MISSISSIPPI STATE UNIVERSITY

Department of Agricultural Economics

Working Paper

January 15 – 1 | January 2015

Is Efficiency Analysis all there is with Data Envelopment Analysis?

Bryon J. Parman (corresponding author)
Mississippi State University
Parman@agecon.msstate.edu

Allen M. Featherstone
Kansas State University
afeather@agecon.ksu.edu

Brian K. Coffey
Keiin Institute of Business and Economics

Department of Agricultural Economics
Mississippi State University
Box 5187
Mississippi State, MS 39762
Phone: (662) 325-2049
Fax: (662) 325-8777
www.agecon.msstate.edu

JEL Classification C13; C14; D20

Keywords: Cost function, Efficiency Analysis, Multi-Product Economies of Scale, Product Specific Economies of Scale, Nonparametric estimation

Abstract: Nonparametric cost frontier estimation has been commonly used to examine the relative efficiency of firms without critically examining the shape of the cost frontier. To examine the shape of the cost frontier has required additional estimation using parametric methods to recover potential cost savings from multi-product and product-specific economies of scale. This paper develops and tests a method for estimating multi-product and product-specific economies of scale using the nonparametric approach by evaluating the difference between scale calculations from an assumed cost frontier and those estimated using data envelopment analysis. The results demonstrated that the nonparametric approach is able to accurately estimate multi-product economies of scale and product-specific economies of scale under alternative inefficiency distributional assumptions.

Is Efficiency Analysis all there is with Data Envelopment Analysis?

1. Introduction

Efficiency analysis has been a common technique used to examine or explain relative costs among alternative economic agents (Färe, Groskopf, and Lovell 1985, Cooper, Seiford, and Tone 2007). Typically, a production frontier is measured either parametrically using the stochastic frontier approach (SFA) (Aigner, Lovell, and Schmidt 1977), or nonparametrically using the Data Envelopment Analysis (DEA) approach and analysis proceeds by estimating how far individual economics agents are off that frontier. Much analysis (Featherstone et al. 1997, Hoff 2007) then uses a second stage to examine whether correlations exist between measures of efficiency and economic agent characteristics. DEA and SFA methods in particular have not traditionally examined the behavior of the frontier production or cost function. However, the theory of the firm provides the potential for exploring the structure of cost. Thus estimates of frontier functions, and the distance firms are from the frontier should provide insights into how firms with similar technological access and marketing achieve different levels of production efficiency and average costs. These methods allow firms operating off the frontier to understand the potential disadvantages due to sub-optimal output and input bundling choices and the effects on firm performance. In addition, these estimates should provide insight into ultimate industry structure.

Multi-product and product-specific economies of scale, and economies of scope are traditionally estimated parametrically using two-sided error systems though specification of a cost function and estimation of parameters (Christenson et. al 1973). The traditional approach is limited from the assumed error structure used in the estimation of a cost frontier function since negative errors imply that some firms are actually producing at a lower cost or higher quantities than the frontier that was being estimated which is not consistent with the economic definition of a cost function (Farrell 1957).

Another potential issue occurs with the estimation of the indirect cost function[1]. To accurately trace out the technology that ultimately determines the shape of the cost frontier, relative price variability becomes important. Lusk et al. (2002) examined the relative variability needed in the estimation of dual cost functions using a Monte Carlo approach. They found that the relative variability necessary to accurately estimate a dual cost function requires more than 20 years of data based on observed data. Thus, dual cost functions may have difficulty recovering the underlying technology. In addition, Featherstone and Moss (1994) note that parametric frontier estimations may often violate the necessary curvature conditions needed for the indirect cost function to exist. Therefore, the lack of ability to accurately measure the underlying technology given data availability and the estimated frontier not being consistent with the necessary cost function conditions are issues that may affect economic analysis and the ultimate recommendations that are based upon that analysis.

The nonparametric or data envelopment analysis approach constructs a frontier from a series of line segments using a linear cost minimization program (Färe, Groskopf, and Lovell 1985). With this method, it is not necessary to restrict the production technology by imposing a functional form and the frontier conforms to economic theory because curvature restrictions on the production/cost function are imposed in the estimation process using inequality constraints. Further, the nonparametric method of Färe, Grosskopf, and Lovell may allow technology to be

[1] An indirect function is a function where the choice variable(s) has been optimized and is held constant while parameter change is evaluated.

measured using a single year's data; thus, reducing the need of relative price variability to accurately measure technology using the dual approach.

Numerous studies have used nonparametric methods to analyze efficiency in various industries including Banker and Maindiratta (1988), Jaforullah and Whiteman (1998), and Chavas and Cox (1988). In these studies, several types of efficiency are estimated to determine if a firm is producing on the production or cost frontier, whether the firm is optimally allocating inputs, or if the firm is operating at the most efficient size. Chavas and Aliber (1993) measure scope economies to determine cost savings from production portfolio diversification in the nonparametric framework while Cummins et al. (2010) use a DEA cost frontier to determine if efficiency measures are related to portfolio diversification.

Other studies have proposed methods for statistical inference and hypothesis testing from nonparametric methods. Cazals et al. (2002) describe a method for robust nonparametric estimation offering a technique for the treatment of outliers. Florens and Simar (2005) follow Cazals et al. with a two-step estimation process describing a method for parametric approximation from a nonparametric frontier. However the Cazals et al. method does not completely envelope the data as outliers remain below the cost frontier, and still require a second step to recover the parameters to explore the structure of that frontier.

Traditional measures of multi-product scale and product-specific scale measures have not yet been developed in the nonparametric DEA framework for the indirect cost function. For example, Paul et. al. (2004) and Kumar and Gulati (2008) use the DEA method to estimate scale efficiency which takes on values of less than, equal to, or greater than one giving an indication of returns to scale. This measure follows from Ray (1998) and Cooper et. al. (2007) where the DEA method is estimated assuming constant returns to scale, and then again assuming variable returns to scale and takes the ratio of the two measures. Paul et al. explain that the interpretation of scale efficiency is not as strait forward as a traditional scale economy measure explained by Baumol et al. (1984). Specifically, they note that these measures only indicate if average per-unit costs are increasing, decreasing, or constant, but not necessarily the magnitude of cost savings from scaling. In both Paul et al., and Kumar and Gulati, it was necessary to perform a parametric estimation to recover traditional estimations of economies of scale, and compare the results to their DEA estimation. Førsund and Hjalmarsson (2004) do calculate scale elasticities, however their method is based on a DEA production frontier estimation and not by using the nonparametric dual approach. Further, techniques for estimating product-specific economies of scale have not been reported for the nonparametric method.

This research derives, operationalizes, and tests estimation techniques for multi-product and product-specific economies of scale for the nonparametric cost frontier estimation method. Specifically, this research develops a multi-product and product-specific scale measure using the definition of Baumol et al. from nonparametrically estimated marginal costs, incremental costs, and output quantities. The estimated measures are then compared to an assumed known cost frontier. From this comparison, it is possible to assess the accuracy of the nonparametric approach estimates in measuring the appropriate economic measures.

Previous research that estimated economies of scope with the nonparametric approach has dropped one or more of the output constraints when estimating the cost of producing a single output (Chavas and Aliber). This research compares the Chavas and Aliber procedure by comparing a method that requires that output to be zero as required in Baumol et al's. definition of an incremental cost. The principle advantage to forcing the output to zero rather than dropping it is that it should more closely measure the theoretically defined incremental cost of each output, where dropping the appropriate constraint often allows a small amount of that output to be produced. In addition, a modification to the Chavas and Aliber technique is proposed when dropping constraints to estimate scope economies. Finally, we evaluate the nonparametric approach under alternative efficiency distributions to investigate the robustness of the results.

