k$  is based on the same metric under CCR and FPCCR. Second, the point of maximum efficiency is estimated by the DMU(s) that have the most efficient aggregated output to aggregated input ratio. The FPCCR measure does not account for the fact that the DMU with the highest average aggregated ratio may not have the highest efficiency for each individual output/input combination, and indeed, may not be the most efficient for <u>any</u> individual output/input combination. Unlike the FPR measure discussed earlier, which uses the most efficient DMU for each output/input set for identifying the overall efficient point, the FPCCR procedure estimates the point efficiency frontier using the DMU with the highest efficiency from the aggregated outputs/aggregated inputs. Third, unlike the FPR, the FPCCR measure bases its efficiency measures on an actual DMU rather than a composite DMU, thus removing the concern that the composite DMU might not be obtainable in the real world.

# **5** Applications comparing CCR to FPR and FPCCR

In this section, we compare the scores of the FPR and FPCCR efficiency indicators with those of the CCR measure. Because the FPR and FPCCR measures incorporate both technical and scale efficiencies, we compare them to a conventional ratio DEA model that also does so, rather than considering ratio models that assume variable returns to scale and therefore measure only technical efficiency. In our comparisons, we are interested mainly in any differences among the models for those DMUs that each model identifies as efficient, and in differences among models' rankings of efficiencies. The key comparisons are efficiency rankings and correlations, and categorization of DMUs as efficient or inefficient.

In this paper, we use a simulated dataset whose parameters insure nonsubstitutability, and a transit dataset in which the inputs and outputs are nonsubstitutable for technological reasons. So, it is not necessary to perform statistical testing herein, but testing is necessary for those situations where a technological imperative for nonsubstitutability is absent, but may occur.

#### 5.1 Monte Carlo dataset

We illustrate the FPR and FPCCR measures vs. the CCR measure with randomlygenerated data sets of 66 DMUs, each DMU with three nonsubstitutable inputs and three nonsubstitutable outputs. We use three of each because examples using fewer than three inputs and three outputs are less likely to be generalizable to higher dimensions. And, we use 66 DMUs so the results are more easily compared with our transit dataset, which includes 66 transit systems.

Suppose, for full efficiency, the ratio among the three inputs  $(x_1, x_2, x_3)$  must be 1:2:5, which we show as an input vector [1,2,5], and the ratio among the three outputs  $(y_1, y_2, y_3)$  must be 1:2:4, which we show as an output vector [1,2,4]. Further, suppose that each unit of the input vector will yield 8 units of the output vector, so the fixed proportion technology that converts z units of the input vector to 8z units of the output vector can be shown by

$$z[1,2,5] \rightarrow 8z[1,2,4] \tag{16}$$

So, for example, if z = 3 units, then, for full efficiency,  $(x_1, x_2, x_3) = 3(1,2,5) = (3,6,15)$ , which implies  $(y_1, y_2, y_3) = 8(3)(1,2,4) = (24,48,96)$ . Further, when inputs and outputs are nonsubstitutable, by knowing the value of any one of the inputs or outputs, we can easily determine the other input and output values that are required for full efficiency.

Each dataset was created as follows. The  $x_1$  values were randomly generated from a uniform distribution over (0, 1]. The values of the remaining inputs and outputs were computed using the relationships specified above, resulting in a data set with 66 completely efficient DMUs that embody the fixed proportions.

To render these DMUs inefficient to various degrees, input amounts must be increased for some inputs and/or output amounts must be decreased for some outputs. Therefore, for each of the  $66 \times 3$  inputs, a random value was generated from a uniform distribution over [1, 1.5]. For each of the  $66 \times 3$  outputs, a random value was generated from a uniform distribution over [0.5, 1]. The products of each originally efficient (input or output) value and its corresponding random value form a dataset. There is no random noise in these data. All of the variations from perfect efficiency are the results of using too much input or producing too little output.

The entire procedure of creating a dataset for 66 DMUs, and computing their FPR, FPCCR and CCR scores, was repeated 1000 times. The Monte Carlo simulation was written in Visual Basic Applications for Excel, using Excel 2007's random number generator. FPR and FPCCR scores were computed with Excel routines, and the DEA scores were calculated with the Premium Solver Platform for Microsoft Excel (Frontline 2009).

#### 5.2 Transit dataset

As noted earlier, the inputs and outputs frequently used in bus transit DEA studies are nonsubstitutable. The inputs of fuel, labor and buses are not substitutes but must be used in fixed proportions in efficient systems. Excess labor or excess buses are simply wasted. Likewise, the common outputs of vehicle miles of service and passenger miles are not substitutes, and inputs cannot be shifted from one to the other in order to increase one by decreasing the other. Thus, for this set of inputs and outputs, the FPR and FPCCR models are applicable, and the CCR model is not.

We use 1999-2004 data for buses operated by agencies that reported at least one million vehicle miles of annual service, excluding the New York City MTA\_(United\_States\_Federal\_Transit\_Administration 1997- 2008). This results in 66 usable DMUs, with six observations for each.

## 5.3 Comparing FPR and FPCCR results

We begin by reporting the comparison between the two indicators designed to measure efficiency when neither inputs nor outputs are substitutable. The two indicators were developed from entirely different starting points, and differ somewhat in how they measure efficiency. FPR efficiency represents a composite that is composed of the set of the maximum output/input ratios, while the FPCCR efficiency is based on the actual DMU that has the maximum ratio of aggregated outputs to aggregated inputs. Even if the two ratios were to measure efficiency from the same efficient point(s), they would often yield different values and efficiency rankings. It is well known that the mean of output/input efficiency ratios does not necessarily equal the ratio of the output to input sums (Hitch and McKean 1960; Gleason and Barnum 1982, 1999; Fox 1999; Färe and Grosskopf 2004). Although each indicator has unique characteristics, we would hope that their results would be reasonably

consistent in order to obtain triangulation, thereby improving confidence in their ability to estimate efficiency when inputs and outputs are not substitutable.

