ECONOMIES OF SCALE FOR DATA ENVELOPMENT ANALYSIS WITH A KANSAS FARM APPLICATION

by

BRYON JAMES PARMAN

B.A., Peru State College, 2008
M.S., University of Nebraska-Omaha, 2010

AN ABSTRACT OF A DISSERTATION

Submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Agricultural Economics
College of Agriculture

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2013

Abstract

Estimation of cost functions can provide useful economic information to producers, economists, and policy makers. From the estimation of a cost function, it is possible to calculate cost efficiency, economies of scope, and economies of scale. Economic theory specifies the cost function as a frontier since firms cannot operate at lower cost than the cost minimizing input/output bundle. However, traditional parametric estimation techniques often violate economic theory using two sided-error systems. The stochastic frontier method has allowed the estimation of a frontier but continues to restrict the technology through functional assumption.

Nonparametric frontier estimation is an alternative approach to estimate a cost frontier by enveloping the data which by its construct, conforms to economic theory. This research expands the economic information available by deriving multi-product scale economies and product-specific scale economies from the nonparametric approach. It also tests its ability to accurately recover theses important economic measures under different assumptions of the cost function, and cost inefficiency distributions. Next, this new method is compared to other methods used to estimate cost functions and associated economic measures including a two-sided error system, stochastic frontier method, and an OLS model restricting the errors to take on only positive values. Finally, the nonparametric approach with the new measures is applied to a sample of Kansas farms.

The nonparametric approach is able to closely estimate economies of scale and scope from estimation of a cost frontier. Comparison reveals that the nonparametric approach is closer to the “true” economic measures than some parametric methods and that it is better able to extrapolate out of sample when there are no zero output firms. Finally, the nonparametric approach shows that potential cost savings from economies of scale and economies of scope exist for small Kansas farms. However, cost savings from economies of scale become exhausted when farms exceed gross annual revenues of $500k, while economies of scope also diminish as farms grow larger. Results also show from annual frontier estimations that estimates of economies of scale, scope, and cost efficiency have remained relatively stable from 2002 to 2011.

ECONOMIES OF SCALE FOR DATA ENVELOPMENT ANALYSIS WITH A KANSAS FARM APPLICATION

by

BRYON JAMES PARMAN

B.A., Peru State College, 2008
M.S., University of Nebraska-Omaha, 2010

A DISSERTATION

Submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Agricultural Economics
College of Agriculture

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2013

Approved by:

Co-Major Professor
Vincent Amanor-Boadu

Approved by:

Co-Major Professor
Allen M. Featherstone

Abstract

Estimation of cost functions can provide useful economic information to producers, economists, and policy makers. From the estimation of a cost function, it is possible to calculate cost efficiency, economies of scope, and economies of scale. Economic theory specifies the cost function as a frontier since firms cannot operate at lower cost than the cost minimizing input/output bundle. However, traditional parametric estimation techniques often violate economic theory using two sided-error systems. The stochastic frontier method has allowed the estimation of a frontier but continues to restrict the technology through functional assumption.

Nonparametric frontier estimation is an alternative approach to estimate a cost frontier by enveloping the data which by its construct, conforms to economic theory. This research expands the economic information available by deriving multi-product scale economies and product-specific scale economies from the nonparametric approach. It also tests its ability to accurately recover theses important economic measures under different assumptions of the cost function, and cost inefficiency distributions. Next, this new method is compared to other methods used to estimate cost functions and associated economic measures including a two-sided error system, stochastic frontier method, and an OLS model restricting the errors to take on only positive values. Finally, the nonparametric approach with the new measures is applied to a sample of Kansas farms.

The nonparametric approach is able to closely estimate economies of scale and scope from estimation of a cost frontier. Comparison reveals that the nonparametric approach is closer to the “true” economic measures than some parametric methods and that it is better able to extrapolate out of sample when there are no zero output firms. Finally, the nonparametric

approach shows that potential cost savings from economies of scale and economies of scope exist for small Kansas farms. However, cost savings from economies of scale become exhausted when farms exceed gross annual revenues of $500k, while economies of scope also diminish as farms grow larger. Results also show from annual frontier estimations that estimates of economies of scale, scope, and cost efficiency have remained relatively stable from 2002 to 2011.

Table of Contents

List of Figures

List of Tables

Acknowledgements

This material is based upon work supported by National Science Foundation Grant # 0903701: "Integrating the Socioeconomic, Technical, and Agricultural Aspects of Renewable and Sustainable Biorefining Program, awarded to Kansas State University." Funding by the Center for Sustainable Energy, Kansas State University is kindly acknowledged

I would first like to thank Dr. Vincent Amanor-Boadu for his sound advice, guidance, and mentoring as I worked toward completion of my degree. For me, the latitude and encouragement he has provided allowed me to pursue my own research interests with the opportunity to work on multiple projects over the years. I am very grateful for the conversations we have had which have helped give me perspective in many areas I had not previously considered.

I would like to thank Dr. Allen Featherstone for recently setting me on a research path that I have found interesting and challenging, and believing that I was capable of completing this latest research. I would like to acknowledge all of his hard work and patience during the development of this dissertation, and I am especially grateful for the very long hours he has put in editing and refining the work contained herein.

To my committee members, Dr. Jason Bergtold, and Dr. Todd Easton I would like to thank them both for helping me through this process and for their advice on my research. I would like to especially thank Dr. Peter Pfromm for his advice and encouragement for not only this research, but on previous work and associated presentations conducted during pursuit of my degree.

I am also very grateful to Dr. Mary Rezac and Keith Rutlin for accepting me into the IGERT program at Kansas State University. They were instrumental in providing the resources I

needed to complete my goals and my degree and provided some unique opportunities for travel and scholarly collaboration not available to many graduate students in my position.

Chapter 1 - Introduction

The economic definition of a cost frontier is that it represents the lowest cost for producing a given level of output. There are fundamental elements in the study and evaluation of industry structure. Firms on the frontier change their cost by changing their output level or output bundle. These firms are unable to improve cost through alterations to their input mix. Firms above the frontier are not efficient and can reduce their cost by changing the output levels or their input bundle.

Figure 1.1 illustrates a single output cost frontier where the cost curve is the minimum cost to produce a given output level. Points A, B, and C represent the actual total cost for three firms where firms A and B are producing at costs higher than the frontier cost for their respective output levels. Point C is operating on the cost frontier. The calculation of cost efficiency (CE_i) of a firm, i, is the ratio of minimum cost (TC_{min}) to actual total cost incurred in the production of the output (ATC_i) and represents the distance the firm is from the frontier.

$$CE_i = \frac{TC_{\min}}{ATC_i} \tag{1.1}$$

When estimating the cost frontier, economies of scale for firm i can be calculated. Economies of scale refer to the cost reductions obtained as the firms size approaches constant returns to scale (Figure 1.2). The economies of scale of firm i in the production of output Y, (S_{iY}) may be determined for a single product Y, produced by the firm as follows:

$$S_{iY} = \frac{C(Y)}{Y\left[\frac{\partial C(Y)}{\partial Y}\right]} \tag{1.2}$$

where $C(Y)$ is the total cost and $\partial C(Y)/\partial Y$ is the marginal cost for producing Y.

Figure 1.2 represents the average cost of firms A, B, and C. Firm C is on the frontier operating at the minimum average total cost, yielding a scale economy measure equal to one and a cost efficiency measure equal to one. Firms A and B are off the frontier by distance α_A and α_B, respectively such that both firms can reduce average total cost by moving closer to the frontier. However, firm A can reduce costs more by exploiting economies of scale rather than improving cost efficiency. The measure θ_A is the distance from the frontier to the line tangent to the minimum average total cost that defines potential savings from increasing output. Since $\theta_A > \alpha_A$, holding CE (α_A) constant and increasing output to the same level as firm C reduces average cost more than improving cost efficiency. Firm B, since $\theta_B < \alpha_B$, improves cost efficiency while holding output constant leading to a reduction in its average total cost more than through output growth.

For the case of a multi-output firm, economies of scale (MPSE) for a firm producing i products, is defined as follows:

$$MPSE = \frac{C(Y)}{\sum_i Y_i \left[\frac{\partial C(\mathrm{Y})}{\partial Y_i} \right]} \tag{1.3}$$

The cost frontier for multi-output firms allows the calculation of product-specific economies of scale (PSE) and economies of scope as well. PSEs are calculated holding all other outputs constant while examining cost as one of the other outputs is varied. The calculation of PSE uses the marginal cost in addition to the incremental cost and average incremental cost for the output of interest. The incremental cost (IC_i) represents the cost the firm would incur were it to produce only output i. That is:

$$IC_i(Y) = C(\mathrm{Y}) - \mathrm{C}(\mathrm{Y}_{N-i}) \tag{1.4}$$

where Y_{N-i} is a vector with a zero component in place of Y_{i} and components equal to Y elsewhere. The average incremental cost (AIC_i) is the incremental cost divided by the output

$$AIC_i = \frac{IC_i}{Y_i} \tag{1.5}$$

Product-specific scale economies for firm i are defined as the ratio of AIC and marginal cost:

$$PSE_i = \frac{AIC_i}{\frac{\partial C(Y)}{\partial Y_i}} \tag{1.6}$$

Economies of scope (SC) are the potential cost savings that exist from simultaneous production of more than a single output by a single firm. Economies of scope measure the relative increase in cost should the firm split and produce each product individually. Mathematically, economies of scope for product Y is:

$$SC(Y) = \frac{\left[C(Y_T) + C(Y_{N-T}) - C(Y)\right]}{C(Y)} \tag{1.7}$$

where $C(Y_T)$ and $C(Y_{N-T})$ respectively define the cost of producing product Y_T and the remaining products Y_{N-T}.