2. Theory

Typical economic measures calculated from a cost function include economies of scale and economies of scope. Measures of scale economies include both multi-product economies of scale (MPSE), and product-specific scale economies (PSE) differing only in that MPSE refers to changes in cost relative to more than one output in a multi-output firm, while PSE refers to proportionate changes in cost relative to a single output (Baumol et al.). Mathematically these measures are defined as follows where $C(Y)$ represents the cost of production with $\partial C(Y)/\partial Y_p$ representing the marginal cost of the p^{th} output.

$$MPSE = \frac{C(Y)}{\sum_p Y_p \left[\frac{\partial C(\mathrm{Y})}{\partial Y_p} \right]} \tag{1}$$

To calculate PSE, the average incremental cost (AIC_p) of producing p must be calculated where the incremental cost (IC) for the p^{th} output is defined as:

$$IC_p = C_p - \sum_j C_{j \neq p} \forall j \tag{2}$$

Thus,

$$AIC_p = \frac{IC_p}{y_p} \tag{3}$$

Product-specific economies of scale are the ratio of the average incremental cost of output *p* and the marginal cost of the p^{th} output.

$$PSE_p = \frac{AIC_p}{\dfrac{\partial C(\mathrm{Y})}{\partial Y_p}} \tag{4}$$

Estimates of economies of scope (*SC*) represent the cost savings of producing multiple outputs within a single firm versus producing outputs individually. Economies of scope may be expressed in the following manner where *C(Y)* is total production cost, $C(Y_T)$ is the cost of producing output Y_T, and $C(Y_{N-T})$ represents the cost of producing the remaining outputs where $Y_{N-T} = (Y_1,\ldots Y_{k-1},0\ldots0)$.

$$SC = \frac{[C(Y_T) + C(Y_{N-T}) - C(Y)]}{C(\mathrm{Y})} \tag{5}$$

Measures of multi-product economies of scale, product-specific economies of scale and economies of scope are related. The relationship between multi-product scale economies (MPSE), product-specific scale economies (PSE), and economies of scope (*SC)* can be determined by defining:

$$\alpha_i = \frac{Y_i\left[\dfrac{\partial C(Y)}{\partial Y_i}\right]}{\sum_{i=1}^{N} Y_i\left[\dfrac{\partial C(Y)}{\partial Y_i}\right]} \tag{6}$$

where α_i is the weight placed on the PSE of interest based upon its relative contribution to total output. Thus:

$$MPSE = \frac{\alpha_i PSE_i(Y) + (1-\alpha_i) PSE_{N-i}(Y)}{1 - SC(Y)} \tag{7}$$

MPSE can take one of three values: decreasing, constant or increasing returns to scale. Equation 7 examines the relationship among factors affecting MPSE. If *SC* is zero and the numerator is less than 1, equal to 1 or greater than 1, then there are decreasing, constant and increasing returns to scale. If *SC* is greater than zero and the PSEs are at constant returns to scale, MPSE is in a region of increasing returns (>1).

3. Data and Methods

3.1 The Nonparametric Method

To derive the scale measures nonparametrically, the cost (C_i) is determined for each firm following Färe, Grosskopf, and Lovell where costs are minimized for a given vector of input prices (w_i) and outputs (y_i) with the choice being the optimal input bundle (x_i^*).

$$\begin{aligned} &\min Ci = w_i' x_i^* \\ &s.t \\ &Xz \leq x_i^* \\ &y' z \geq y_i \\ &z_1 + z_2 + \ldots + z_n = 1 \\ &z_i \in \mathbb{R}^+ \end{aligned} \tag{8}$$

where there are "n" producers. The vector Z represents the weight of a particular firm with the sum of Z_i*'s* equal to 1 for variable returns to scale. From the above model, the costs and output quantities can be estimated. The output quantities (y_i) constrain the cost minimizing input bundle to be at or below that observed in the data. Total cost from the model (C_i) is the solution to the cost minimization problem including the production of all outputs for the i^{th} firm. The cost of producing all outputs except one ($C_{i,all-p}$) where p is the dropped output and is determined by either forcing one of the outputs to equal zero or by dropping one of the p^{th} output constraints.

To calculate multi-product scale measures, marginal costs must be determined. The marginal costs ($MC_{i,p}$) for the p^{th} output are obtained from the shadow prices on the output constraints on the base model (equation 8). The calculation of multi-product economies of scale (MPSE) uses the total cost of producing all outputs ($C_{i,all}$), the marginal costs ($MC_{i,p}$), and the output levels produced ($Y_{i,p}$) (equation 1). Using the nonparametric approach, there is an issue with the nonparametric marginal cost because the linear structure results in "Kink Points" on the frontier that results in non-unique marginal costs. Thus, the marginal costs for efficient firms may not be unique. In practice this results in a relatively small number of firms. In addition, a range of estimates of marginal costs at a kink can be calculated.

Product specific economies of scale (PSE) require the calculation of incremental costs ($IC_{i,p}$) which are the cost of producing all outputs minus the sum of the costs of all individual outputs except output (p) for firm i (equation 2). The Chavas and Aliber method to calculate incremental costs using the nonparametric method is to drop one or more of the output constraints from equation 8 to determine the cost of producing the output alone. For example, if a firm produces four different products, four different linear programs would be estimated excluding one of the outputs at a time. We examine the results from dropping one of the output constraints are compared with constraining the appropriate output to zero.

Using equation 2, average incremental costs ($AIC_{i,p}$) are determined by dividing incremental costs by individual output as shown in equation 3. From the average incremental cost (equation 3) and the marginal cost calculations from the shadow prices, it is possible to calculate PSEs (equation 4) where PSEs are interpreted similar to MPSEs except that PSEs pertain to only one output.

The calculation for scope economies (SC_i) follows from equation 5 where $C_{i,p}$ is the cost of producing output *p* for firm *i,* and $C_{i,all}$ is the cost of joint production of all outputs for firm *i*. This measure identifies the potential for cost savings through product diversification. Generally, $SC_i > 0$ implies that scope economies exist and average per-unit costs are reduced with diversification. A scope measure of 0.5 implies that jointly producing multiple outputs in a two good case would reduce costs of producing these outputs by 50% compared to producing them individually.

Cost efficiency (*CE*) identifies a firm's proximity to the cost frontier for a given input/output bundle. It is the quotient of the estimated frontier cost (C_i) and the actual total cost (*ATC*) the firm incurred producing their output bundle.

$$CE_i = \left[\frac{C_i}{ATC_i} \right] \tag{9}$$

This measure must be greater than 0 but less than or equal to 1. A cost efficiency of 1 implies that the firm is operating on the frontier at the lowest possible cost for a given output bundle. However, a cost efficiency less than 1 implies that cost can be reduced by altering the input bundle.

This section has derived and operationalized the measure of marginal costs and incremental costs necessary for the measurement of multi-product and product-specific scale economies. The next section examines the methods used to compare the accuracy of the nonparametric measures with those from a "true" cost frontier.

3.2 Data Simulation

The data for the analysis were generated utilizing an economic data generation procedure[2] found in Gao and Featherstone (2008) run on the SHAZAM software platform (SHAZAM Analytics Ltd.). A normalized quadratic cost function involving 3 inputs (x_1, x_2, x_3) with corresponding prices (w_1, w_2, w_3), and 2 outputs (y_1, y_2) with corresponding prices (p_1, p_2) was used. The normalized quadratic cost/profit function is used because it is a self-dual cost function and a flexible functional form (Lusk et al., Lau 1976). The input and output prices (w_i, p_i) are randomly generated following a normal distribution. Assumed distributions for the output prices and input prices provide observed prices strictly greater than zero with different means and standard deviations to ensure some variability in input/output quantity demands and relative prices. They are:

$$\begin{aligned} &w_1 \sim N\,(9, 0.99) \\ &w_2 \sim N\,(18, 1.98) \\ &w_3 \sim N\,(7, 0.77) \\ &p_1 \sim N\,(325, 99) \\ &p_2 \sim N\,(800, 99)^3 \end{aligned} \tag{10}$$

The input price variability was set proportionate to its mean while the output prices have different relative variability to represent products in markets with different volatilities.

The outputs (y_i) and inputs (x_j) are determined as a function of input and output prices using an assumed underlying production technology. All prices are normalized on w_3 and the cost function is divided by w_3 to impose homogeneity. To ensure the curvature condition is met, the "true" cost function is assumed to be concave in input prices and convex in output quantities.

---

[2] The procedure randomly generates prices for inputs and outputs and then outputs are chosen based upon the production technology using the assumption of profit maximization.

[3] The distributions are arbitrarily chosen.