We compared the correlation between the two measures, the DMUs each identified as inefficient and efficient, and the similarity of their efficiency rankings. The mean R-square between the indicators was 0.996 for the simulation data and 0.959 for the transit data. For the 1000 Monte Carlo trials and the 6 years of transit data, each measure for each dataset always reported one DMU as efficient. The DMUs ranked as efficient were virtually identical. For the transit dataset, both measures ranked the same DMU as efficient in all six trials. For the simulation dataset, the same DMU was ranked as efficient in 94 percent of the 1000 trials, and the DMU that the FPCCR ranked as most efficient was ranked in the top 6 in the remaining trials. For the entire 66 observations, estimated efficiencies differed by 6 ranks or less for 94 percent of transit ranks and 99.7 percent of simulation ranks.

In sum, although the scores of the two measures are not identical, as expected, they very closely approximate each other, with the differences that do exist attributable to the different method by which efficiencies are computed. Further discussion of differences between these two measures is beyond the scope of this paper, but their relative closeness in scores makes it possible to use them to triangulate true efficiency in the presence of nonsubstitutable variables.

#### 5.4 Comparing CCR results with FPR and FPCCR

Although the FPR and FPCCR scores were much alike, they both differed greatly from the CCR scores (Table 2). R-square values between CCR, and FPR and FPCCR respectively, averaged 0.61 and 0.58 over six trials for the transit data and 0.69 and 0.67 over 1000 trials for the simulation data.

#### -Put Table 2 about here-

Of more concern is that of the 66 observations, for DMUs reported as inefficient by the FPR and FPCCR measures, CCR reported a mean of 10.33 DMUs as efficient for the transit data and a mean of 15.29 DMUs as efficient for the simulation data.

As shown in Table 3, of the DMUs that CCR reported as efficient but FPR and FPCCR did not, over half were not in the top decile of efficiency scores in either the FPR or FPCCR, and a mean of 7 percent were in the bottom half of the efficiency

rankings of the FPR and FPCCR indicators. Indeed, in all four cases, DMUs that CCR reported as <u>efficient</u> were ranked in the <u>least efficient decile</u> by the fixed proportion measures!

-Put Table 3 about here-

Looking at absolute differences in ranks (Table 4), CCR's reported ranks differed from the other two by over 6 ranks in more than half of the cases, and, in all four scenarios, differed by more than 33 of the 66 ranks about 2 percent of the time.

-Put Table 4 about here-

In sum, based on both the transit and the Monte Carlo data, CCR is an invalid measure of efficiency when inputs and outputs are not substitutable. Not only is the CCR's overall relationship to measures requiring nonsubstitutable poor, but, in terms of evaluation of individual DMUs, its efficiency estimates sometimes differ drastically from those of the two other measures.

It is beyond the scope of this paper to attempt an empirical analysis to identify the reasons for the differences, but we suspect that the following phenomenon is key. DEA efficiency scores are primarily determined by the largest outputs and the smallest inputs, with these variables being assigned the heaviest weights. Lower outputs (higher inputs) are assumed to be at points on the production frontier that are necessary to remain on the frontier. Substituting their relatively small (large) sizes allows other variables to maintain their relative sizes of the opposite magnitude. But, if there is no substitutability, then the low outputs and high inputs are <u>not</u> the result of substitution but the result of inefficiency, and their excess should be recorded as inefficiency instead of being "swept under the rug" by assigning them low weights. When there is no substitutability, then there is no justification for varying weights.

## 6 Other issues

## 6.1 Practicality of applying FPR and FPCCR to real-world data

An important concern remains regarding identification of the point frontier; namely, will it be the appropriate benchmark for measuring the efficiency of all DMUs? The issue was well-stated by a referee for the original version of this paper, using as an example transit systems with fixed product proportion outputs of vehicle

miles and passenger miles, and fixed factor proportion inputs of available buses, fuel, and employees:

Even if vehicle miles and passenger miles are complements, the degree of complementarily may differ substantially across DMUs. . . . Firms may have different mixes of bus fleets (and ages of the busses in those fleets), and those busses may have different fuel mileage characteristics. There may be differences in union contracts (or other factors not variable or under the manager's control) in which firms (even the truly efficient ones) must use different ratios of labor hours and available vehicles. In at least some instances, this implies different efficient points. One can certainly account for this by restricting the data set and comparing a smaller, more narrowly defined group of DMUs, but this simply leads to other problems, such as the curse of dimensionality.

First, although it often is not explicitly considered in applications, in theory all DMUs in a set to be analyzed by DEA are expected to use the same technology. So, if an industry utilizes several substantially different production technologies, then only those DMUs using a particular technology should be analyzed together. In fixed-route bus public transit, technological differences are related to fleet size. It has been shown that to validly compare transit systems, it is necessary to group observations by system size, as Karlaftis and his colleagues have demonstrated with a variety of different samples and diverse statistical methodologies (Karlaftis and McCarthy 1997; Karlaftis et al. 1999; Karlaftis and McCarthy 2002; Karlaftis 2003, 2004). For that reason, we have limited our sample to the nation's largest systems, excluding New York, by using only transit agencies producing at least 1 million vehicle miles annually (United\_States\_Federal\_Transit\_Administration 1997- 2008).