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

$$\alpha_i = \frac{Y_i\left[\frac{\partial C(Y)}{\partial Y_i}\right]}{\sum_{i=1}^{2} Y_i\left[\frac{\partial C(Y)}{\partial Y_i}\right]} \tag{1.8}$$

Where α_i is the weight placed on the PSE of interest based upon its relative contribution to total output from a two output firm:

$$MPSE = \frac{\alpha_i PSE_i(Y) + (1-\alpha_i)PSE_{N-i}(Y)}{1-SC(Y)} \tag{1.9}$$

MPSE can take one of three values: decreasing, constant or increasing returns to scale. Equation 1.9 allows factors underlying the measures of MPSE. If *SC(Y*) 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(Y)* is greater than zero and the PSEs are at constant returns to scale, MPSE is increasing (>1).

Research Motivation

Historically, cost frontiers were econometrically estimated assuming a functional form such as a translog (Christensen et. al. 1973), normalized quadratic (Diewert and Wales 1988), or Generalized Leontief (Diewert 1971) using standard two-sided error systems where some errors are positive (above the frontier), and others are negative (below the frontier). Multiproduct-scale economies, product-specific scale economies, and economies of scope can be estimated from the parameter estimates. This approach allows firms to operate at lower costs than the cost frontier.

Frontier estimation research has addressed the issue of two-sided error estimations by ensuring that frontiers do not have firms below the frontier using both parametric and nonparametric approaches. The parametric approach is the stochastic frontier method by Aigner, Lovell and Schmidt with a nonparametric alternative, the data envelopment analysis method (DEA), proposed by Farrell.

Frontier approaches have typically analyzed the relative efficiency of firms in relationship to the frontier. Measures of economies of scale and/or scope are not typically reported. Thus, economic analysis has focused on firms above the cost frontier or the behavior of the frontier using dual methods and not both as illustrated in Figures 1.1 and 1.2.

This dissertation will examine methods to unify the measurement of scope, scale, and cost efficiency. The stochastic error methods can be used to measure scope and scale, but have typically focused on cost efficiency. The nonparametric method (DEA) has also focused primarily on cost efficiency, although Chavas and Aliber propose a method for measuring economies of scope. Measures of product-specific economies of scale and multi-product economies of scale have not been formalized in the literature for the nonparametric method. Finally, the literature does not contain analysis that compares the accuracy of alternative techniques to measures of cost efficiency, economies of scale and economies of scope from a "true" cost frontier.

This dissertation has been organized into three papers. The first formalizes and tests a method for calculating multi-product and product-specific scale economies from an estimated nonparametric cost frontier. This contributes to the literature by increasing the amount of information that may be reported from nonparametric estimation. The analysis uses two datasets assuming a “true” cost function generated using Monte Carlo. The simulations are conducted using half-normal and uniform distributions of cost efficiency. This allows for a comparison between the two distributions of the errors for the “true” cost frontier. Two “true” cost functions were assumed as well and two data sets examined with one containing single output firms and multiple output firms, and one with only multiple output firms.

The second study compares the accuracy of the nonparametric approach developed in the first paper to three parametric approaches. The three parametric approaches evaluated are a two-sided error system, the stochastic frontier method, and an OLS model in which all errors are restricted to be positive. The data sets used for the first objective are used for the second comparison. The different models abilities to accurately calculate the cost efficiency, economies of scope, multi-product economies of scale, and product-specific economies of scale measures are then evaluated.

The final paper uses the methods and techniques developed and tested under the first two papers on farm level data instead of simulated data. The data are obtained from the Kansas Farm Management Association for 241 farms from 2002 to 2011. Under this objective, the foregoing methods are used to estimate the cost efficiency measures along with multi-product and product-specific scale economies, and economies of scope for these Kansas farms.

The dissertation is presented in the next three chapters. The information presented in these chapters will be useful to researchers, economists, managers, and policymakers as it provides sound economic tools for quantifying the cost advantages farms have due to their relative size. The tools allow the determination of the extent that small farms may improve their cost savings from increasing output exploiting economies of scale. It will then be determined how much costs are reduced by producing multiple outputs (scope) rather than each output individually. In addition, cost efficiency reveals the reduction in overall costs that can be obtained by appropriately adjusting the input mix.

Chapter 1 References

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

Baumol, William J., John C. Panzar, and Robert O. Willig. *Contestable Markets and the Theory of Industry Structure*, Harcourt Brace Javanovich, Inc., New York, NY 1982.

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

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

Diewert, W.E. "An Application of the Shephard Duality Theorem: A Generalized Leontief Production Function" *Journal of Political Economy*, 79,3(1971): 481-507.

Diewert, W.E. and T.J. Wales. "A Normalized Quadratic Semiflexible Functional Form" *Journal of Econometrics*, 37, 3(1988): 327-342

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

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

Chapter 1 Figures

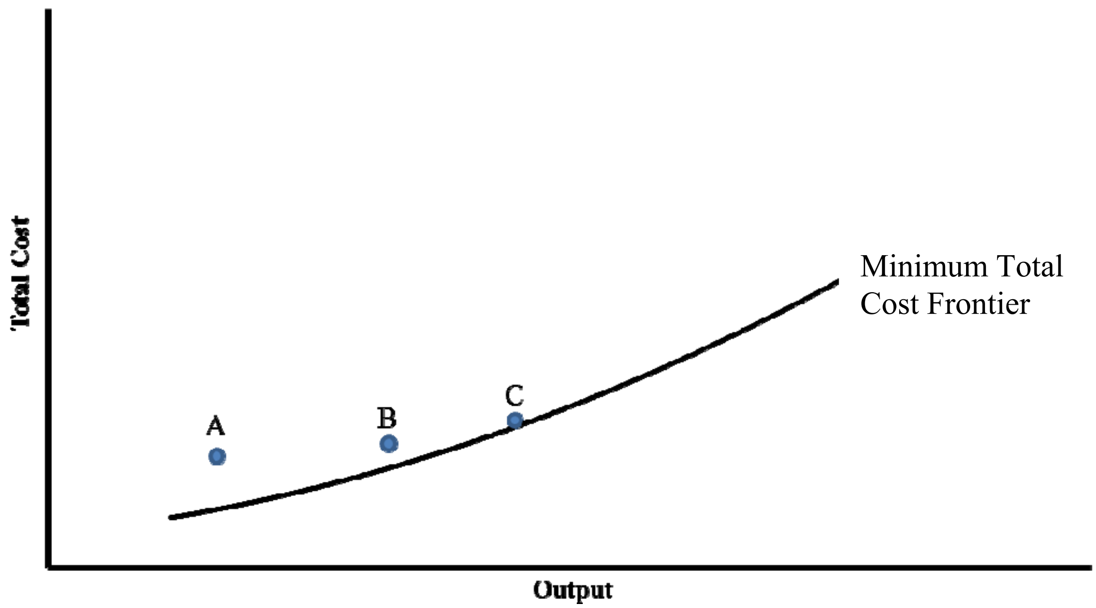


Figure 1.1 Actual Total Cost for Three Firms and the Total Cost Frontier

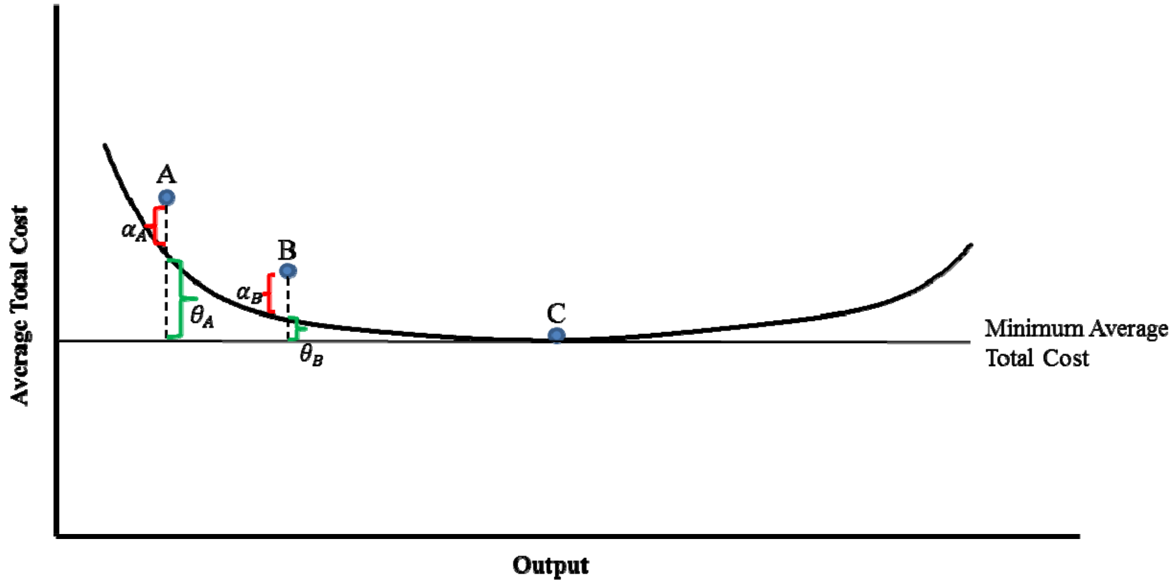


Figure 1.2 Minimum Average Total Cost and Actual Average Total Cost for Three Firms

Chapter 2 - A Nonparametric Approach to Multi-product and Product-specific Economies of Scale

Introduction

Producer theory provides useful tools for exploring the structure of cost. Estimates of frontier functions, and the distance of firms from the frontier 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.

Traditionally, multi-product and product-specific economies of scale and economies of scope are estimated parametrically using two-sided error systems through specification of a cost function and estimation of parameters (Christenson et. al.). The error structure is important 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). The stochastic frontier method has addressed the concerns of two-sided error systems by restricting the errors using positive error models (Aigner, Lovell, and Schmidt).

Lusk et al. examined the relative variability needed in the estimation of dual cost functions. They found that the relative variability necessary to accurately estimate a dual cost function requires more than 20 years of data based on observations. Thus, dual cost functions may have difficulty recovering the underlying technology. Featherstone and Moss note that parametric frontier estimations may also violate curvature of the cost function. Therefore, the lack of ability to accurately measure the underlying technology given data availability and the frontiers not maintaining the necessary cost function conditions are issues that may affect parametric frontier methods.