The eigenvalues are calculated for the "*b*" (price) and "*c*" (output) matrices where the eigenvalues for "*b*" should be negative (concave in prices) and "*c*" values should be positive (convex in outputs). The assumed parameters are set to satisfy the following theory based condition: $b_{ij}=b_{ji}$ (symmetry in input prices). The assumed parameters (Table 1) are used to generate the output quantities y_1 and y_2[4]. The general form of the normalized quadratic cost function is:

$$
\begin{aligned}
&C(W,Y) = b_0 + \begin{bmatrix} b_1 & b_2 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} + \begin{bmatrix} a_1 & a_2 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} + \\
&\frac{1}{2}\left\{ \begin{bmatrix} w_1 & w_2 \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} + \begin{bmatrix} y_1 & y_2 \end{bmatrix} \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \right\} + \begin{bmatrix} w_1 & w_2 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}
\end{aligned} \tag{11}
$$

Output quantities (shown below) are calculated using the assumed parameters of the cost function (Table 1) and the random prices defined in equation 10.

$$
\begin{aligned}
y_1 &= \frac{c_{22}p_1 - c_{12}p_2 + \left(a_{12}c_{12} - a_{11}c_{22}\right)w_1 + \left(a_{22}c_{12} - a_{21}c_{22}\right)w_2 + \left(a_2c_{12} - a_1c_{22}\right)}{\left(c_{22}c_{11} - c_{12}c_{12}\right)} \\
y_2 &= \frac{c_{12}p_1 - c_{11}p_2 + \left(a_{12}c_{11} - a_{11}c_{12}\right)w_1 + \left(a_{22}c_{11} - a_{21}c_{12}\right)w_2 + \left(a_2c_{11} - a_1c_{12}\right)}{-\left(c_{22}c_{11} - c_{12}c_{12}\right)}
\end{aligned} \tag{12}
$$

Using the above cost function (equation 11), a positive random cost deviation term is added to the cost function following a half-normal distribution that alters the cost efficiency where the absolute value of e is distributed $e \sim N\ (0,1000)$[5]. The inclusion of this term adds cost inefficiencies in the data such that firms are off the frontier effectively increasing their cost while keeping the output quantities the same. The level of inefficiency is half-normally distributed.

---

[4] The analysis also was completed for alternative assumptions on price distributions.

[5] The analysis also examined alternative normal standard deviations.

An additional data set[6] is generated assuming a uniform distribution. The uniform deviation ranged from zero to 900. The normal distribution standard deviation of 1000 generates a mean and standard deviation for cost efficiency roughly equivalent to a uniform distribution with a range from zero to 900.

From equation (11), and using Shephard's Lemma (Shephard 1970) where (∂C(W,Y)/∂w$_i$)=x_i, the factor demands for inputs x_1 and x_2 are recovered. Factor demand for x_3 is found by subtracting the product of quantities and prices for x_2 and x_3 from the total cost (equation 13).

$$\begin{aligned} x_1 &= b_1 + b_{11}w_1 + b_{12}w_2 + a_{11}y_1 + a_{12}y_2 \\ x_2 &= b_2 + b_{12}w_1 + b_{22}w_2 + a_{21}y_1 + a_{22}y_2 \\ x_3 &= C(W,Y) - x_1w_1 - x_2w_2 \end{aligned} \tag{13}$$

The input quantities (x_i's) are then adjusted (x_i^a) by the cost efficiency (CE) effectively increasing the input demands proportionate to the costs generated for each firm.

$$x_i^a = \frac{x_i}{CE} \tag{14}$$

Using the above method, 400 observations were generated where firms produce a combination of both outputs. Fifty firms were generated producing only y_1 with another 50 firms producing only y_2 which is accomplished by restricting either y_1 or y_2 to equal zero and re-running the simulation for 50 separate observations each. Thus, a total of 500 observations were simulated with descriptive statistics shown in Table 2. In Table 2, x_i^n represents inefficient input quantities for the normal error distribution and x_i^u represent the inefficient input quantities for the uniform distribution. The summary statistics for the multi-product scale, product-specific scale, scope, and cost efficiencies for each data point from the "true" cost function are shown in Table

[6] The analysis was run using 2500 observations with little difference in the results.

3. Summary statistics on scale and scope are independent of the distribution of cost "inefficiency". Figures 1 through 4 provide a visual representation of the multi-product scale and scope economies as well as cost efficiencies and product-specific scale economies calculated from the "true" cost function. These calculations are used to examine the accuracy of the proposed nonparametric approach.

Economies of scale and economies of scope calculations are made at efficient data points projected to the DEA-estimated cost frontier. While the cost efficiency for each firm is simulated under a uniform and a half-normal distribution (Figure 2), the MPSE, PSE's, and Economies of Scope are identical for each data point (Table 3) for the "true" cost function. This occurs because the input prices (w_i*'s*) and output prices (p_i*'s*) remain unchanged and thus, the output quantities (y_i*'s*) remain unchanged (Equation 14). The input quantities (x_i*'s*) are different in that the deviation in input quantity is uniformly distributed. The uniformly distributed data more evenly distribute the quantity of firms at each relative distance from the frontier, rather than many firms being clustered around the mean distance as in the half-normal case.

The difference between the "true" and the nonparametric approach is evaluated by subtracting each nonparametric calculation from the "true" measure calculated with economic simulation. Since the approximation of the "true" measure is key, the statistics reported hereafter are the difference between the "true" measures and what was estimated nonparametrically. Using this approach, any possible bias from the DEA approach can be determined. A positive number implies that the nonparametric approach underestimates the measure being evaluated and conversely, a negative difference indicates the nonparametric method overestimates the measure. The mean absolute deviation is also reported for all three models allowing for the comparison of average absolute deviation from zero

Cumulative density functions are presented for the differences between the true measures and the estimated measures to produce visual representation of both bias and deviation. If there is no difference between the estimated measure and the true measure, the cumulative density function is a vertical line at zero (see Figure 10 for the No Inefficiencies model).

4. Results

Three comparisons were conducted using the half-normal distribution for cost inefficiency, and three identical comparisons using the uniform distribution for cost inefficiency. The first comparison for both distributions uses the economic simulation data with only cost inefficiencies in the cost function (No Inefficiency). The purpose of this simulation is to ensure the model is estimating the measures correctly, and to examine the nonparametric procedure estimates of scale and scope when all firms are efficient in input quantities. The second and third comparisons for both distributions involve introducing technical inefficiencies into the input quantities (equation 16), and are more consistent with observed data. Since efficient firms have a cost efficiency of 1, and less cost efficient firms have a cost efficiency between 0 and 1, an efficient firm uses optimal input quantities. However firms may use additional inputs to produce output if the firm is not efficient. Inputs x_1, x_2, and x_3 are adjusted upwards by the proportionate cost inefficiency to reflect this.

The second nonparametric comparison for the half-normal and uniform distributions assume the appropriate constraints are dropped (Dropped) for the estimation of incremental costs. The third simulation forces the appropriate output to be zero (Constrained). The estimation was done using the General Algebraic Modeling System (GAMS Development Corporation, Washington D.C.).

Twenty-four frontier points are identified from the nonparametric estimates for the half-normal distribution and twenty-five using the uniform data. For each distribution, the firms found on the frontier were the same for the Dropped Model and the Constrained Model. These points have non-unique marginal cost estimates. Due to the non-uniqueness of the marginal costs from these observations, MPSEs were not reported. For single output observations, PSEs cannot be calculated for the output not being produced. Economies of scope are also not reported for single output observations.

4.1 Multi-product Economies of Scale

The differences results for MPSE are found in Table 4 and Figure 5. The No Inefficiencies model for both distributions shows little difference from the actual frontier function (Figure 5). The average bias was close zero for both distributions with a standard deviation of 0.11 in the half-normal case and a standard deviation of 0.023 for the uniform case (Table 4). The mean absolute deviation was nearly zero as well. This result indicates that since MPSE is a function of total costs, marginal costs, and output levels, the marginal costs are estimated closely to the "true" marginal costs.

For the two models estimated where technical inefficiencies were introduced, with half-normal distribution nearly 85% of the MPSE difference calculations were within 0.1 in absolute value to the "true MPSE" (Figure 5). The standard deviations for both the Constrained and Dropped models were small (Table 4) and the mean absolute deviation was less than 0.05 for both models. For the uniform distribution, the average for both models was nearly zero with the Constrained model being slightly closer to zero than the Dropped Model in terms of bias but the mean absolute for the Constrained model was 0.04 higher than the Dropped model. The standard deviations for both models was approximately 0.05. When comparing the distributions, the models with the uniform distribution estimated the MPSE's closer to the "true MPSE's" for each observation with greater than 99% within between -0.1 and 0.1 and mean absolute deviations less than for the half-normal distribution.

The nonparametric approach showed a very close proximity to the calculations from the frontier function with respect to the MPSE. The model with outputs constrained to zero results in slightly more accurate estimate of MPSE compared to those estimates dropping a constraint.

4.2 Product-specific Economies of Scale

Product specific scale economies estimated from the No Inefficiencies model showed slight differences from that of the actual frontier function for both distributions (Table 5, and Figures 6 and 7). The averages and mean absolute deviations for both PSE_{y1} and PSE_{y2} were nearly zero and the standard deviations were also low. This result concurs with the results from the other measures where deviations from the frontier function were small. Though the averages were nearly zero for both distributions, the bias for both PSEs y_1 and y_2 were negative in the half-normal case showing that the nonparametric approach slightly overestimated PSE while the average difference in the uniform case was positive showing a slight underestimation of the PSEs.