Second, issues such as differences in union contracts may influence labor/vehicle ratios. Just as is true for most first-stage DEA models, no attempt is made to categorize efficiency as controllable or not controllable by management. That is, we are measuring overall relative efficiency, not efficiency under management's control. It certainly is the case for both DEA and fixed proportion models that differences in union work rules will cause lower efficiency in some systems than in others, and that environmental conditions will cause huge differences in ratios of passenger miles to vehicle miles. But, in these and all similar cases, total efficiency <u>is</u> lower when work rules hold down productivity or when buses are operated in low-density, high-income suburbs versus high-density, low-income central cities. A second-stage analysis in

both DEA and fixed proportion models might be able to isolate the efficiency under management's control. Herein, we have sought to measure only total efficiency, which includes efficiency differences caused by the environment and/or by management activities, and, for that matter, technology. Therefore, the same efficient point is applicable to all.

# 6.2 Alternatives allowing utilization of DEA models under fixed proportion technologies

Our goal in developing indicators to measure efficiency in the presence of nonsubstitutability is the correct identification of the production frontier using the same inputs and outputs as conventional DEA, so the efficiency scores of the two types of models would be comparable.

However, beyond such comparisons, or if some variables have fixed relationships but others are substitutable, there are adaptations of the input and output variables that permit the use of conventional DEA models with fixed proportion technologies. One method is to aggregate nonsubstitutable variables if they are measured with comparable units, or, if not, aggregate them using their prices (or some other logical choice) as weights. We believe that this solution is a good one for inputs if prices are available and can be adjusted for price differences over time and among DMUs.

A second solution is to use conventional DEA models but utilize only one of the nonsubstitutable inputs and one of the nonsubstitutable outputs. Because nonsubstitutable variables occur in a fixed proportion for efficient DMUs, they will increase and decrease together. So, one can serve as a rough proxy for all. This solution has been used in at least two cases (Banker et al. 2004; Casu and Thanassoulis 2006). The problem is that this approach does not account for differences in a DMU's efficiency in producing different outputs or in using different inputs. But, in the absence of comparable prices or other acceptable weighting, it may be the best choice available if one wishes to use conventional DEA models.

A third possibility is to combine variables using statistical methodologies such as Exploratory Factor Analysis, Principle Components Analysis, Confirmatory Factor Analysis, or a variation of two-stage regression analysis (Greene 2008; Wooldridge 2002). A variation on this solution was the use of Multi-Directional Efficiency Analysis instead of DEA because of the substitutability issue (Holvad et al. 2004)

There is at least one conventional DEA solution that is not appropriate for dealing with fixed proportions, namely models based on Banker and Morey's efficiency analysis for exogenously fixed inputs and outputs (1986). It is probably obvious that the word "fixed" is being used in two different contexts. The former refers to a *constant proportional relationship* between two variables, while the latter means that *one variable value remains unchanged* as the other variable increases or decreases. That is, the former refers to a fixed relationship between two variables, while the latter refers to a refers to one variable's value being fixed while the other variable changes.

The main purpose of our fixed proportion models is to identify biases in conventional DEA models when inputs (outputs) are not substitutable. Therefore we have tried to maintain as many DEA characteristics as possible in our model development. Certainly the FPA model fully achieves this objective when compared to the VPA model. Both the FPA and VPA models use the same variables as input, and both use the same metric to measure inefficiency. The only difference between them is the point(s) identified as efficient. Therefore, their results are directly comparable and true differences in their efficiency estimates can be identified.

The FPR and FPCCR models approximate the CCR model as closely as possible. All three measure dimensionless, proportional efficiency ranging from zero to 1, and all use identical inputs and outputs. The FPR and CCR efficiency levels are not directly comparable because they are based on different metrics, although the FPCCR and the CCR are more directly comparable because they are based on the same metric. In any case, differences in efficiency rankings between CCR and the other two will identify bias present under the CCR model.

All of the alternatives allowing use of the conventional DEA models are worthy of consideration when estimating efficiency. However, none of them are as useful as the FPR and FPCCR models for <u>comparisons</u>, because all reduce the number of inputs (outputs), sometimes by combining them and sometimes by eliminating them. Changing the number of variables can have a significant impact on the estimated efficiency, as is well known. More damaging for comparisons, all of the alternative DEA models either eliminate some of the inefficiencies or combine variables in ways that make it impossible to know the impact of the various inefficiencies on the final

score. The FPR and FPCCR models take all actual inefficiencies into account. The FPR has one other advantage when compared with both the conventional and alternative measures: it identifies the precise sources of the inefficiencies. For example, if the overall efficiency is the result of inefficiencies in the use of input 1 and the production of output 3, then that can be specifically identified and marked for correction.

In short, we believe the FPR and the FPCCR are the most valid measures currently available for comparison to the CCR. In some cases, they also might be the measures of choice if a researcher's objective were solely to estimate the relative efficiencies of the DMUs, but in other cases one of the other alternatives might better fit the researcher's needs. Because our primary objective is comparison, we only present comparisons between CRR and the two new models.

# **7** Conclusions

In this paper we identify the effects on efficiency measurement if conventional DEA models are used when the inputs and outputs are nonsubstitutable. Using data illustrating fixed proportion technologies, we compare the scores of popular DEA measures that assume substitutability to the scores of new measures that assume nonsubstitutability. The key analyses are between the conventional CCR measure (which assumes substitutability) and two new measures, FPR and FPCCR (designed to be comparable to the CCR except for assuming nonsubstitutability). Two datasets are used to compare the measures. The first includes 66 transit systems with two outputs and three inputs, with data for each of 6 years; contemporaneous efficiency scores are computed for each system for each of the six years. The second involves a Monte Carlo simulation. Three nonsubstitutable inputs and three nonsubstitutable outputs are used (with each of the six rendered inefficient to random degrees), and input-output data are computed for 66 DMUs. This simulation was repeated 1000 times, with contemporaneous efficiency scores computed for each repetition.