One of the methods used for frontier estimation is the nonparametric method that constructs a frontier from a series of line segments using a linear cost minimization program (Färe, Groskopf, and Lovell). With this method, it is not necessary to restrict the production technology by imposing a functional form. The nonparametric approach conforms to economic theory because curvature restrictions on the production/cost function are imposed in the estimation process. Further, the nonparametric method of Färe et al. may allow technology to be 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, Jaforullah and Whiteman, and Chavas and Cox. In these studies, several types of efficiencies 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 measure scope economies to determine cost savings from production portfolio diversification in the nonparametric framework.

Typical parametric measures of multi-product scale and product-specific scale measures have not yet been developed in the nonparametric DEA framework. 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. However, Paul et al. explain that the interpretation of scale efficiency is not as straight forward as a traditional scale economy measure explained by Baumol et. al. 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. Further, techniques for estimating product-specific economies of scale have not been reported for the nonparametric method.

This research develops and tests estimation techniques for multi-product and product-specific economies of scale for the nonparametric 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 and proposed economic measures.

In addition, previous research that estimates 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 examines that procedure by comparing a method which requires that output to be zero as required in the theory 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. Further, we evaluate the nonparametric approach under alternative efficiency distributions to investigate the robustness of the results.

Theory

Typical economic measures calculated from a cost function estimation 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.

$$MPSE = \frac{C(Y)}{\sum_p Y_p \left[\frac{\partial C(\mathrm{Y})}{\partial Y_p} \right]} \tag{2.1}$$

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:

$$IC_p = C_p - \sum_j C_{j \neq p} \forall j \tag{2.2}$$

Thus,

$$AIC_p = \frac{IC_p}{y_p} \tag{2.3}$$

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.

$$PSE_p = \frac{AIC_p}{\frac{\partial C(\mathrm{Y})}{\partial Y_p}} \tag{2.4}$$

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 $Y_{N-T} = (Y_1,\ldots Y_{k-1},0\ldots0)$.

$$SC(Y) = \frac{\left[C(Y_T) + C(Y_{N-T}) - C(Y)\right]}{C(Y)} \tag{2.5}$$

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:

$$\alpha_i = \frac{Y_i\left[\dfrac{\partial C(Y)}{\partial Y_i}\right]}{\sum_{i=1}^{N} Y_i\left[\dfrac{\partial C(Y)}{\partial Y_i}\right]} \tag{2.6}$$

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

$$MPSE = \frac{\alpha_i PSE_i(Y) + (1-\alpha_i) PSE_{N-i}(Y)}{1 - SC(Y)} \tag{2.7}$$

MPSE can take one of three values: decreasing, constant or increasing returns to scale. Equation 2.7 examines the relationship among factors affecting MPSE. If $SC(Y)$ 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(Y)* is greater than zero and the PSEs are at constant returns to scale, MPSE is in a region of increasing returns (>1).

Data and Methods

The Nonparametric Method

To estimate the new scale measures, 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^*).

$$
\begin{aligned}
&\min Ci = w_i' x_i^* \\
&s.t \\
&Xz \leq x_i^* \\
&y' z \geq y_i \\
&z_1 + z_2 + \ldots + z_n = 1 \\
&z_i \in \mathbb{R}^+
\end{aligned}
\qquad (2.8)
$$

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 ($C_{i,all\text{-}p}$) 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 2.8). 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 2.1). 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 the most efficient firms may not be unique. In practice this is usually a relatively small number of firms. In addition, a range of estimates of marginal costs 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 2.2). Previous methods to calculate incremental costs using the nonparametric method drop one or more of the output constraints from equation 2.8 to determine the cost of producing the output alone (Chavas and Aliber). For example if a firm produces four different products, four different linear programs would be estimated excluding one of the outputs at a time. In this research, results from dropping one of the output constraints are compared with constraining the appropriate output to zero.

Using equation 2.2, average incremental costs ($AIC_{i,p}$) are determined by dividing incremental costs by individual output as shown in equation 2.3. From the average incremental cost (equation 2.3) and the marginal cost calculations from the shadow prices, it is possible to calculate PSEs (equation 2.4) 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 2.5 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 goods 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.

$$CE_i = \left[\frac{C_i}{ATC_i} \right] \tag{2.9}$$

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 sub-section has 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.

Data Simulation

The data for the analysis were generated utilizing a modified Monte Carlo procedure found in Gao and Featherstone (2008) run on the SHAZAM software platform with the code found in Appendix A. 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.). 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:

$$\begin{aligned} &w_1 \sim N(9, 0.99) \\ &w_2 \sim N(18, 1.98) \\ &w_3 \sim N(7, 0.77) \\ &p_1 \sim N(325, 99) \\ &p_2 \sim N(800, 99) \end{aligned} \tag{2.10}$$

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. The assumed parameters are set to satisfy the following theory based condition: $b_{ij}=b_{ji}$ (symmetry in input prices). The assumed parameters (Table 2.1) are used to generate the output quantities y_1 *and* y_2[1]. The general form of the normalized quadratic cost function is:

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

$$
\begin{aligned}
&C(W,Y) = b_0 + \begin{bmatrix} b_1 & b_2 \end{bmatrix}\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} + \begin{bmatrix} a_1 & a_2 \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} + \\
&\frac{1}{2}\left\{\begin{bmatrix} w_1 & w_2 \end{bmatrix}\begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} + \begin{bmatrix} y_1 & y_2 \end{bmatrix}\begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}\right\} + \begin{bmatrix} w_1 & w_2 \end{bmatrix}\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}
\end{aligned} \tag{2.11}
$$

Output quantities (shown below) are calculated using the assumed parameters of the cost function (Table 2.1) and the random prices defined in equation 2.10.

$$
\begin{aligned}
y_1 &= \frac{c_{22}p_1 - c_{12}p_2 + \left(a_{12}c_{12} - a_{11}c_{22}\right)w_1 + \left(a_{22}c_{12} - a_{21}c_{22}\right)w_2 + \left(a_2c_{12} - a_1c_{22}\right)}{\left(c_{22}c_{11} - c_{12}c_{12}\right)} \\
y_2 &= \frac{c_{12}p_1 - c_{11}p_2 + \left(a_{12}c_{11} - a_{11}c_{12}\right)w_1 + \left(a_{22}c_{11} - a_{21}c_{12}\right)w_2 + \left(a_2c_{11} - a_1c_{12}\right)}{-\left(c_{22}c_{11} - c_{12}c_{12}\right)}
\end{aligned} \tag{2.12}
$$

Using the above cost function (Equation 2.11), 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 $e \sim N(0,1000)$[2]. 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.

An additional data set[3] is generated assuming a uniform distribution. The uniform deviation ranged from zero to 900. The normal distribution standard deviation of 1,000 generates a mean and standard deviation for cost efficiency roughly equivalent to a uniform distribution with a range from zero to 900.

From equation (2.11), and using Shephard's Lemma where $(\partial C(W,Y)/\partial w_i) = x_i$, 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 2.13).

---

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

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

$$
\begin{aligned}
x_1 &= b_1 + b_{11}w_1 + b_{12}w_2 + a_{11}y_1 + a_{12}y_2 \\
x_2 &= b_2 + b_{12}w_1 + b_{22}w_2 + a_{21}y_1 + a_{22}y_2 \\
x_3 &= \left(C\left(W,Y\right) - \left|e\right|\right) - x_1w_1 - x_2w_2\right)
\end{aligned}
\tag{2.13}
$$

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.

$$
x_i^a = \frac{x_i}{CE} \tag{2.14}
$$

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 2.2. In Table 2.2, 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 2.3. Summary statistics on scale and scope are independent of the distribution of cost "inefficiency". Figures 2.1 through 2.4 provide visual representations 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.

While the cost efficiency for each firm is simulated under a uniform and a half-normal distribution (Figure 2.2), the MPSE, PSE's, and Economies of Scope are identical for each data point (Table 2.3) for the "true" cost function. This is due to the fact that 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 2.14). The input quantities (x_i*'s*) are different in that the deviation in input quantity is uniformly distributed. In the uniformly distributed data more evenly distributes 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 Monte Carlo 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 nonparametric 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 2.10 for the No Inefficiencies model).

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 Monte Carlo 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 2.16), 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 0 (Constrained). The estimation was done using the General Algebraic Modeling Software and the code can be found in Appendix A.

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 cannot be calculated. For single output observations, PSEs cannot be calculated for the output not being produced. Economies of scope are also not calculated for single output observations.

Multi-product Economies of Scale

The differences results for MPSE are found in Table 2.4 and Figure 2.5. The No Inefficiencies model for both distributions shows little difference from the actual frontier function (Figure 2.5). The average bias was close to 0 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 2.4). 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 2.5). The standard deviations for both the Constrained and Dropped models were small (Table 2.4) 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.

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 2.5, and Figures 2.6 and 2.7). 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 0 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 2.5). 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 2.6 and 2.7. 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 2.6 and Figure 2.8 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 key to improving the accuracy of PSEs.

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 2.9**.** 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 0. For the uniform distribution, the differences were small as well with a standard deviation of 0.020 and mean absolute deviation nearly zero (Table 2.7). 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 2.7). The estimates were identical for both models. This indicates that the calculations for costs for producing zero output ($C_{i,all-p}$) 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 2.9 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 2.9 had a slightly tighter estimation of economies of scope than the Constrained model.

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 2.10). 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 2.9). 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 2.10). This is also confirmed in Figure 2.9. 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.

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. 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. The results are compared to a "true" cost function using Monte Carlo 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 do 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 nonparametric approach developed in this article has been shown to accurately estimate scope, multi-product scale, and product-specific economies. It’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.