The differences for the estimates with technical inefficiencies in the input quantities were highest for the PSE estimates compared to the other measures. For the half-normal distribution, the average of PSE_1 for both the Constrained and Dropped models was about 0.13 showing negative bias with standard deviations and mean absolute deviations of approximately 0.22. For PSE_2 the average was much closer to zero at approximately 0.03 for the Dropped model and 0.02 for the Constrained model with standard deviations for both around 0.11. The mean absolute deviations were also lower with both being around 0.085. The direction of bias was negative in that the models with technical inefficiency overestimated the PSE estimates.

The estimations for PSE_1 and PSE_2 were closer to the "true PSE's" for both models with the uniform distribution having lower standard deviations, and averages closer to zero. The average PSE_1 for the Dropped model was nearly zero in the uniform case with a standard deviation of 0.106, while the average for the Constrained model was 0.05 and a standard deviation of 0.19 (Table 5). Mean absolute deviations for the uniform distribution were also closer to zero however the dropped model's mean absolute was nearly halved while the constrained model changed by only 0.01. Like the half-normal case, the differences for the uniform distribution were positive on average indicating that both models slightly underestimated the PSE's. The differences for PSE estimates are also evident in Figures 6 and 7. PSE measures are relatively less accurate than the measurement of MPSE.

The cause of the higher error in the PSE estimates in both the half-normal case and the uniform case occurs due to variations in the incremental cost calculations. The total costs estimated by the nonparametric methods were nearly the "true" costs, as were output quantities with only slight variations in marginal cost. Thus, the MPSE differences were small. Product-specific economies are calculated using total cost and incremental costs. Incremental costs exhibit some, albeit small variation.

The concern with the incremental cost was hypothesized to be due to missing frontier observations with zero quantities. This results in the frontier estimation for regions with missing data to shift upward for an inefficient firm reflecting that a firm is on the frontier when it is not. This conclusion is apparent in that the No Inefficiency model under both distributions which has no inefficiency shows less deviation from the frontier function than the two models with cost inefficiency.

To examine the importance of the single output firms, the 24 efficient observations from the models with the half-normal distribution were set to be efficient. This puts them on the true frontier and the model is re-estimated with the remaining observations unchanged. Table 6 and Figure 8 show the results for PSE_1 which had the largest deviation for both distributions. The standard deviations for both models decreased from 0.221 to approximately 0.169 and 0.164 while the averages were reduced from 0.133 to 0.065 and 0.069 respectively. The mean absolute deviations also fell for both models indicating a closer proximity to the true PSE values than in the initial estimation of PSE_1. Thus, obtaining correct measures of the frontier for zero output observations is important to improving the accuracy of PSEs.

4.3 Economies of Scope

The distribution of the difference for scope between the frontier function and the nonparametric estimates for both distributions are shown in Figure 9**.** For the half-normal distribution, differences in scope for the No Inefficiency model were very small yielding a standard deviation of about 0.017 and an average and mean deviation close to zero. For the uniform distribution, the differences were small as well with a standard deviation of 0.020 and mean absolute deviation nearly zero (Table 7). The implication is that in the absence of input inefficiency, the individual cost estimates from the nonparametric method are close to that of the actual frontier function.

The estimates for models where inefficiencies were introduced were also very close to that of the frontier function for both distributions. For the half-normal case, the standard deviation for the Constrained, and Dropped models were small. The average differences were both less than 0.1 in absolute, value as were the mean absolute deviations (Table 7). The estimates were identical for both models. This indicates that the calculations for costs for producing zero output ($C_{i,all\text{-}p}$) are the same. Thus, both approaches including dropping a constraint, or constraining an output to zero appear to do equally well estimating economies of scope.

In the uniform distribution case, both models with technical inefficiency did not estimate an identical scope. The absolute values for average and standard deviation for the Dropped model was -0.017 and 0.066 respectively for the while the absolute values for average and standard deviation for the Constrained model was -0.028 and 0.095, respectively. Also, the mean absolute deviation was more than twice as high for the Constrained model than the Dropped model. Thus, under a uniform distribution, dropping the appropriate constraint reduces the mean absolute deviation of economies of scope more than constraining the appropriate output to zero.

Figure 9 shows that most of the differences in scope estimates from the models with technical inefficiency for both distributions are negative. This implies that the economies of scope measures for the Constrained and Dropped models slightly over estimate economies of scope. The average scope difference with inefficiency is less than 0.1 in absolute value and over 70% of the differences are within this proximity range to the “true” scope measure in the half-normal case for both models. In the uniform case, the models with technical inefficiency have an average difference of nearly zero with all but five observations within 0.03 of the true scope calculation in absolute value. The results demonstrate small differences between the economies of scope estimates between the Dropped model and Constrained model in the half-normal case.

However, the uniform case shows that the Dropped model in Figure 9 had a slightly tighter estimation of economies of scope than the Constrained model.

4.4 Cost Efficiency

The difference of cost efficiency estimates from the nonparametric models without technical inefficiency in quantities (No Inefficiency) for both distributions and the actual frontier were identical in that every single observation yielded the exact same cost efficiency estimate (Table 10). This implies that the minimum cost estimated from the nonparametric system was the same as that of the actual frontier. Thus, the No Inefficiency procedure correctly estimated the "true" cost frontier for the half-normal and uniform distributions.

With inefficiencies introduced in the input quantities, the Constrained and Dropped model's differences for the half-normal distribution were small. Approximately 80% of the observations had a difference of less than 0.05 in absolute value from the true cost efficiency with approximately 12% having a difference of less than 0.1 but greater than 0.05 in absolute value (Figure 9). The implication is that, an introduction of technical inefficiency in the input variables does not significantly reduce the accuracy of the nonparametric models estimates of cost efficiency in the half-normal case.

In the case of the uniform distribution, both the Constrained and Dropped model estimated the same frontier as illustrated by the same mean and standard deviation for the differences. However, the half uniform estimated the frontier more closely than the half-normal with a mean, mean absolute deviation, and standard deviation for both models of nearly zero (Table 10). This is also confirmed in Figure 9. Both models slightly over estimated the cost efficiency on average with an average for both models in the half-normal case being negative and in the uniform case however, they are both close to zero.

5. Conclusions

This paper develops, and tests a method for estimating product specific scale economies and multi-product scale economies using Färe's nonparametric method using two efficiency distributions. Results indicate that much more economic information can be obtained from Data Envelopment Analysis (DEA) than has occurred in the literature. Thus, there is more to DEA analysis than just efficiency analysis.

Alternative specifications of the nonparametric approach to measure incremental costs by forcing the appropriate output to equal zero rather than dropping the constraint as suggested by Chavas and Aliber are tested. The results are compared to a "true" cost function using economic simulation where the difference between the "true" measures and the estimated values are used to evaluate the accuracy of the approach.

When measuring observations with inefficiency, the nonparametric approach with the uniform and half-normal distributions does well in estimating scope, multi-product scale economies, cost efficiency, and product-specific scale economies. The mean differences were close to zero as were the mean absolute deviations. While the PSE estimates are close to the "true" frontier PSEs, in the half-normal case, the deviations for PSE calculations using a uniform distribution illustrate the importance of having observations from efficient firms producing a single output. Since PSE is based on a ratio of incremental costs to marginal costs, the PSE measures are sensitive to these calculations. In areas where there are few single output observations where observations are not on the "true" frontier, the estimated incremental costs for these observations may deviate from the "true" incremental cost. In areas of the data where there are many observations, the likelihood that observations do not come close to the frontier is small. Thus, areas where the data are clustered yield more precise results than areas where observations are sparse. It should be noted that this is important when estimating these measures using parametric estimation.

The approach developed in this article to obtain more economic analysis out of DEA has been shown to accurately estimate scope, multi-product scale, and product-specific economies. DEA's consistency with economic theory without restrictions on technology make it particularly attractive empirically along with the ability to estimate from a primal rather than a dual approach due to the higher need for relative price variability in the dual approach. Thus, with the new approaches developed to measures economies of scale, there are many more economic effects that can be obtained from DEA analysis than has been reported in the literature.

References

Aigner, D.J., C.A.K. Lovell, and P. Schmidt (1977) Formulation and estimation of Stochastic Frontier Production Models. Journal of Econometrics 6: 21-37.

Banker, Rajiv D. and Ajay Maindiratta (1988) Nonparametric Analysis of Technical and Allocative Efficiencies in Production. Econometrica 46 (69): 1315-1332.