On average, the CCR model identified 17 percent (11.33) of the transit systems efficient each year, and 25 percent (16.22) of the observations in each repetition of the simulation. In almost all cases, one of the DMUs identified as efficient by CCR was also identified as efficient by FPCCR and FPR. However, of the remaining

DMUs that CCR identified as efficient, FPCCR and FPR found <u>all</u> to be inefficient. CCR rendered the DMUs efficient by placing them at various points along a production frontier, but, given that we <u>know</u> that substitutability is impossible, there in truth can be no convex production frontier with varying combinations of the inputs and outputs. In short, whether or not one accepts the FPCCR and FPR results, the extra DMUs identified as efficient by CCR must be misidentified, because it is simply wrong to assume that they can be placed along a production frontier, or that the variables can be assigned varying weights, because such variances depend on substitutability being possible. The bottom line is that CCR identified large numbers of DMUs as efficient when they were not.

Further, CCR did a poor job of ranking DMUs by their true efficiency. While the FPCCR and FPR indicators provided almost-identical efficiency estimations, the R-squared between each of the two and CCR averaged less than two-thirds over both datasets. Looking at absolute differences in ranks, CCR's reported ranks differed from the other two by over a decile in more than half of the cases, and, in all four scenarios, differed by more than 33 of the 66 ranks in 2 percent of the cases. Indeed, some of the DMUs that CCR ranked as efficient were to be found in the least-efficient decile of the other two measures.

# Acknowledgements

We are very appreciative of the valuable assistance that we received from the anonymous referees. They helped us think through many critical issues. As a result, the final paper is very different from our initial submission, and infinitely better.

## Appendix: Identifying fixed vs. variable proportion relationships

A rigorous, comprehensive examination of statistical methodologies for identifying fixed vs. variable proportion relationships is far beyond the scope of this paper, but a brief discussion of some of the issues may be worthwhile.

If variables are nonsubstitutable because of the technology, then no further evidence is necessary and models accounting for their fixed proportions must be used. However, inputs (outputs) that are <u>technologically substitutable</u> may, in practice, <u>not</u> be substituted for each other, for reasons such as those identified in the opening paragraphs of this paper. Reasons for nonsubstitutability in a dataset are immaterial for deciding whether fixed or variable proportion efficiency models should be used.

If fixed proportion relationships are not mandated by the technology, then empirical evidence supporting the hypothesized relationship could be important. Olsen and Petersen's 2009 *JPA* paper illustrates the problem. Although reading and math score outputs <u>could</u> be substitutes as input resources are shifted from one to the other, they are <u>not</u> substituted for each other in Olsen and Petersen's dataset.

Recall that different measures of inputs, outputs and environmental factors might affect the results, as might adjusting for endogeneity and other statistical concerns. However, it is <u>not</u> the true population relationships that we seek to identify, but the <u>actual relationships</u> in the data being used for the DEA. DEA estimates are based on the extant data's characteristics and relationships, and it is for these relationships that the presence or absence of substitutability must be determined.

Econometricians long ago developed sophisticated methods for measuring elasticity between inputs and between outputs (Fuss 1977; Fuss and McFadden 1978), and such methods may sometimes be desirable. Herein we only illustrate some patterns that might be observed, using artificial datasets in which two inputs and one constant output illustrate either fixed or variable proportion relationships.

If any two inputs (outputs) are substitutes, their relationship must be negative when all remaining DEA inputs and outputs are held constant. *That is, if output is constant for all observations, the relationship between the inputs <u>must</u> be negative if they are <i>substitutes* (Beattie and Taylor 1985). If a relationship between two inputs (outputs) is not negative, then the variables in question have not been substituted for each other, and valid DEA scores cannot be estimated.

Because DEA constructs the production frontier by forming convex combinations of extreme observations, the DEA frontier will <u>always</u> exhibit a negative slope between inputs (outputs), whether the variables are substitutable or nonsubstitutable. Further, negative relationships on the DEA frontier would be present if the DMUs forming it exhibit different degrees of efficiency for each variable. For example, suppose that in truth a fixed ratio of 10 units of Input A and 5 units of Input B are necessary to efficiently produce one unit of output. If one DMU used 11 units of the first input and 5 units of the second, and another DMU used 10 units of the first input and 6 of the second, then both would be on the DEA frontier and would exhibit a negative relationship between the inputs. In short, negative relationships between variables on the frontier provide <u>no</u> evidence whatsoever concerning their substitutability.

It is sometimes assumed that relations among variables on the production frontier differ from their relations within the production possibility set. So, for example, for a fixed amount of output, it sometimes is contended that even if labor and capital are, in truth, substituted for each other on the frontier, they might not exhibit a negative relationship elsewhere in the production space. If this surmise were true, we would find negative relationships between inputs (outputs) if only efficient DMUs are analyzed, but not if inefficient DMUs are analyzed.

Unfortunately, determining whether a frontier technology in truth differs from the technology used by inefficient units is a <u>very</u> difficult statistical problem. As has been observed by Schmidt, "I have tried (and failed) to find a reasonable specification which would allow one to say how features of the technology vary with distance from the frontier" (Schmidt 1985-86, p. 322). Even if a difference does exist in some datasets, assuming that it is present for a particular dataset would be problematic.

In estimating elasticity among inputs or outputs, econometricians, including those employing stochastic frontier analysis, use all observations in the data set, making the assumption that even inefficient units will yield similar relationships among inputs (outputs). Indeed, we asked Schmidt that, if a sample as a whole shows no substitutability between two inputs, would it be reasonable to assume that little or no substitutability between these inputs exists on the frontier. He replied that, at an intuitive level, the answer should be yes (email from Peter Schmidt, 6 June 2010).