Chapter 2 References

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

Atkinson, Scott E. and Robert Halvorsen. “Parametric Efficiency Tests, Economies of Scale, and Input Demand in U.S. Electric Power Generation” *International Economic Review.* 25(4)(October 1984):647-662.

Banker, Rajiv D. and Ajay Maindiratta. “Nonparametric Analysis of Technical and Allocative Efficiencies in Production.” *Econometrica*, 46,69 (November 1988): 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

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

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

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

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

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

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

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

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

Jaforullah, Mohammad and John Whiteman. "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. "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)(2008):33-69.

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

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

Paul, Catherine Morrison, Richard Nehring, David Banker, and Agapi Somwaru. "Scale Economies and Efficiency in U.S. Agriculture: Are Traditional Farms History?" *Journal of Productivity Analysis*, 22(2004): 185-2005.

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

Sahin, Ismet, Yasar A. Ozcan, Hacer Ozgen. "Assessment of Hospital Efficiency under Health Transformation Program in Turkey" *Central European Journal of Operations Research* 19,(1) (Mar 2011): 19-37.

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

Chapter 2 Tables

Table 2.1 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 2.2 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 2.3 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 2.4 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 Technical 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 Technical 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 2.5 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 Technical 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 Technical 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 Technical 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 2.6 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 Technical 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 2.7 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 Technical 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 Technical 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 2.8 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 Technical 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 Technical 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

Chapter 2 Figures

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

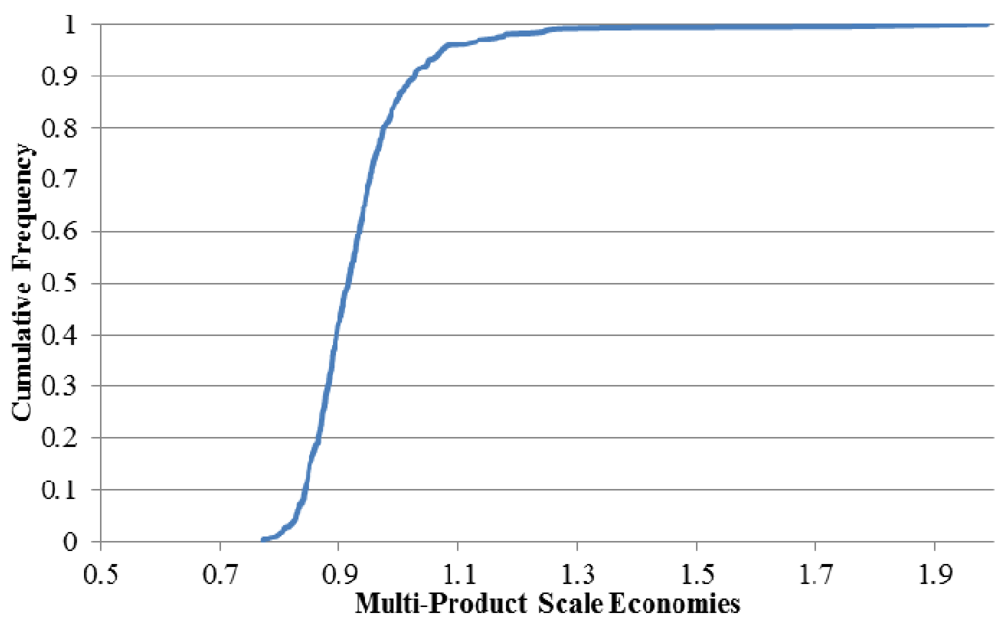


Figure 2.1 Frontier Multi-Product Scale Economies Cumulative Frequency for Generated Data.

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

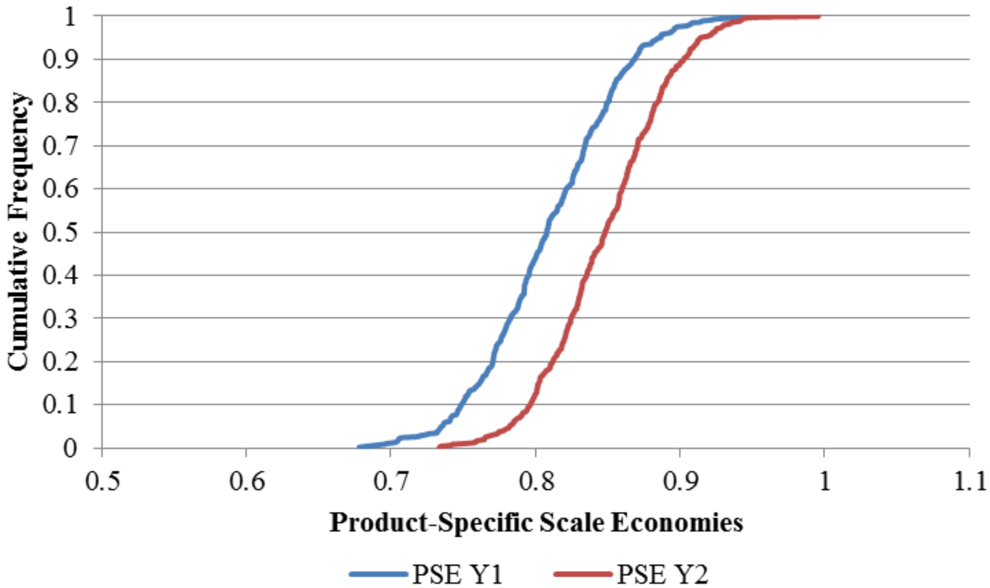


Figure 2.2 Frontier Product-Specific Scale Economies

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

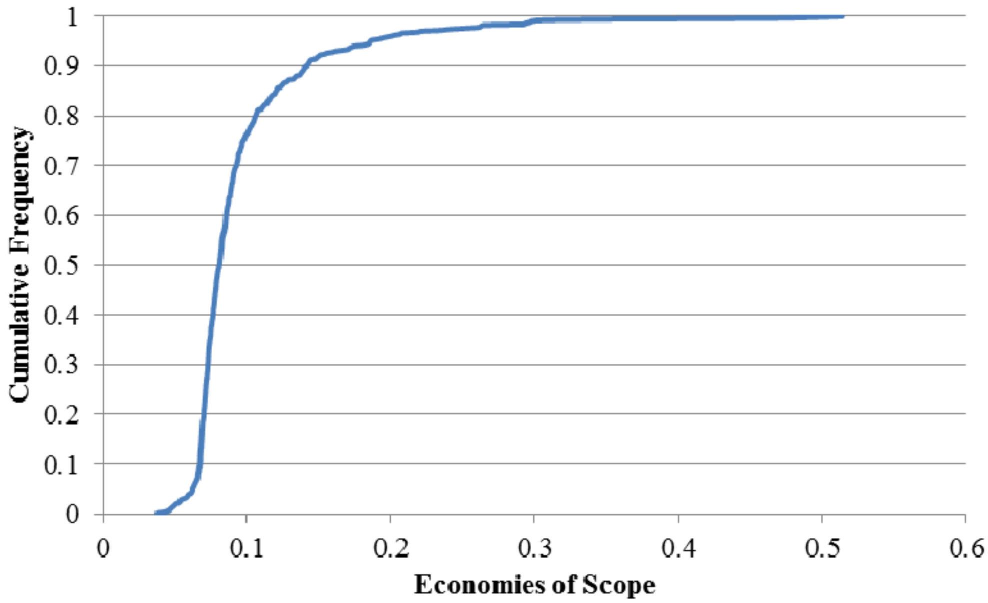


Figure 2.3 Frontier Economies of Scope Cumulative Frequency

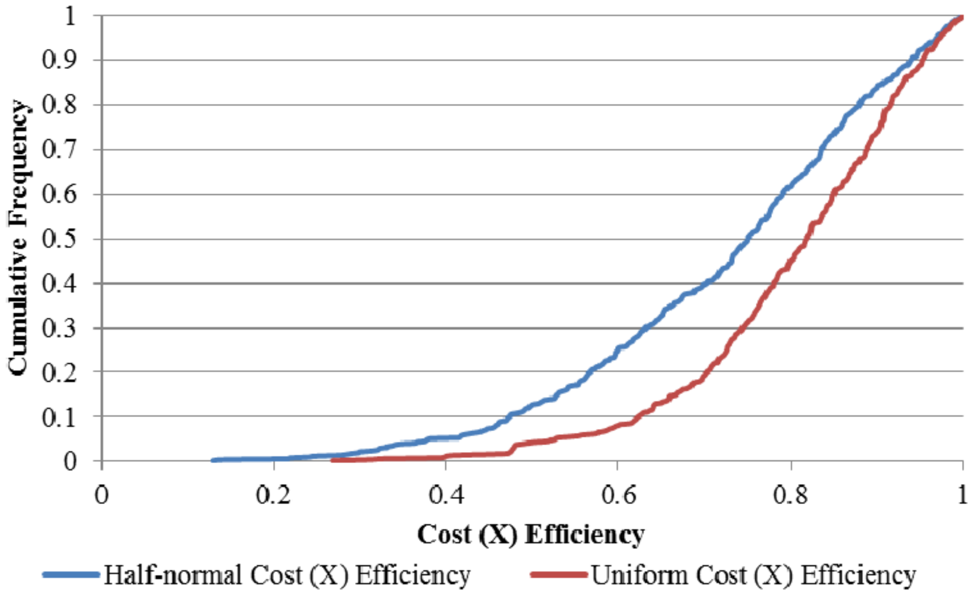
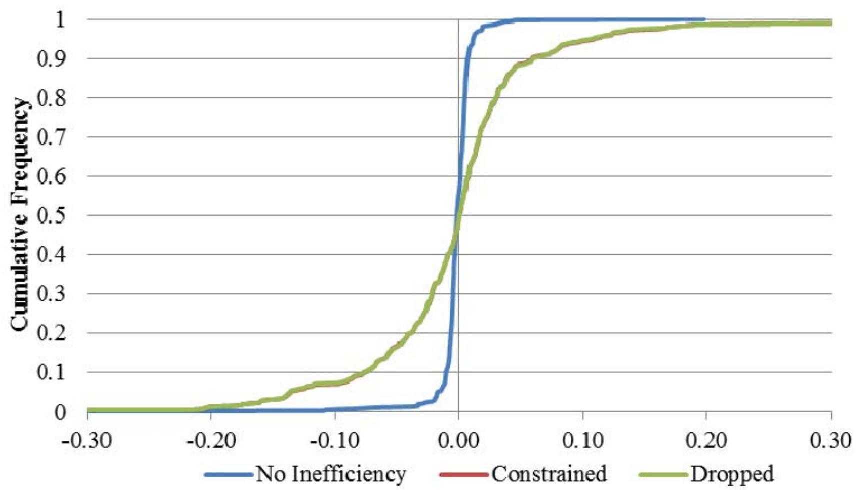


Figure 2.4 Frontier Cost Efficiencies Cumulative Frequency for both Half-normal and Uniform Distributions.