Baumol, W. J., Panzar, J. C. and Willig, R. D. (1982) Contestable Markets and the Theory of Industry Structure. Harcourt Brace Janaovich, Inc. New York

Cazals, C., Jean-Pierre Florens, and Lepold Simar. (2002) Nonparametric Frontier Estimation: A Robust Approach. Journal of Econometrics 106: 1-25.

Chavas, Jean-Paul and M. Aliber (1993) An Analysis of Economic Efficiency in Agriculture: A Nonparametric Approach. Journal of Agricultural and Resource Economics 18: 1-16.

Chavas, Jean-Paul and T.L. Cox (1988) A Nonparametric Analysis of Agricultural Technology. American Journal of Agricultural Economics 70: 303-310.

Christensen, Lauritis R. Dale W. Jorgenson, and Lawrence J. Lau. (1973) Transcendental Logarithmic Production Frontiers. The Review of Economics and Statistics 55 (1): 28-45.

Cooper, Wiliam W., Lawrence M. Seiford, and Kaoru Tone. (2007) Data Envelopment Analysis. Springer Science+Media LLC. New York, New York.

Cummins, J. David, Mary A. Weiss, Xiaoying Xie, and Hongmin Zi. (2010) Economies of Scope in Financial Services: A DEA Efficiency Analysis of the US Insurance Industry. Journal of Banking & Finance 34: 1525-1539.

Färe, R., S. Groskopf, and C.A.K Lovell (1985) The Measurement of Efficiency of Production. Boston: Kluwer-Nijhoff.

Färe, R., R. Grabowski, S. Grosskopf, S. Kraft (1997) Efficiency of a Fixed but Allocable Input: A Nonparametric Approach. Economic Letters 56: 187-193.

Farrell, M. J. (1957) The Measurement of Productive Efficiency. Journal of the Royal Statistical Society 120 (3): 253-290.

Featherstone, Allen M., Michael R. Langemeier, and Mohammad Ismet. (1997) A Nonparametric Analysis of Efficiency for a Sample of Kansas Beef Cow Farms. Journal of Agricultural and Applied Economics 29 (1): 175-184.

Featherstone, Allen M., Charles B. Moss (1994) Measuring Economies of Scale and Scope in Agricultural Banking. American Journal of Agricultural Economics 76 (3): 655-661.

Florens, Jean-Pierre, and Lepold Simar (2005) Parametric Approximations of Nonparametric Frontiers. Journal of Econometrics 124: 91-116.

Forsund, Finn, R., Lennart Hjalmarsson. (2004) Calculating the Scale Elasticity in DEA Models. Journal of Operational Research Society 55(10): 1023-1038.

Gao, Zhifeng and Allen M. Featherstone (2008) Estimating Economies of Scope Using the Profit Function: A Dual Approach for the Normalized Quadratic Profit Function. Economic Letters 100: 418-421.

General Algebraic Modeling System 24.2.1. Released 9 Jan 2013. Gams Development Corporation. Washington, DC. USA. Website: http://www.gams.com/index.htm

Hoff, A., (2007) Second Stage DEA: Comparison of Approaches for Modeling the DEA Score. European Journal of Operational Research 181: 425-435.

Jaforullah, Mohammad and John Whiteman. (1998) Scale Efficiency in the New Zealand Dairy Industry: A Nonparametric Approach. The Australian Journal of Agricultural and Resource Economics 43 (4): 523-541.

Kumar, Sunil, and Richita Gulati (2008) An Examination of Technical, Pure Technical, and Scale Efficiencies in Indian Public Sector Banks using Data Envelopment Analysis. Eurasian Journal of Business and Economics 1 (2): 33-69.

Lau, L.J. (1976) A Characterization of the Normalized Restricted Profit Function. Journal of Economic Theory. 12: 131-163.

Lusk, Jayson L., Allen M. Featherstone, Thomas L. Mash and Abdullahi O. Abdulkadri (2002) The Empirical Properties of Duality Theory. The Australian Journal of Agricultural and Resource Economics. 46: 45-68.

Paul, Catherine Morrison, Richard Nehring, David Banker, and Agapi Somwaru (2004) Scale Economies and Efficiency in U.S. Agriculture: Are Traditional Farms History? Journal of Productivity Analysis 22: 185-205.

Ray, S.C. (1998) Measuring Scale Efficiency from a Translog Production Function. Journal of Productivity Analysis 11: 183-194

SHAZAM Professional Edition 11. Copyright 2012 SHAZAM Analytics Ltd. Norbury, London. Website: http://store.econometrics.com/

Shephard, R.W. (1970) Theory of Cost and Production Functions. Princeton, NJ: Princeton University Press.

Tables

Table 1 Coefficients used in cost function for data simulation for half-normal and uniform distributions

Coefficient	Value
A_1	30.0
A_2	80.0
A_{11}	0.50
A_{12}	1.00
A_{21}	0.60
A_{22}	0.50
B_0	20.0
B_1	10.0
B_2	35.0
B_{11}	-0.09
B_{12}	-0.15
B_{22}	-0.47
C_{11}	1.44
C_{12}	-0.24
C_{22}	2.29

Table 2 The average, standard deviation, minimum and maximum for the input/output quantities and input prices in half-normal (x_i^n) and uniform (x_i^u) cases

N=500	Average	Standard Deviation	Minimum	Maximum
x_1^n	42.29	11.95	13.35	88.33
x_2^n	69.85	23.29	38.44	268.76
x_3^n	2602.60	1154.75	152.95	8083.87
x_1^u	36.93	8.644	14.06	68.89
x_2^u	60.16	10.25	38.43	136.13
x_3^u	2302.06	1027.79	147.92	6585.05
w_1	9.05	0.98	5.42	11.98
w_2	17.95	1.88	13.15	24.70
w_3	6.98	0.78	4.85	9.75
y_1	11.67	5.90	0.00	30.19
y_2	14.31	7.53	0.00	37.92

Table 3 Summary statistics for MPSE, PSE, economies of scope, and cost efficiency

Economic Measure	Average	Standard Deviation	Minimum	Maximum
		------Half-normal Distribution------		
Multi-product Scale Economies	0.931	0.108	0.779	1.989
Cost Efficiency	0.721	0.177	0.129	1.000
Scope	0.096	0.051	0.037	0.513
Product-specific Scale Economies for y1	0.728	0.246	0.000	0.957
Product-specific Scale Economies for y2	0.763	0.257	0.000	0.995
		------Uniform Distribution------		
Cost Efficiency	0.799	0.133	0.268	1

Note: Economies of Scope, Multi-product Scale Economies, and Product-specific Scale economies are identical for the Half-normal and Uniform distributions

Table 4 Statistics for simulated multi-product scale economies estimates minus multi-product scale economies estimated nonparametrically for the half-normal and uniform distributions

N=476	Average	Standard Deviation	Minimum	Maximum	Mean Absolute Deviation
		------Half-normal Distribution-----			
Nonparametric No Inefficiency	-0.002	0.011	-0.108	0.046	0.008
Dropped	0.001	0.047	-0.145	0.235	0.049
Constrained	0.001	0.047	-0.145	0.235	0.049
		------Uniform Distribution-----			
Nonparametric No Inefficiency	-0.003	0.023	-0.336	0.198	0.008
Dropped	-0.012	0.054	-0.277	0.114	0.027
Constrained	-0.006	0.055	-0.266	0.320	0.031

Table 5 Statistics for simulated product-specific scale economies estimates minus product-specific scale economies estimated nonparametrically for the half-normal and uniform distributions for outputs 1 and 2

N=426		Average	Standard Deviation	Minimum	Maximum	Mean Absolute Deviation
		------Half-normal Distribution-----				
y_1	Nonparametric No Inefficiency	0.000	0.046	-0.325	0.629	0.016
	Dropped	0.133	0.221	-0.257	0.633	0.219
	Constrained	0.133	0.221	-0.257	0.633	0.219
y_2	Nonparametric no Technical Inefficiency	-0.002	0.020	-0.080	0.260	0.007
	Dropped	0.032	0.108	-0.344	0.712	0.086
	Constrained	0.022	0.109	-0.204	0.723	0.085
		-----Uniform Distribution------				
y_1	Nonparametric No Inefficiency	0.002	0.024	-0.084	0.109	0.016
	Dropped	0.003	0.106	-0.214	0.185	0.107
	Constrained	0.050	0.191	-0.239	0.735	0.259
y_2	Nonparametric No Inefficiency	0.002	0.021	-0.148	0.082	0.014
	Dropped	-0.003	0.051	-0.169	0.109	0.042
	Constrained	0.019	0.088	-0.294	0.294	0.068

Table 6 Statistics for simulated product-specific scale economies estimates minus product-specific scale economies estimated nonparametrically for y_1 removing the technical inefficiency in the input quantities