Further, it seems to us that even firms inefficiently using substitutable inputs (for example) would adjust the input ratio if the price ratio between the inputs changed

significantly. To believe otherwise would mean that inefficient DMUs completely ignore economics in making decisions, which seems unlikely to the extreme. Further, even if the elasticity of substitution was unequal on the frontier and in the interior of the production space, it again seems unlikely to the extreme that elasticity would change directions, going from negative to either zero or positive, rather than just becoming weaker or stronger in the interior as compared to the frontier.

We illustrate possible outcomes with artificial data for two inputs, assuming other inputs and all outputs have been held constant, so only the relationship between the two inputs remains. For substitutable inputs, Figure A1 shows efficient DMUs, and the resulting piecewise isoquant, which we use in the simulation.

We simulate inefficiencies in using each of the two inputs for 200 DMUs, randomly generating percentage inefficiencies from a half-normal distribution with a mean of 0 and a standard deviation on 1. This means that a third of the inputs will be between 100 and 200 percent of the efficient amount, with about a 1 percent chance that an input will be more than 3 times the efficient amount. It is unlikely that such substantial proportions of DMUs would use inputs so inefficiently, but this will make the resulting graphic results more credible. Further, we used the exact same set of percentage inefficiencies for both the substitutable input scenario and the nonsubstitutable input scenario, so any differences observed would not be related to the magnitude of the inefficiencies involved.

For the substitutable inputs scenario, we randomly generated 200 efficient input combinations on the frontier of the isoquant shown in Figure A1. Then, we multiplied each of the two inputs by one of the randomly-generated inefficiency proportions noted above, with the products indicating a DMU's actual use of the two inputs. The results are shown in Figure A2. Figure A2a illustrates the case when the inefficient percentages are independent, A2b shows the situation when the inefficient percentages are perfectly correlated, and A2c illustrates the case when there is an R-square of 0.42 between the inefficient percentages. Note that it is the inefficiencies only that are correlated, not the aggregated efficient and inefficient amounts.

For the scenario in which the inputs are not substitutable, the sole efficient input combination is chosen to be 0.35 of Input 1 and 0.75 of Input 2, which would be the point efficiency coordinate for the observations in Figure 1A. We multiplied each of the two inputs by one of the randomly-generated inefficiency proportions noted above, with the products indicating a DMU's actual use of the two inputs. The results are shown in

Figure A3. Figure A3a illustrates the case when the inefficiencies are independent, A3b shows the situation when the inefficiencies are perfectly correlated, and A3c illustrates the case when there is an R-squared of 0.42 between the inefficiencies.

Recall that the same proportional inefficiencies are used for both scenarios, the only difference being that efficient points along the entire isoquant were used for substitutable inputs, and the one efficient point was used for nonsubstitutable inputs. Even though the inefficiencies are much greater than would usually be expected, the patterns are clear. In all cases the substitutable inputs reflect strong down-sloping relationships between inefficiencies when the inefficiencies are independent but an increasing strong positive relationship as correlation increases.

On the other hand, in our artificial dataset, even though the inefficiencies are relatively large compared to the efficient input amounts, the length of the frontier is relatively large compared to the inefficiencies. Assuming that in truth the inputs are substitutable, the shorter the relative length of the frontier when compared to the values of the inefficiencies, the less likely it will be to find a statistically significant negative relationship between the inputs.

However, as the frontier gets shorter, the de-facto technology approaches nonsubstitutability. So, DEA might be inappropriate in such cases even if there is a minute amount of substitutability.

Further, if substitutable inputs are <u>not</u> substituted to any great degree in the dataset at hand, then a negative relationship may not be detectable even if present in truth. Indeed, if the range of substitution is so small when compared to the aggregation of noise and random inefficiencies that a negative relationship cannot be detected, then attempts to empirically estimate a frontier with DEA may be futile.

Finally, recall that the purpose of this appendix is only to introduce some of the estimation issues, and present simple scenarios to better illustrate the concepts involved. We certainly do not purport to offer definitive answers to what often are very complex estimation questions!

# References

- Banker RD, Chang H, Janakiraman SN, Konstans C (2004) A balanced scorecard analysis of performance metrics. European Journal of Operational Research 154 (2):423-436
- Banker RD, Morey RC (1986) Efficiency analysis for exogenously fixed inputs and outputs. Operations Research 34 (4):513-521
- Barnum D, Walton S, Shields K, Schumock G (2010a) Measuring hospital efficiency with data envelopment analysis: Nonsubstitutable vs.Substitutable inputs and outputs. Journal of Medical Systems. doi:doi: 10.1007/s10916-009-9416-0
- Barnum DT (2009) Bibliography of urban transit dea publications. Social Science Research Network. doi:http://ssrn.com/abstract=1350583

Barnum DT, Gleason JM, Hemily B (2008) Using panel data analysis to estimate dea confidence intervals adjusted for the environment. Journal of Transportation Engineering 134 (5):215-223