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

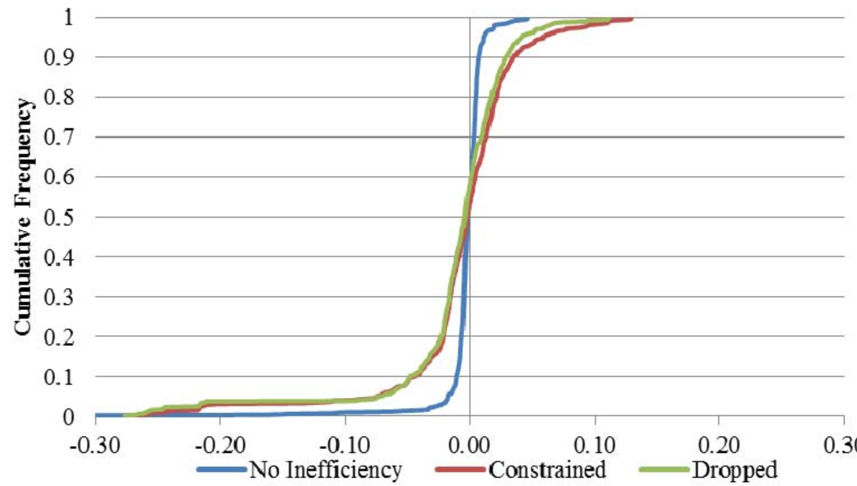


Figure 2.5 Differences between frontier Multiproduct Scale Economies and nonparametric estimates of Multiproduct Economies of Scale for Half-normal and Uniform Cumulative Distributions.

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

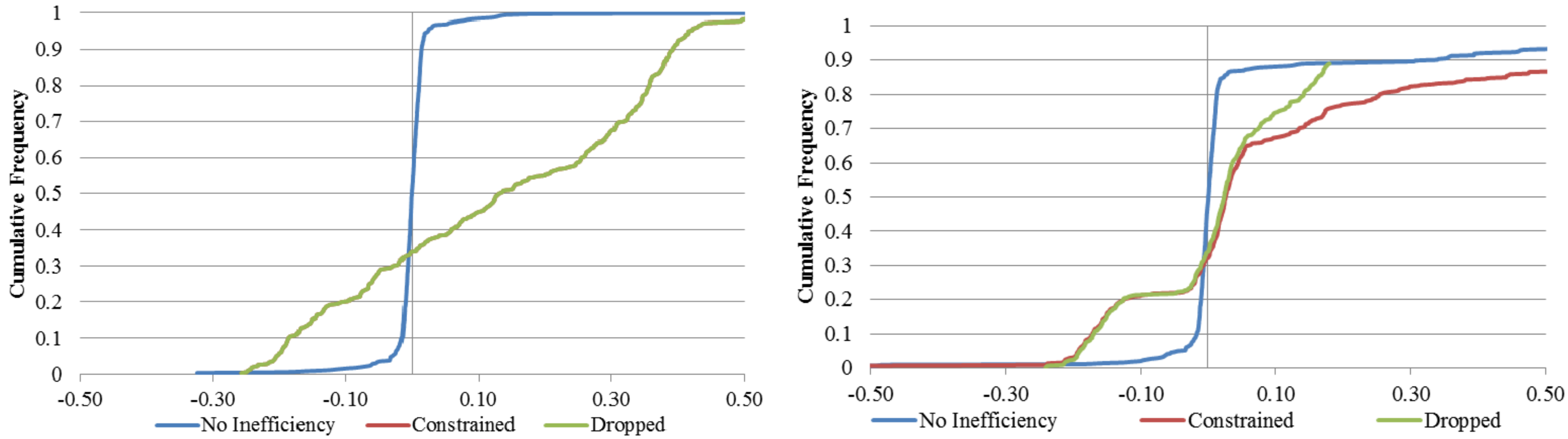
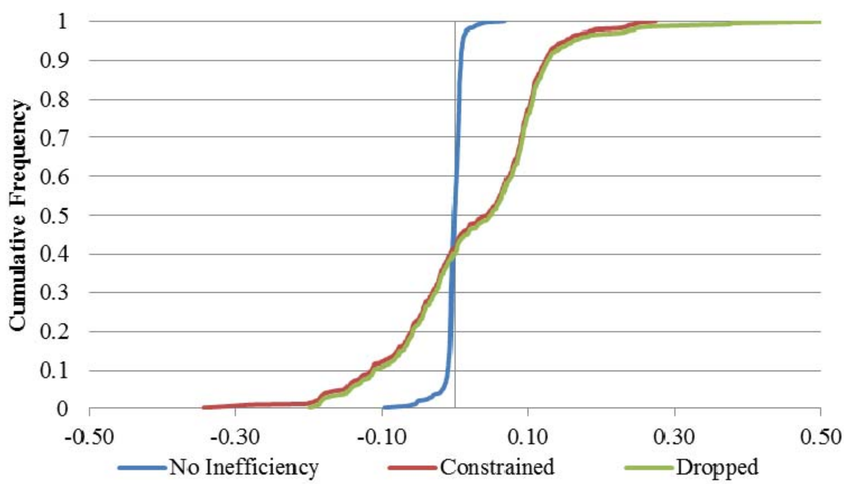


Figure 2.6 Differences between frontier Product-specific Economies of Scale for Y1 and nonparametric estimates of Product-specific Economies of Scale for Y1 for Half-normal and Uniform Cumulative Distributions.

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



Panel B: Product-specific Scale Economies Uniform Y_2 Distribution

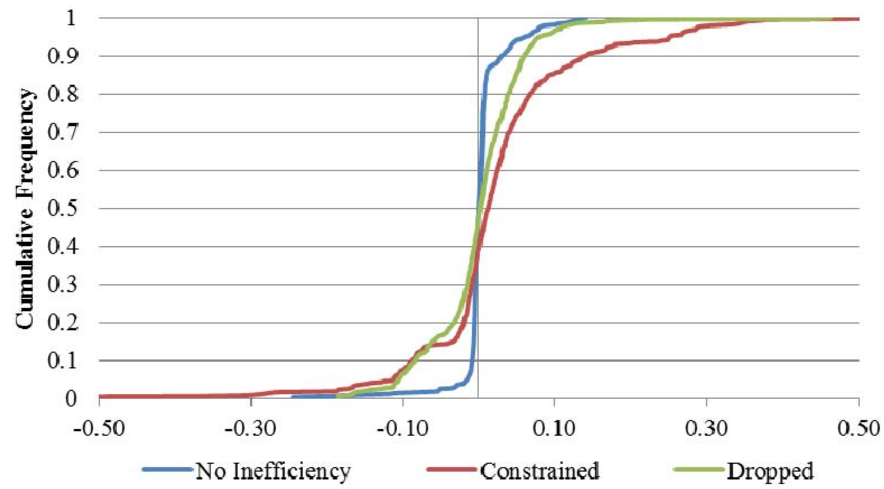


Figure 2.7 Differences between frontier Product-specific Economies of Scale for Y2 and nonparametric estimates of Product-specific Economies of Scale for Y2 for the Half-normal and Uniform Cumulative distributions.

Note: Constrained and Dropped trace out nearly identically

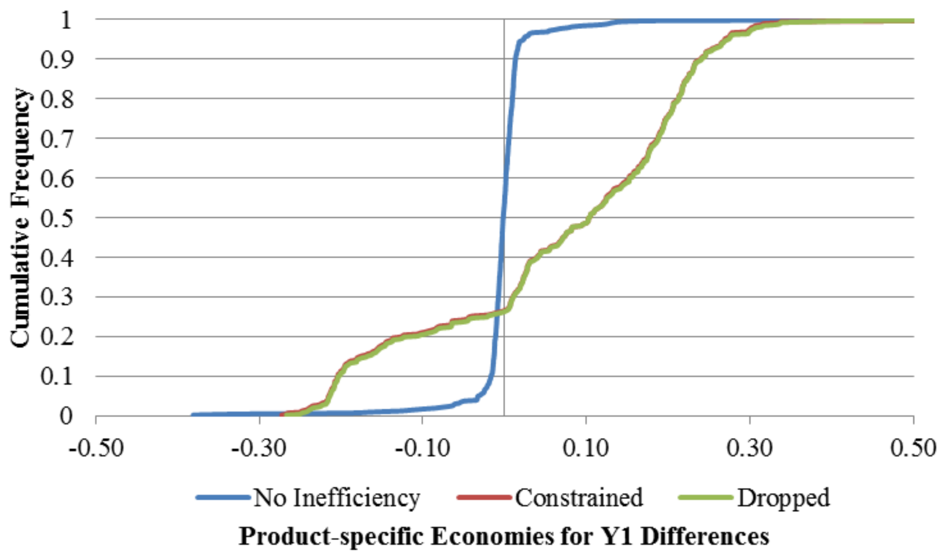


Figure 2.8 Differences between frontier Product-specific Economies of Scale for Y1 and nonparametric estimates of Product-specific Economies of Scale for Y1 removing technical inefficiency from frontier firms.