N=426	Average	Standard Deviation	Minimum	Maximum	Mean Absolute Deviation
Nonparametric No Inefficiency	-0.002	0.053	-0.399	0.629	0.016
Dropped	0.065	0.169	-0.744	0.583	0.154
Constrained	0.069	0.164	-0.272	0.671	0.154

Table 7 Statistics for simulated scope economies estimates minus scope economies estimated nonparametrically

N=397	Average	Standard Deviation	Minimum	Maximum	Mean Absolute Deviation
			------*Half-normal Distribution*------		
Nonparametric No Inefficiency	0.000	0.017	-0.201	0.193	0.003
Dropped	-0.098	0.070	-0.709	-0.043	0.098
Constrained	-0.089	0.034	-0.249	0.234	0.098
			------*Uniform Distribution*------		
Nonparametric No Inefficiency	0.001	0.020	-0.201	0.206	0.008
Dropped	-0.017	0.066	-0.712	0.017	0.020
Constrained	-0.028	0.095	-0.813	0.206	0.044

Table 8 Statistics for simulated cost efficiency minus cost efficiencies estimated nonparametrically for half-normal and uniform distributions

N=500	Average	Standard Deviation	Minimum	Maximum	Mean Absolute Deviation
		------*Half-normal Distribution*------			
Nonparametric No Inefficiency	0.000	0.000	0.000	0.000	0.000
Dropped	-0.025	0.041	-0.530	-0.003	0.026
Constrained	-0.025	0.041	-0.530	-0.003	0.026
		------*Uniform Distribution*------			
Nonparametric No Inefficiency	0.000	0.000	0.000	0.000	0.000
Dropped	-0.004	0.007	-0.079	0.000	0.004
Constrained	-0.004	0.007	-0.079	0.000	0.004

Figures

Note: the MPSE calculations for both the half-normal and uniform error distribution is identical.

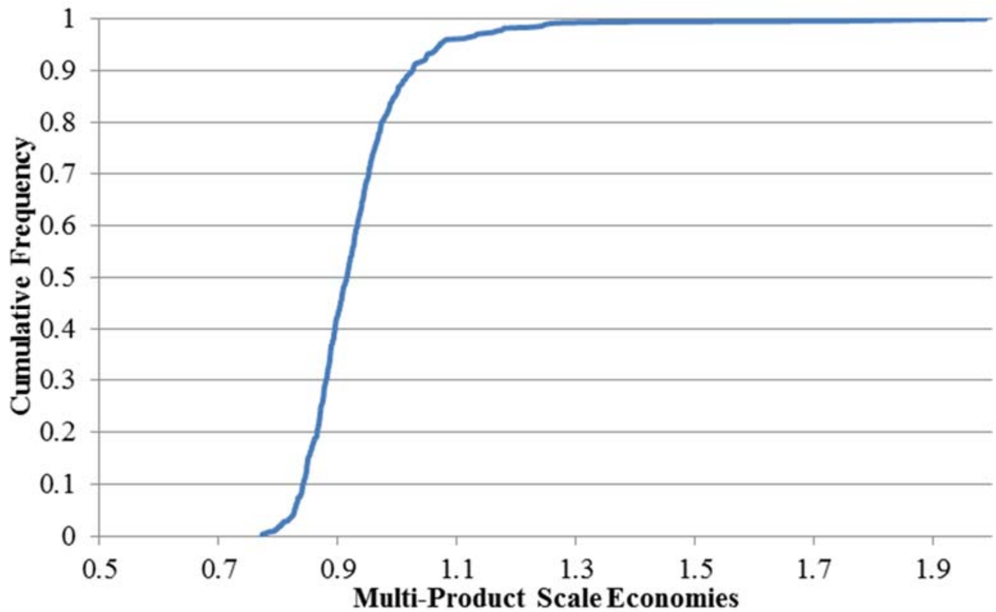


Figure 1 Frontier Multi-Product Scale Economies Cumulative Frequency for Generated Data

Note: The PSE calculations for Y1 and Y2 for both the half-normal and uniform error distribution are identical

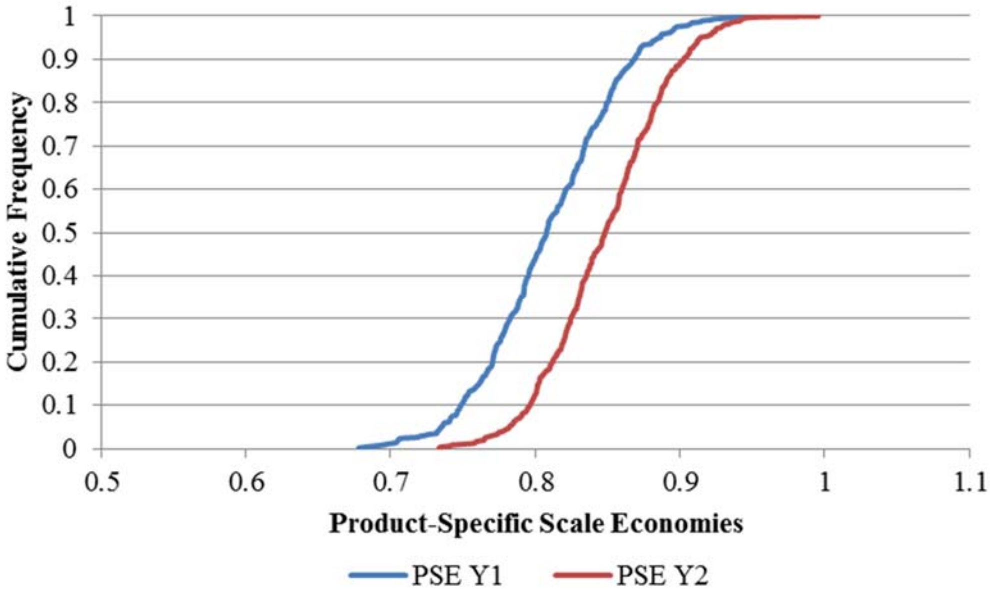


Figure 2 Frontier Product-Specific Scale Economies

Note: The Economies of Scope calculations for both the half-normal and uniform error distribution is identical.