- Barnum DT, Gleason JM, Hemily B, Lin J, Wang P (2010b) Progressing from uncertainty to risk for dea-based decisions. Journal of the Operational Research Society. doi:doi:10.1057/jors.2009.120
- Beattie BR, Taylor CR (1985) The economics of production. vol xvi, 258 p. Wiley, New York
- Casu B, Thanassoulis E (2006) Evaluating cost efficiency in central administrative services in uk universities. Omega 34 (5):417-426
- Charnes A, Cooper WW, Rhodes E (1978) Measuring the efficiency of decision making units. European Journal of Operational Research 2 (6):429-444
- Cherchye L, Moesen W, Rogge N, Puyenbroeck TV, Saisana M, Saltelli A, Liska R, Tarantola S (2007) Creating composite indicators with dea and robustness analysis: The case of the technology achievement index. Journal of the Operational Research Society 59 (2):239-251
- Cooper WW, Deng H, Huang Z, Li SX (2002) Chance constrained programming approaches to technical efficiencies and inefficiencies in stochastic data envelopment analysis. Journal of the Operational Research Society 53 (12):1347-1356
- Cooper WW, Li S, Seiford LM, Zhu J (2004a) Sensitivity analysis in dea. In: Cooper WW, Seiford LM, Zhu J (eds) Handbook on data envelopment analysis. Kluwer Academic Publishers, Boston, pp 75-97
- Cooper WW, Seiford LM, Tone K (2007a) Data envelopment analysis: A comprehensive text with models, applications, references and dea-solver software, vol 2nd ed. vol xxxviii, 490 p. Springer, New York
- Cooper WW, Seiford LM, Tone K, Zhu J (2007b) Some models and measures for evaluating performances with dea: Past accomplishments and future prospects. Journal of Productivity Analysis 28 (3):151-163
- Cooper WW, Seiford LM, Zhu J (2004b) Data envelopment analysis: History, models and interpretations. In: Cooper WW, Seiford LM, Zhu J (eds) Handbook on data envelopment analysis. Kluwer Academic Publishers, Boston, pp 1-40
- Cullmann A, von Hirschhausen C (2008) Efficiency analysis of east european electricity distribution in transition: Legacy of the past? Journal of Productivity Analysis 29 (2):155-167
- Denizer C, Dinc M, Tarimcilar M (2007) Financial liberalization and banking efficiency: Evidence from turkey. Journal of Productivity Analysis 27 (3):177-195
- Färe R, Grosskopf S (2004) New directions: Efficiency and productivity. Studies in productivity and efficiency, vol 174 p. Kluwer Academic Publishers, Boston
- Färe R, Grosskopf S, Lovell CAK (1994) Production frontiers. vol xv, 296 p. Cambridge University Press, Cambridge, England

	1
	2
	3
	4
	5
	б
	7
	8
	9
1	0
1	1
1	2
1	2
1	Δ
1	5
1	5
1	0
1	/
1	8
1	9
2	0
2	1
2	2
2	3
2	4
2	5
2	б
2	7
2	8
2	9
2	0
2	1
2 2	л Т
3	⊿ つ
3	3
3	4
3	5
3	6
3	7
3	8
3	9
4	0
4	1
4	2
4	3
4	4
4	5
4	6
4	7
т Л	, 0
4	0
4	9
5	0
5	Ţ
5	2
5	3
5	4
5	5
5	6
5	7
5	8
5	9
6	0
6	1
6	2
6	3
6	4
0	-

For 1	Society, Series A 120 (3):253-290 XI (1999) Efficiency at different levels of aggregation: Public vs. Private soctor firm
FUX I	Economics Letters 65 (2):173-176
Fried	HO Levell CAK Schmidt SS (2008) The measurement of productive efficiency an
I IICu	productivity growth vol xiv 638 p Oxford University Press Oxford
Front	line S (2009) Premium solver nlatform 9.5 for use with microsoft excel
Fuss	MA (1977) The structure of technology over time: A model for testing the "Putty-cl:
1 455	Hypothesis. Econometrica 45 (8):1797-1821
Fuss	MA. McFadden D (1978) Production economics: A dual approach to theory and
1 400	applications. Contributions to economic analysis, vol 2 v. North-Holland Pub. Co
	Amsterdam
Gleas	son JM. Barnum DT (1982) Toward valid measures of public sector productivity:
	Performance indicators in urban transit. Management Science 28 (4):379-386
Gleas	son JM, Barnum DT (1999) Analyzing proficiency criteria of health technology syste
	The case of drug testing. IEEE Transactions on Engineering Management 46 (3):
	369
Golu	b GH, Van Loan CF (1989) Matrix computations, vol Second. Johns Hopkins Unive
-	Press, Baltimore, MD
Gree	ne WH (2008) Econometric analysis, 6th ed, vol 5. Prentice Hall, Upper Saddle Rive
	NJ
Gröse	che P (2009) Measuring residential energy efficiency improvements with dea. Journa
	Productivity Analysis 31 (2):87-94
Hitch	CJ, McKean RN (1960) The economics of defense in the nuclear age. Harvard
	University Press, Cambridge, MA
Holli	ngsworth B (2008) The measurement of efficiency and productivity of health care
	delivery. Health Economics 17 (10):1107-1128
Holv	ad T, Hougaard JL, Kronborg D, Kvist HK (2004) Measuring inefficiency in the
	norwegian bus industry using multi-directional efficiency analysis. Transportation
	(3):349-369
Houg	gaard JL (2005) A simple approximation of productivity scores of fuzzy production s
	Fuzzy Sets and Systems 152:455-465
Iwasa	aki N, Tremblay V (2009) The effect of marketing regulations on efficiency: Lechate
	versus coordination effects. Journal of Productivity Analysis 32 (1):41-54
Karla	ftis MG (2003) Investigating transit production and performance: A programming
	approach. Transportation Research Part A: Policy and Practice 37 (3):225-240
Karla	attis MG (2004) A dea approach for evaluating the efficiency and effectiveness of ur
	transit systems. European Journal of Operational Research 152 (2):354-364
Karla	iftis MG, McCarthy PS (1997) Subsidy and public transit performance: A factor ana
	approach. Transportation 24 (3):253-270
Karla	iftis MG, McCarthy PS (2002) Cost structures of public transit systems: A panel data
	analysis. Transportation Research Part E: Logistics and Transportation Review 3
17 1	(1):1-18 (1) MO M O (1) DO (1) HO (1000) O (1)
Karla	attis MG, McCarthy PS, Sinha KC (1999) System size and cost structure of transit
т •.	industry. Journal of Transportation Engineering 125 (3):208-215
Litm	an 1 (2004) Transit price elasticities and cross-elasticities. Journal of Public
т	Transportation 7 (2):37-58 $PAD_{2} = PAC_{2} = PAC_{2} = PAC_{2}$
Lorei	nzo JIVIP, Sanchez INIG (2007) Efficiency evaluation in municipal services: An
	application to the street lighting service in spain. Journal of Productivity Analysis
	(3):149-167