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

Panel A: Economies of Scope Half-normal Distribution **Panel B**: Economies of Scope Uniform Distribution

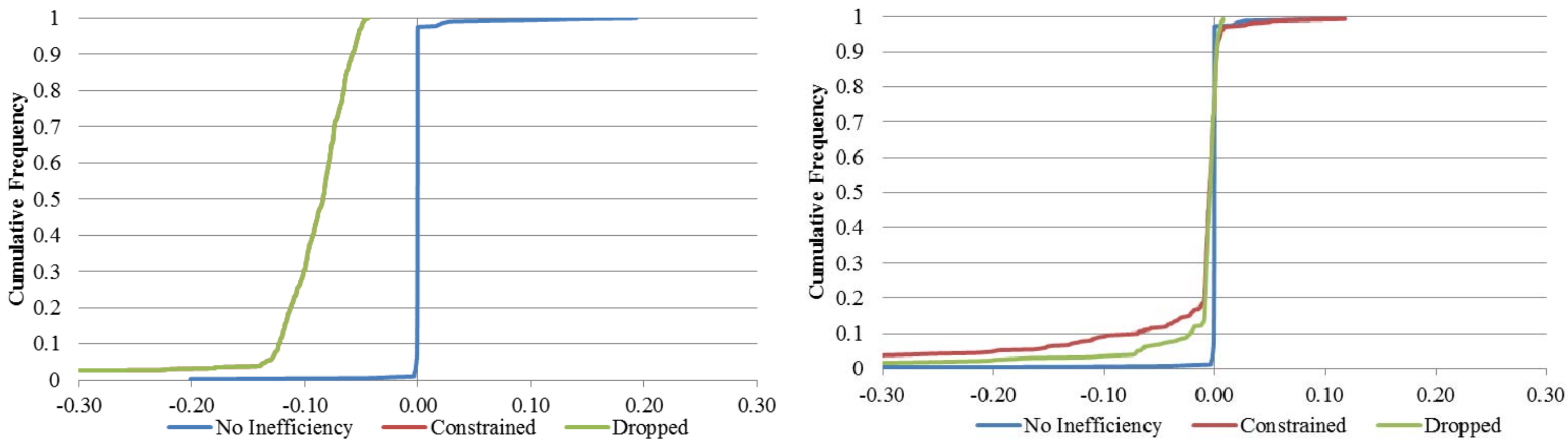
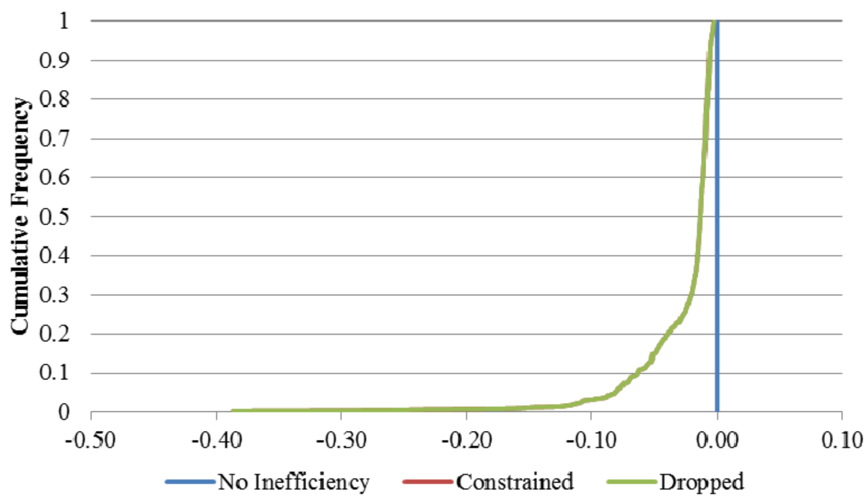


Figure 2.9 Differences between frontier Economies of Scope and nonparametric estimates of Economies of Scope for Half-normal and Uniform distributions.

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

Panel A: Cost Efficiency Half-normal Distribution



Panel B: Cost Efficiency Uniform Distribution

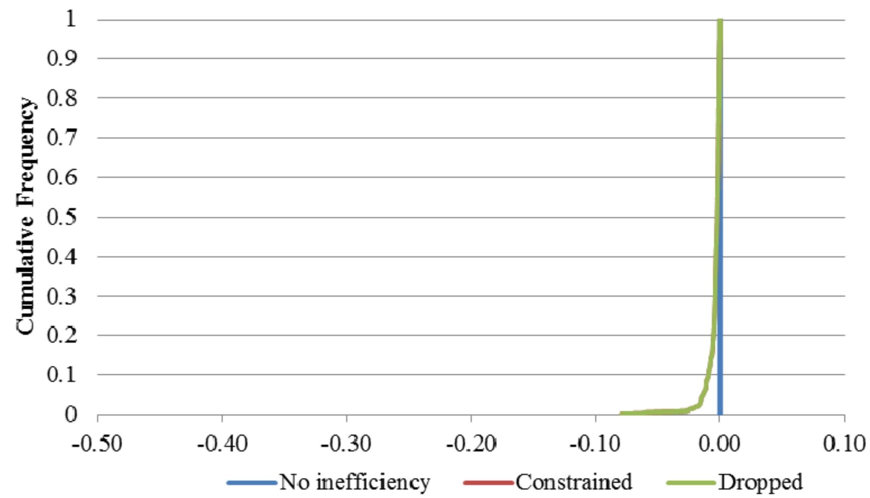


Figure 2.10 Differences between frontier Cost Efficiency and nonparametric estimates of Cost Efficiency for Half-normal and Uniform Cumulative Distribution

Chapter 3 - A Comparison of Parametric and Nonparametric Estimation Methods for Cost Frontiers and Economic Measures

Introduction

The study of producer theory uses several tools for exploring the structure of cost. Estimates of frontier functions, and the distances that firms are from the frontier provides insight into how firms with similar technological access and marketing opportunities achieve different levels of production efficiency. Frontier estimation also provides insight for both managers and economists regarding where cost savings exist for multi-product output firms. Parman et al. illustrate that it is possible to calculate multi-product scale economies and product-specific economies of scale that measure the potential for cost savings through the adjustment of output mix using Data Envelopment Analysis (DEA). Calculations of economies of scope from frontier estimation estimates illustrate how savings are achieved through producing multiple outputs in the same firm versus each output in a separate firm.

Traditionally, cost functions have been estimated using parametric methods with two-sided errors (i.e. OLS) where more efficient firms lie below the "average" frontier and less efficient firms lie above the "average" frontier (Christenson et. al. 1973, Diewert et. al. 1988). The result of such an estimation from a two-sided error model is thus an average cost function for the firms and not truly an estimation of the best practices (Greene 2005). Farrell (1957) used piece-wise linearization to envelope production data. In his analysis, all firms were either on or below the production frontier. In this way, the firms that reside on the frontier are relatively efficient, while those who resided below the frontier experience some amount of inefficiency. The distance from inefficient firms to the estimated frontier is calculated as a ratio of estimated minimum production inputs for a given output to actual production inputs for a given output was then used as a metric to determine relative efficiency among firms. Later works by Farrell and Fieldhouse (1962), and Afriat (1972) eliminated the restriction of constant returns to scale technology using the nonparametric approach. Charnes, Cooper and Rhodes (1978), while evaluating the technical efficiency of decision making units coined the name Data Envelopment Analysis (DEA) used today to describe the evolved method developed by Farrell.

The DEA method was later augmented using the works of Samuelson (1938) and Shephard (1953) to highlight the dual relationship between costs and production to provide an envelope method to estimate relative cost efficiency among firms. Färe, Grosskopf, and Lovell (1985) provided a method using the dual cost approach with DEA to estimate cost efficiency. In this case, a cost frontier (minimum) is calculated rather than a production frontier (maximum) and thus efficient firms lie on the frontier, but inefficient firms lie above the frontier.

Aigner, Lovell and Schmidt and Meeusen and Van den Broeck; and Battese and Coelli suggest a method of estimation known as the stochastic frontier estimation based on maximum likelihood. They argue that the stochastic frontier conforms more closely to economic theory building a frontier where the observations of cost lie either on, or above a cost frontier. Like traditional parametric estimation methods, the stochastic frontier method requires the specification of a functional form, and all the assumptions that traditional parametric estimation methods must satisfy remain for the function to be consistent with economic theory. Battese and Coelli have expanded this method to include panel estimation of a stochastic frontier using the software program Frontier V4.1[4].

---

[4] Frontier V4.1 written by Tim Coelli are available online at : http://www.uq.edu.au/economics/cepa/frontier.php

Regression based methods with two-sided errors have been used to envelope the data such as the Conditional Ordinary Least Squares method (COLS) (Greene 2005), and Modified Ordinary Least Squares Method (MOLS) (Afriat 1972). These methods involve either altering the intercept (COLS) or shifting the production/cost function up/down based upon an expected value of the inefficiency distribution (MOLS). These methods are not without challenges and restrictions since the COLS method requires a homoscedastic distribution and the frontier function may not be the same as the minimized sum of squared errors. Also, the MOLS method cannot guarantee that the data is enveloped. A shift or intercept change only affects the calculation of the distance from the frontier calculations but does not affect calculations of marginal costs or incremental costs.

A less investigated parametric method uses OLS, restricting the errors to take on only positive values in the case of a cost function. This method does not require any prior assumptions of distribution and envelopes the data. Further, since it is not a shift, it allows for the marginal cost calculations to be based off of a parametric curve fitted to frontier firms.

The nonparametric approach to frontier estimation has as a few advantages to parametric methods. The most important is that it envelopes the data such that it conforms to economic theory. That is, the cost function is the minimum cost to produce an output bundle (Mas-Colell et al. 1995). As mentioned above, this is a disadvantage to the traditional parametric methods. Another cited advantage is that it does not require the specification of a function and thus is not technologically restrictive. In addition, the nonparametric method does not require the imposition of curvature required for a cost function (Featherstone and Moss 1994).