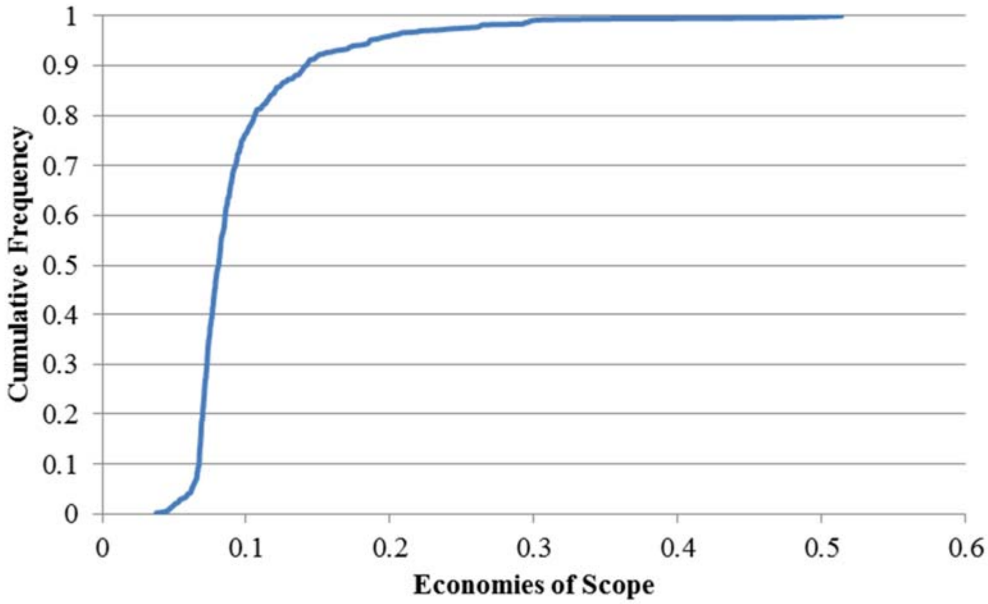


Figure 3 Frontier Economies of Scope Cumulative Frequency

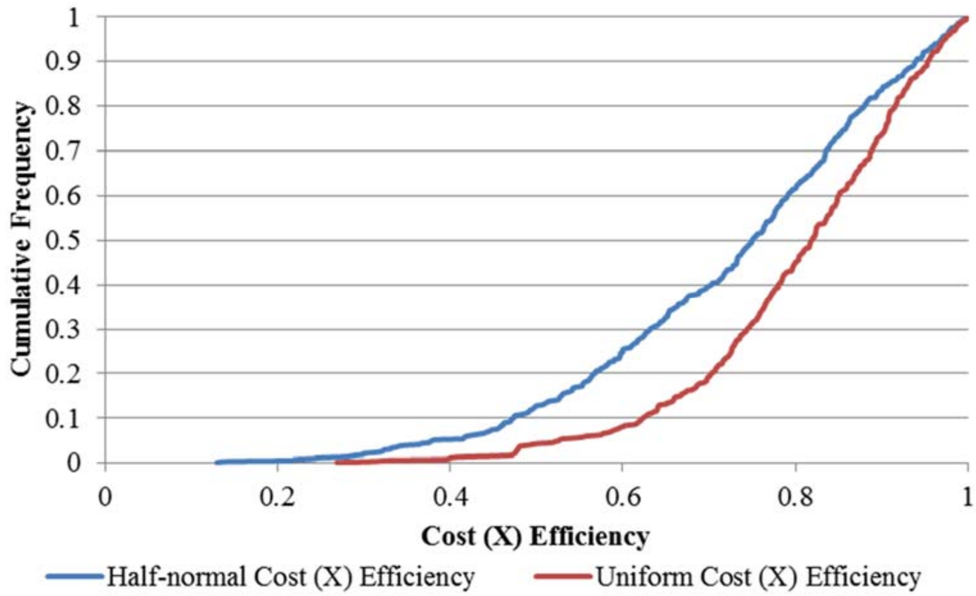
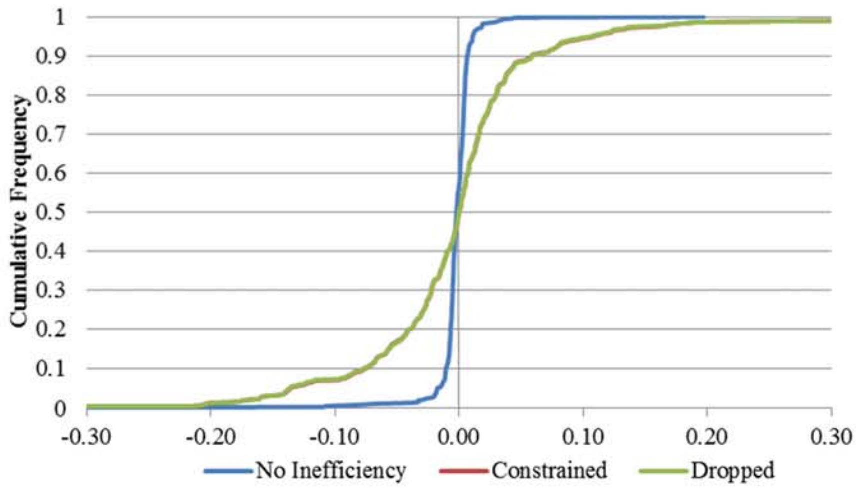


Figure 4 Frontier Cost Efficiencies Cumulative Frequency for both Half-normal and Uniform Distributions

Note: Constrained and Dropped trace out identically for Half-normal distribution.

Panel A: Multi-product Scale Economies Half-normal Distribution



Panel B: Multi-product Scale Economies Uniform Distribution

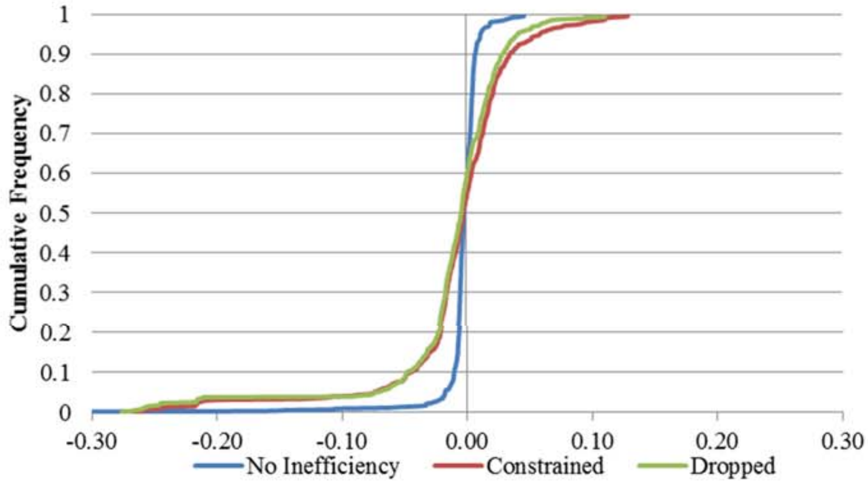
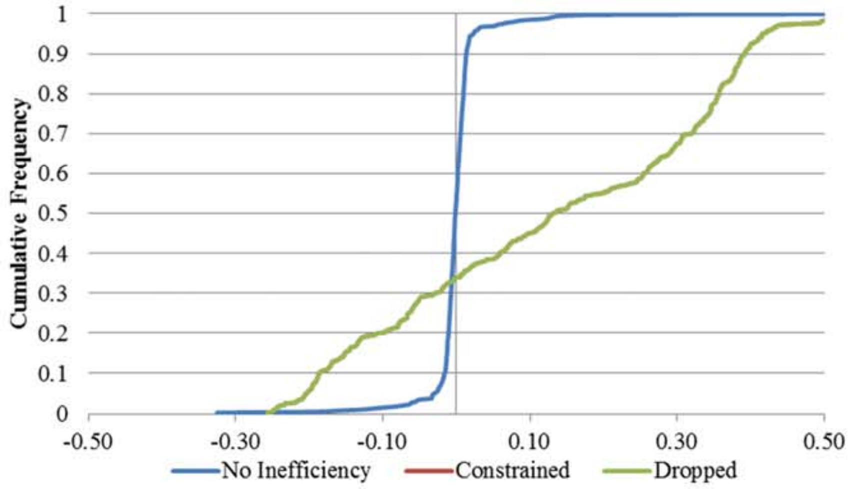


Figure 5 Differences between frontier Multiproduct Scale Economies and nonparametric estimates of Multiproduct Economies of Scale for Half-normal and Uniform Cumulative Distributions

Note: Constrained and Dropped trace out identically for half-normal distribution

Panel A: Product-specific Scale Economies Y_1 Half-normal Distribution



Panel B: Product-specific Scale Economies Y_1 Uniform Distribution

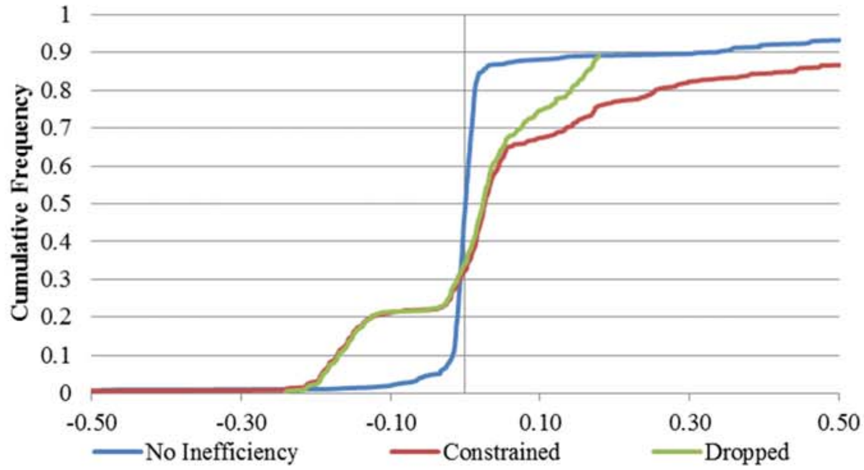


Figure 6 Differences between frontier Product-specific Economies of Scale for Y1 and nonparametric estimates of Product-specific Economies of Scale for Y1 for Half-normal and Uniform Cumulative Distributions