1	
2	
3	
4	
5	
о 7	
8	
9	
10	
11	
12	
13	
14	
15	
10 17	
18	
19	
20	
21	
22	
23	
24	
25	
26	
27 20	
20 29	
30	
31	
32	
33	
34	
35	
36	
37	
38 20	
39 40	
41	
42	
43	
44	
45	
46	
47	
48 40	
49 50	
51	
52	
53	
54	
55	
56	
57	
58	
59 60	
61	
62	
63	
64	
65	

Neralic L, Wendell RE (2004) Sensitivity in data envelopment analysis using an approximate inverse matrix. Journal of the Operational Research Society 55 (11):1187-1193

Olesen O, Petersen N (2009) Target and technical efficiency in dea: Controlling for environmental characteristics. Journal of Productivity Analysis 32 (1):27-40

- Ray S (2007) Are some indian banks too large? An examination of size efficiency in indian banking. Journal of Productivity Analysis 27 (1):41-56
- Ruggiero J (2004) Data envelopment analysis with stochastic data. Journal of the Operational Research Society 55 (9):1008-1012
- Schmidt P (1985-86) Frontier production functions. Econometric Reviews 4 (2):289-328
- Simar L, Wilson PW (2000) Statistical inference in nonparametric frontier models: The state of the art. Journal of Productivity Analysis 13 (1):49-78
- Tser-yieth C (2002) A comparison of chance-constrained dea and stochastic frontier analysis: Bank efficiency in taiwan. Journal of the Operational Research Society 53 (5):492-500

National transit database (1997- 2008) United States Federal Transit Administration,. http://www.ntdprogram.gov/ntdprogram/. Accessed 01 May 2010

Woodruff EB, Lammers HB, Lammers TF (2005) Steam plant operation, vol 8th ed. vol xiv, 850 p. McGraw-Hill, New York

- Wooldridge JM (2002) Econometric analysis of cross section and panel data. MIT Press, Cambridge, MA
- Zhou P, Ang BW, Poh KL (2008) A survey of data envelopment analysis in energy and environmental studies. European Journal of Operational Research 189 (1):1-18





3 4 5 6 7 8 23 















Note: VPA frontier is H-B-Q-R, with B and Q being efficient.

FPA frontier is point P, with composite DMU P being efficient.







Figure A1. Variable factor proportion technology, frontier DMUs









Figure A3. Fixed factor proportion technology, inefficient DMUs



	EDA		FPA-	%	FPA	VPA	Rank
DIVIO	гга	VFA	VPA	Difference	rank	rank	Diff
А	1.06	1.04216	0.02	2.03	9	18	-9
В	1.07	0	1.07	100.00	10	1	9
С	0.99	0.96344	0.02	2.20	6	13	-7
D	1.13	0.79697	0.34	29.77	18	9	9
Е	1.02	1.0019	0.02	2.11	8	16	-8
F	1.42	0.80695	0.61	43.24	56	10	46
G	2.19	1.48344	0.71	32.33	196	70	126

Table 1. Selected DMUs assessed by VPA and FPA from sample of 216 DMUs

Note: DMU B is VPA-efficient; the higher the value, the lower the efficiency

## Table 2. Comparisons among CCR, FPCCR, and FPR

Means		Transit Data			Monte Carlo Simulation			
$R^2$ (FPR vs FPCCR)		0.959			0.996			
$R^2$ (FPR vs CCR)		0.612			0.692			
R <sup>2</sup> (FPCCR vs CCR)		0.577			0.670			
	FPR	FPCCR	CCR		FPR	FPCCR	CCR	
Maximum Score	1	1	1		1	1	1	
Number at Maximum	1	1	11 33		1	1	16 29	
Mean Minimum Score	0.523	0.494	0.623		0.546	0.546	0.614	

Mean values from the six trials of real-world transit data and 1000 trials of simulation data, with 66 observations per trial. The FPR and FPCCR minimum values do not differ from each other to a statistically significant degree. The CCR minimums differ to a statistically significant degree from both the FPR and FPCCR minimums for both the transit and Monte Carlo datasets.

Means	Transit		Simulation	
wieans	FPR	FPCCR	FPR	FPCCR
1	1	1	1	1
first decile (2-6)	4.00	3.83	4.49	4.42
second decile (7-13)	2.17	2.67	4.40	4.25
third decile (14-19)	1.67	1.33	2.50	2.47
fourth decile (20-26)	1.17	1.33	1.83	1.86
fifth decile (27-33)	0.50	0.33	1.16	1.23
sixth decile (34-39)	0.50	0.00	0.52	0.54
seventh decile (40-46)	0.00	0.00	0.27	0.33
eighth decile (47-52)	0.00	0.00	0.08	0.12
ninth decile (53-59)	0.17	0.17	0.06	0.05
tenth decile (60-66)	0.17	0.67	0.01	0.02
Total Efficient CCR Observations	11.33	11.33	16.30	16.29
Proportion in Bottom Half	0.07	0.07	0.06	0.07
Proportion Below first decile	0.56	0.57	0.66	0.67