Recently, studies by Chavas and Aliber (using the dual DEA method shown by Färe et al. 1995) and Chavas et al. (2012) discuss methods for calculating economies of scope. These articles developed nonparametric frontier estimation and associated incremental cost calculations to determine cost savings from producing multiple outputs simultaneously. However, the methods for calculating multi-product and product-specific scale economies nonparametrically are relatively new (Parman et. al.) and have not been compared to other methods. Such a comparison will evaluate the relative efficiency of the nonparametric approach to estimate the economies of scale measures.

This research examines the robustness of four different estimation approaches to evaluate their ability to estimate a "true" cost frontier and associated economic measures. The manuscript will evaluate three parametric methods including a two-sided error system, OLS with only positive errors, and the stochastic frontier method. The fourth method will be the DEA method (Färe et. al.) augmented to calculate multi-product and product-specific economies of scale (Parman et. al.). The robustness of the four estimation methods is examined using simulated data sets from two different distributions and two different observation quantity levels.

Data

The data for the analysis were generated using a modified Monte Carlo procedure found in Gao and Featherstone (2008) run on the SHAZAM software platform with the code found in Appendix A at the end of this document. 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 since it is a self-dual cost function and a flexible functional form (Lusk et al.). The input and output prices (w_i, p_i) are simulated randomly following a normal distribution. The assumed distributions for the output prices and input prices shown below were set to 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:

$$
\begin{aligned}
&w_1 \sim N\,(9,\ 0.99)\\
&w_2 \sim N\,(18,\ 1.98)\\
&w_3 \sim N\,(7,\ 0.77)\\
&p_1 \sim N\,(325,\ 99)\\
&p_2 \sim N\,(800,\ 99)
\end{aligned}
\tag{3.1}
$$

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 the input price w_3 and cost is scaled by w_3 to impose homogeneity. To ensure curvature holds, the "true" cost function is concave in input prices and convex in output quantities. The assumed parameters also satisfy symmetry ($b_{ij}=b_{ji}$). The assumed parameters (Table 3.1) are used to determine the output quantities y_1 *and* y_2[5]. The general form of the normalized quadratic cost function is:

$$
\begin{aligned}
&C(W,Y) = b_0 + \begin{bmatrix} b_1 & b_2 \end{bmatrix}\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} + \begin{bmatrix} a_1 & a_2 \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} + \\
&\frac{1}{2}\left\{ \begin{bmatrix} w_1 & w_2 \end{bmatrix}\begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} + \begin{bmatrix} y_1 & y_2 \end{bmatrix}\begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \right\} + \begin{bmatrix} w_1 & w_2 \end{bmatrix}\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}
\end{aligned}
\tag{3.2}
$$

Output quantities (shown below) are calculated using the assumed parameters of the cost function (Table 3.1) and the output prices generated in equation 3.1.

$$
\begin{aligned}
y_1 &= \frac{c_{22}p_1 - c_{12}p_2 + \left(a_{12}c_{12} - a_{11}c_{22}\right)w_1 + \left(a_{22}c_{12} - a_{21}c_{22}\right)w_2 + \left(a_2c_{12} - a_1c_{22}\right)}{\left(c_{22}c_{11} - c_{12}c_{12}\right)}\\
y_2 &= \frac{c_{12}p_1 - c_{11}p_2 + \left(a_{12}c_{11} - a_{11}c_{12}\right)w_1 + \left(a_{22}c_{11} - a_{21}c_{12}\right)w_2 + \left(a_2c_{11} - a_1c_{12}\right)}{-\left(c_{22}c_{11} - c_{12}c_{12}\right)}
\end{aligned}
\tag{3.3}
$$

[5] The analysis also was completed for alternative assumptions on input.

Using Equation 3.2, 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~N (0,1000)*[6]. The inclusion of this term adds cost inefficiency to the data such that firms are off the frontier effectively increasing their production cost while keeping the output quantities the same. The level of inefficiency is half-normally distributed.

An additional data set[7] is generated assuming a uniform distribution. The uniform deviation ranged from zero to 900. The normal distribution standard deviation of 1,000 generates a mean and standard deviation for cost efficiency roughly equivalent to a uniform distribution with a range from zero to 900.

From equation 3.2, and using Shephard's Lemma where $(\partial C(W,Y)/\partial w_i)=x_i$, 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.

$$
\begin{aligned}
x_1 &= b_1 + b_{11}w_1 + b_{12}w_2 + a_{11}y_1 + a_{12}y_2 \\
x_2 &= b_2 + b_{12}w_1 + b_{22}w_2 + a_{21}y_1 + a_{22}y_2 \\
x_3 &= \left(C\left(W,Y\right) - |e|\right) - x_1w_1 - x_2w_2)
\end{aligned} \tag{3.4}
$$

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 (equation 3.5).

$$x_i^a = \frac{x_i}{CE} \tag{3.5}$$

Using the above method, 400 observations were simulated 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-

---

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

[7] The analysis was run using 2500 observations. The results were robust for 500 and 2500 observations

running the simulation for 50 separate observations each. Thus, a total of 500 observations were generated with summary statistics shown in Table 3.2. In Table 3.2, 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 3.3. Summary statistics for the economic measures are independent of the distribution of cost "inefficiency". Figures 3.1 through 3.4 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.

While the cost efficiency for each firm is altered under a uniform versus a half-normal distribution (Figure 3.2), the MPSE, PSE's, and economies of scope are identical for each data point (Table 3.3) for the "true" cost function due to the input prices (w_i*'s*) and output prices (p_i*'s*) being the same. Thus, the output quantities (y_i*'s*) remain unchanged (Equation 3.3). The input quantities (x_i*'s*) are adjusted such that the deviation in input quantity used by each firm is uniformly distributed. In effect, the uniform data evenly distributes the quantity of firms at each relative distance from the frontier, rather than most firms being clustered around the mean distance as in the half-normal case.

A third data set is simulated using the half-normal distribution. This set uses the same data points as the half-normal case but excludes the single output firms. In this set, there are 400 firms each producing both y_1 and y_2. This data is used to evaluate each method's ability to estimate incremental costs accurately when no zero output firms are observed in the data.

The difference between the "true" estimates and each of the four methods are evaluated. This is done by subtracting each model's estimate from the "true" measure calculated with

Monte Carlo simulation. Since an approximation of the "true" measure is key, the statistics reported are the difference between the "true" measures and what was estimated by each method. Using this approach, any possible bias from each approach can be determined. A positive difference implies that the model underestimates the measure being evaluated and conversely, a negative difference indicates the model overestimates the measure being evaluated. The mean absolute deviation is also reported for all four methods 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.

Estimation Methods

The Two-Sided Error System Equation

The traditional two-sided error system involves specification of a cost function and single frontier of input quantities and costs from observed prices and outputs. This method fits a curve with observations residing both above and below the fitted curve. The two-sided error method for this study was estimated using the SHAZAM software package using a normalized quadratic cost function with input prices normalized on w_3 (equation 3.6).

$$C(W,Y) = b_0 + \begin{bmatrix} b_1 & b_2 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} + \begin{bmatrix} a_1 & a_2 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} + \frac{1}{2}\left\{ \begin{bmatrix} w_1 & w_2 \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} + \begin{bmatrix} y_1 & y_2 \end{bmatrix} \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \right\} + \begin{bmatrix} w_1 & w_2 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} + e_1 \quad (3.6)$$

$$
\begin{aligned}
x_1 &= b_1 + b_{11}w_1 + b_{12}w_2 + a_{11}y_1 + a_{12}y_2 + e_2 \\
x_2 &= b_2 + b_{12}w_1 + b_{22}w_2 + a_{21}y_1 + a_{22}y_2 + e_3
\end{aligned}
\tag{3.7}
$$

Once the parameters shown in Equations 3.6 and 3.7 are estimated, the marginal costs are calculated by:

$$
\begin{aligned}
mcy_1 &= a_1 + c_{11}y_1 + c_{12}y_2 + a_{11}w_1 + a_{21}w_2; and \\
mcy_2 &= a_2 + c_{22}y_2 + c_{12}y_1 + a_{12}w_1 + a_{22}w_2
\end{aligned}
\tag{3.8}
$$

For the normalized quadratic function with two outputs, the incremental costs for each output are:

$$
\begin{aligned}
ICy_1 &= a_1y_1 + \frac{1}{2}c_{11}y_1^2 + c_{12}y_1y_2 + a_{11}w_1y_1 + a_{21}w_2y_1; and \\
ICy_2 &= a_2y_2 + \frac{1}{2}c_{22}y_2^2 + c_{12}y_1y_2 + a_{12}w_1y_2 + a_{22}w_2y_2
\end{aligned}
\tag{3.9}
$$

The costs of producing a single output are:

$$
\begin{aligned}
CY_1 &= C(W,Y) - ICy_2; and \\
CY_2 &= C(W,Y) - ICy_1
\end{aligned}
\tag{3.10}
$$

Once the marginal costs, incremental costs, and single output costs have been estimated, the multi-product scale economies (MPSE), economies of scope (SC), and product-specific scale economies ($PSEy_i$) can be calculated:

$$
MPSE = \frac{C(W,Y)}{\left[\sum_{i=1}^{2} mcy_i * y_i\right]}
\tag{3.11}
$$

$$
SC_i = \frac{\left[\sum_{i=1}^{2} CY_i - C(W,Y)\right]}{C(W,Y)}
\tag{3.12}
$$

$$
PSEy_i = \frac{ICy_i}{y_i * mcy_i}
\tag{3.13}
$$

Cost efficiency is not calculated for the two-sided error system since the deviations from the "frontier" are two sided.