Panel a: Product-specific Scale Economies Y_2 Half-normal Distribution

Panel b: Product-specific Scale Economies Uniform Y_2 Distribution

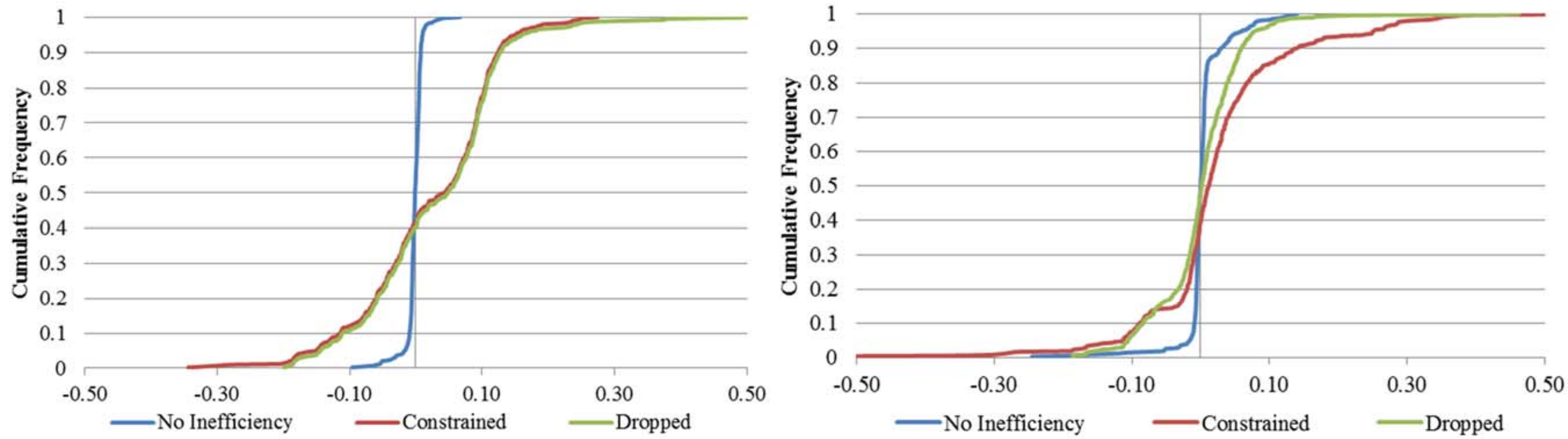


Figure 7 Differences between frontier Product-specific Economies of Scale for Y2 and nonparametric estimates of Product-specific Economies of Scale for Y2 for the Half-normal and Uniform Cumulative distributions.

Note: Constrained and Dropped trace out nearly identically

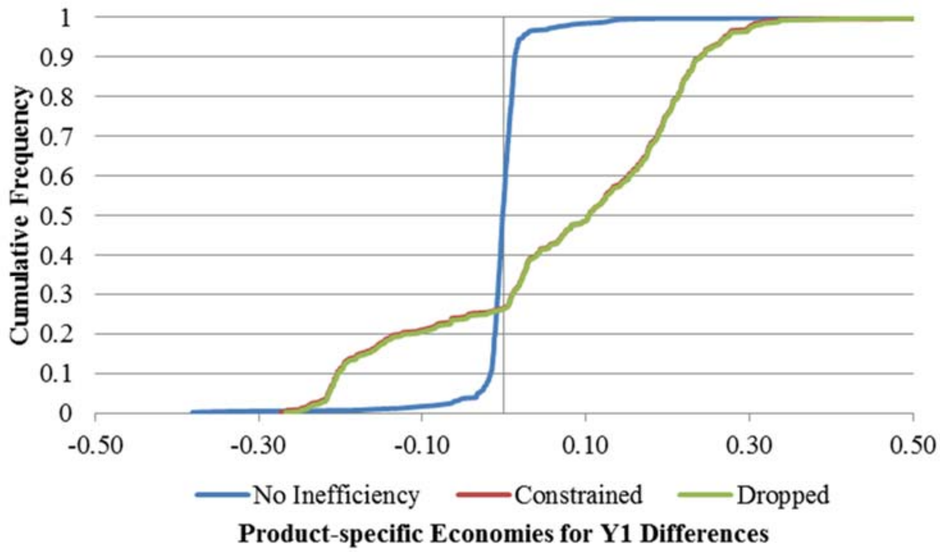
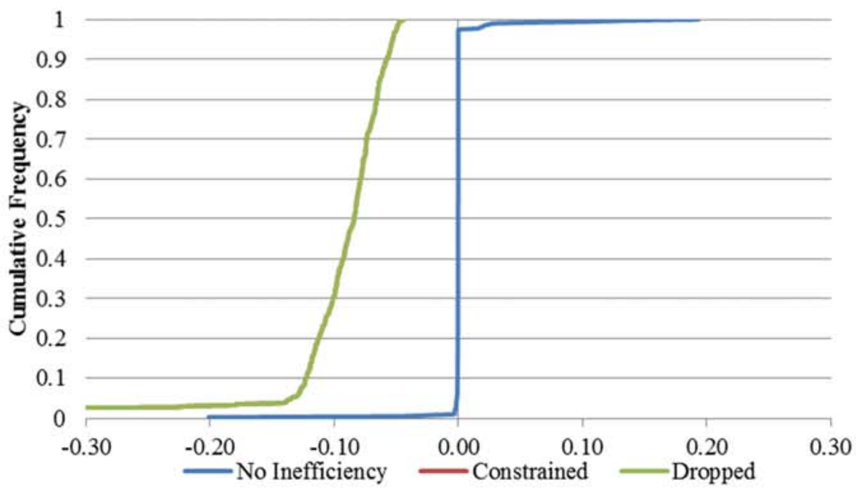


Figure 8 Differences between frontier Product-specific Economies of Scale for Y1 and nonparametric estimates of Product-specific Economies of Scale for Y1 removing technical inefficiency from frontier firms

Note: Constrained and Dropped trace out identically for economies of scope for the half-normal Cumulative distribution

Panel b: Economies of Scope Half-normal Distribution



Panel b: Economies of Scope Uniform Distribution

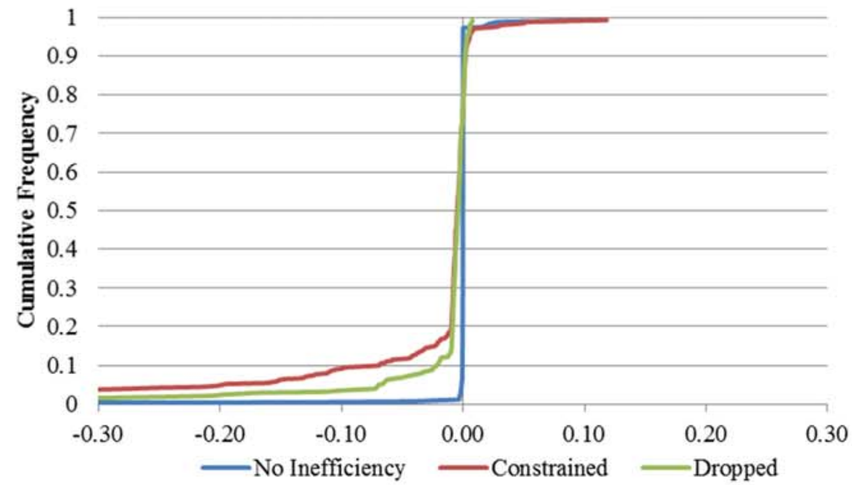
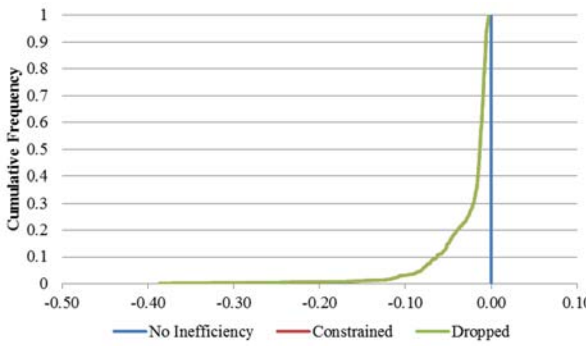


Figure 9 Differences between frontier Economies of Scope and nonparametric estimates of Economies of Scope for Half-normal and Uniform distributions

Note: Constrained and Dropped trace out identically for Cost Efficiency in Half-normal and Uniform cases.

Panel b: Cost Efficiency Half-normal Distribution



Panel b: Cost Efficiency Uniform Distribution

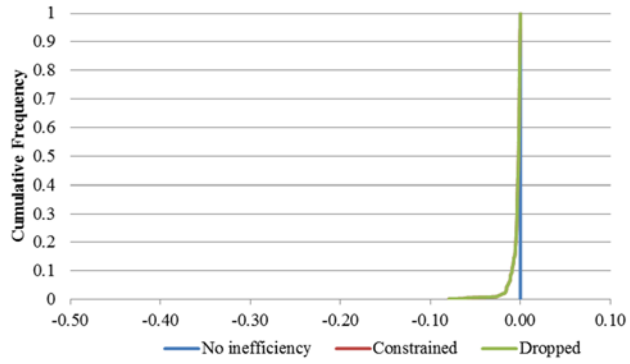


Figure 10 Differences between frontier Cost Efficiency and nonparametric estimates of Cost Efficiency for Half-normal and Uniform Cumulative Distribution