Table 3. Ranks of FPR and FPCCR when CCR = 1 (DMU reported efficient)

Table 4 Absolute differences in ranks of CCR vs. FPR and FPCCR

Maana	Transit	Simulation
Means	FPR FPCCR	FPR FPCCR
Same Rank (zero difference)	2.83 2.00	4.43 4.42
first decile (1-6 ranks different)	24.50 26.83	27.39 26.44
second decile (7-13 ranks different)	24.00 24.83	18.57 18.32
third decile (14-19 ranks different)	9.33 7.50	8.53 8.96
fourth decile (20-26 ranks different)	2.33 2.17	4.54 4.84
fifth decile (27-33 ranks different)	1.50 0.83	1.61 1.92
sixth decile (34-39 ranks different)	0.83 0.67	0.56 0.65
seventh decile (40-46 ranks different)	0.17 0.17	0.24 0.27
eighth decile (47-52 ranks different)	0.17 0.17	0.07 0.12
ninth decile (53-59 ranks different)	0.17 0.33	0.05 0.05
tenth decile (60-66 ranks different)	0.17 0.50	0.01 0.01
Total observations	66.00 66.00	66.00 66.00
Proportion differing by more than 33 ranks	0.02 0.028	0.01 0.017
Proportion differing by more than 6 ranks	0.59 0.563	0.52 0.532

# Authors' Response to the Associate Editor

I feel that the paper is a bit too long and in reading it I feel that it would be possible to reduce the paper without any loss on its content.

- In order to shorten the paper the authors may . . . consider . . . reducing to the minimum the sections where one input one output situations are considered, since in this setting there are no issues of inputs or outputs substitutability. . .
- Re-structuring of the paper is also likely to reduce some pages.
- The appendix is rather long and should be reduced.

We have applied your three suggestions, and made additional deletions as well, which cut the entire paper by two pages (about 6 percent of text pages). Due to additional material we added to the appendix in response to your last suggestion, the appendix's net length did not change much.

Although we would be happy to consider any additional deletions suggested, we couldn't see how to do it without weakening the sections where the cuts were made.

However, there are several sections that are less central to the paper's focus than the others. These are sections 6.1 and 6.2, and the appendix. Although we do think these three add important information, and indeed were written at the suggestion of referees, they seem somewhat less central to the paper. We would prefer that they remain, but, if more shortening is absolutely necessary, we think it would be preferable to cut an entire section rather than trying to cut a little bit from all sections. We will be happy to follow your counsel, however, in whatever you propose.

In addition, I also think that the paper follows a structure that is not usual in JPA papers. In particular it has 11 sections and the standard is about half of it. So I would like to ask the authors to organize their paper in 6 or 7 sessions and consider some subsections for these main sections. For example, sections 2 and 3 do not really need to be two sections - they can be one where the authors introduce fixed proportion technologies and then provide an illustrative example.

We have done this, and now there are 7 sections.

When additive (or similar) models are used the authors call the resulting score 'efficiency', but it is in fact an inefficiency score. Consider re-phrasing it in order to clarify matters.

We have done so. Section 3 deals with the additive models, and, when reporting scores, we now call them inefficiency scores, hopefully clarifying matters.

In page 14 the authors note that the efficiency score of a unit can be obtained as the sum of observed outputs divided by the sum of target outputs. Considering y - observed output and y\*- target output, this happens because  $yi/yi^*$ =theta for both outputs. Therefore  $(y1+y2)/(y1^*+y2^*) = (\text{theta } y1^*+\text{theta } y2^*)/(y1^*+y2^*) = \text{theta}.$  Note that if there were slacks or if the assessment was under VRS this would not be the case.

Cool! We have added your algebraic proof to the paper on what is now page 13:

Our JPA Associate Editor demonstrated the same conclusion algebraically, assuming constant returns to scale and no slacks (which are implicitly present geometrically): Let y1 and y2 be the coordinates of the observed outputs, and y1\* and y2\* be the coordinates of the target outputs (which are on the frontier). Because the observed and target coordinates are both on a ray through the origin,  $y1/y1^* = y2/y2^* = \theta$ . Therefore  $(y1+y2)/(y1^*+y2^*) = (\theta y1^*+\theta y2^*)/(y1^*+y2^*) = \theta$ 

Note that in the situation depicted in Fig A1, where the inputs are substitutable, the authors consider a wide range of input combinations which may not happen in practice. That is, if the substitutability range implied that all units were concentrated above the line linking points (1,2) and (3,1) eventually the pattern of negative correlations would be more difficult to obtain. Please consider briefly discussing this issue.

Excellent point! Here is what we have added as the last four paragraphs in the appendix:

On the other hand, in our artificial dataset, even though the inefficiencies are relatively large compared to the efficient input amounts, the length of the frontier is relatively large compared to the inefficiencies. Assuming that in truth the inputs are substitutable, the shorter the relative length of the frontier when compared to the values of the inefficiencies, the less likely it will be to find a statistically significant negative relationship between the inputs.

However, as the frontier gets shorter, the de-facto technology approaches nonsubstitutability. So, DEA might be inappropriate in such cases even if there is a minute amount of substitutability.

Further, if substitutable inputs are <u>not</u> substituted to any great degree in the dataset at hand, then a negative relationship may not be detectable even if present in truth. Indeed, if the range of substitution is so small when compared to the aggregation of noise and random inefficiencies that a negative relationship cannot be detected, then attempts to empirically estimate a frontier with DEA may be futile.

Finally, recall that the purpose of this appendix is only to introduce some of the estimation issues, and present simple scenarios to better illustrate the concepts involved. We certainly do not purport to offer definitive answers to what often are very complex estimation questions!