The OLS Estimator with Positive Errors

A one-sided error model is estimated similar to the two-sided error model with the difference being the error term is one sided and input demand equations (3.7) are not estimated. Equation 3.6 is estimated with the restriction that $e_i \geq 0$ for all i using the General Algebraic Modeling Software (GAMS) program. The objective function minimizes the sum of squared errors subject to constraints that define the error. Firms on the frontier exhibit errors equal to zero while those with inefficiency exhibit positive errors. The calculations of MPSE, PSE, and *SC* are identical to the two-sided error model using the coefficient estimates from the one-sided error model.

The Stochastic Frontier Cost Function Estimator

The stochastic frontier estimation method uses FRONTIER Version 4.1 by Coelli. It is based off the stochastic frontier methods of Battese and Coelli (1992, 1995) and Schmidt and Lovell (1979). One of the primary differences between the stochastic frontier method and the OLS two-sided error method is the error term. Specifically, the error term consists of two elements, V_{it} which are random variables assumed to be iid N(0,$\boldsymbol{\sigma}^2$), and U_{it} which is a non-negative random variable capturing inefficiency. U_{it} is assumed to be half-normal for this analysis and defines how far above the frontier a firm operates. The resulting cost function is:

$$
\begin{aligned}
C(W,Y) = b_0 + \begin{bmatrix} b_1 & b_2 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} + \begin{bmatrix} a_1 & a_2 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} + \\
\frac{1}{2}\left\{ \begin{bmatrix} w_1 & w_2 \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} + \begin{bmatrix} y_1 & y_2 \end{bmatrix} \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \right\} + \begin{bmatrix} w_1 & w_2 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} + \left(V_i + U_i\right)
\end{aligned}
\tag{3.14}
$$

For simplicity 3.14 can be rewritten as follows:

$$C(W,Y)_i = X_i B + (V_i + U_i) \tag{3.15}$$

The cost efficiency (CE) from the stochastic frontier method takes on a value between one and infinity since $U_i \geq 0$. The cost efficiency from the nonparametric method and the one-sided error model is evaluated by taking the minimized total costs estimate dividing it by the actual total costs resulting in cost efficiency estimates between 0 and 1.

$$CE = \frac{X_i B}{X_i B + U_i} \tag{3.16}$$

The calculations of marginal costs, incremental costs, the MPSEs, the PSEs, and the economies of scope are the same as those shown in the two-sided error model above using the estimated parameters.

Each of the methods used to estimate the dual cost function are parametric. Symmetry and homogeneity are imposed in the estimation process. Curvature and monotonicity are not and in an empirical estimation they would need to be examined to ensure the cost function estimated is consistent with economic theory.

The Nonparametric Approach

The nonparametric approach for estimating multi-product scale, product-specific scale and scope economies follow Parman et al. (2013). The cost (C_i) is determined for each firm 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^*).

$$\begin{aligned} &\min C_i = w_i' x_i^* \\ &s.t \\ &Xz \le x_i^* \\ &y'z \ge y_i \\ &z_1 + z_2 + \ldots + z_n = 1 \\ &z_i \in \mathbb{R}^+ \end{aligned} \tag{3.17}$$

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 above 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 the p^{th} output constraint.

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

$$CE_i = \left[\frac{C_i}{ATC_i} \right] \tag{3.18}$$

The calculation for economies of scope are:

$$SC_i = \left[\frac{(\sum_p C_{i,p}) - C_i}{C_i} \right] \tag{3.19}$$

The calculation of multi-product economies of scale uses the shadow prices on the output constraints (3.17) to calculate marginal cost. MPSE is then:

$$MPSE_i = \left[\frac{C_{i,all}}{\sum_p MC_{i,p} Y_{i,p}} \right] \tag{3.20}$$

Product specific economies of scale (PSE) require the calculation of the incremental costs ($IC_{i,p}$):

$$IC_{i,p} = C_i - \sum_j C_{i,j \neq p} \forall j \tag{3.21}$$

Average incremental costs ($AIC_{i,p}$) are determined by dividing incremental costs by individual output:

$$AIC_{i,p} = \frac{IC_{i,p}}{y_{i,p}} \tag{3.22}$$

Using the average incremental cost and the marginal cost calculation above, the PSEs are:

$$PSE_{i,p} = \frac{AIC_{i,p}}{MC_{i,p}} \tag{3.23}$$

When estimating the frontier nonparametrically using a data set with no single output firms, it is not possible to estimate the incremental costs by forcing one of the output constraints to zero (Equation 3.17). Thus, the only alternative is to drop one of the constraints. However, when an output constraint is dropped, the program may allow some of the output for the dropped constraint to be produced resulting in an overstatement of the cost of that one output ($C_{i,p}$) which will cause an over statement of economies of scope (equation 3.19) and an understatement of product specific scale economies (3.23).

Thus, the additional product-specific production costs from an output being produced when it shouldn't must be. The procedure for adjusting in a two goods case is as follows: the cost of producing y_1 only ($C_{i,1}$) assumes that only (y_1^1) is being produced. However, the optimization program allows some $y_{i,2}^1$ to be produced in this situation overstating the cost of producing y_1

only ($C_{i,1}$). To remove the additional cost, the percentage contribution of $y_{i,1}^{1}$ to cost is multiplied by the cost of producing y_1 only, yielding an adjusted cost ($C^{a}_{i,1}$). This new adjusted cost is then used in the calculation of incremental costs and associated economic measures:

$$C^{a}_{i,1} = C_{i,1}\left(\frac{y^{1}_{i,1}}{y^{1}_{i,1} + y^{2}_{i,2}}\right) \tag{3.24}$$

The analysis evaluates the difference between the “true” measures of cost efficiency, economies of scope (scope), multi-product scale economies (MPSE), and product-specific economies of scale (PSE) from the four modeling approaches. The statistics and results presented are not the economic measure calculations but the difference between the model estimates and the “true” measure.

The parametric estimators are specified knowing the “true” functional form: the normalized quadratic cost function. Therefore, the differences may represent a “best case scenario” for each parametric method in that the true functional form is known with only the parameter estimates being unknown.

Results

Table 3.4 shows the parameter estimates and standard errors for the parametric methods for all three data sets. The parameter estimates from each method were different under the same distributional assumptions, and different for the same method under different distributional assumptions with the exception of the OLS positive errors model which yielded the same parameter estimates for the uniform and half-normal distributions. For both the two-sided error system, and the stochastic frontier estimation, different distributional assumptions yielded changes in magnitude as well as sign changes for various parameter estimates. Also, when comparing the 500 observation half-normal case to the 400 observation half-normal with no zero outputs case, there were changes for all three estimation methods as well as changes in magnitude for the estimated parameters. The calculation for the standard errors using GAMS was conducted using the method from Odeh et. al. (2010).

Curvature was checked for each estimation method and each simulation to ensure that it was not violated (Table 3.5). A curvature violation implies that the shape of the cost frontier estimation does not conform to the "true" cost function which is known in this case, and that it violates the economic theory of the cost function. To check these conditions, 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). Each parametric model violated curvature in every simulation for either the "*b*" or "*c*" matrices or both. The one-sided error model and the two-sided system violated curvature of both the "*b*" and "*c*" matrices for the 400 observations simulation.

Cost Efficiency

Cost efficiency differences evaluate each model's ability to estimate the frontier since it is the ratio of estimated minimum cost to actual total cost. The two-sided error model was not examined because it does not estimate a frontier. The OLS Positive Errors and Nonparametric models performed well for all three data sets in estimating the frontier with average differences below 0.03 in absolute value and standard deviations below 0.04 (Table 3.6). The most accurate estimation of cost efficiency was the nonparametric model under the uniform distribution simulation with the average, standard deviation, and mean absolute deviation close to zero.

The stochastic frontier method performed almost as well under the half-normal simulation with the average closest to zero, and under the 400 observation simulation with an average difference of -0.028 but much worse under the Uniform simulation (Figure 3.5) with an average difference of -0.198, mean absolute deviation of 0.198, and standard deviation of 0.118. This implies that estimating efficiency measures with the stochastic frontier method may be dependent on the correct assumption of the error distribution.

In all cases, the average differences were below zero implying that the OLS positive errors, Stochastic Frontier, and Nonparametric models slightly over estimated the cost efficiencies for most of the firms. This is confirmed by examining the mean absolute deviation in the uniform and 400 observations cases being the same the absolute value of the mean. This is expected given the simulation procedure. Frontier methods envelope the observed data, thus cost efficiencies are overestimated unless there are a significant number of firms where the simulated error is zero. However, the averages were close to zero in most cases with low standard deviations.

Economies of Scope

Differences in estimates of economies of scope for the four different methods raised more issues than the cost efficiency estimates. For the half-normal and uniform simulations, the two-sided error system had an average furthest from zero at -0.30 in with a standard deviation similar to the other methods (Table 3.7). For the 400 observation simulation, the Stochastic Frontier Method was furthest from zero at -2.32. Due to scaling, the stochastic frontier method cumulative density is not visible in Figure 3.6 for the 400 observations case.

The OLS Positive Errors Model and Nonparametric Model estimated economies of scope closely with averages for the half-normal distribution of -0.08 and -0.09 respectively and standard deviations around 0.07 and 0.03 respectively (Table 3.7). The estimates of scope for the uniform distribution from the OLS Positive Errors Model and Nonparametric Model were less than 0.02 in absolute value with low standard deviations. The average and standard deviation for the Nonparametric method under the uniform distribution were affected by a few observations being significantly off (Figure 3.6). For the 400 observation data set, the Nonparametric method had the lowest standard deviation (0.04) and an average closest to zero in absolute value (0.07) (Table 3.7).

The three parametric estimation methods over estimated economies of scope in all simulations except for the case of a normal distribution where the OLS Positive Errors Model under estimated economies of scope slightly. In many cases, the parametric methods strictly over estimated scope where the absolute values of the means were the same as the mean absolute deviations (Table 3.7). The Nonparametric Model slightly over estimated scope in both the half-normal and uniform simulations but slightly underestimated scope in the 400 observations data set.

The most robust estimator of economies of scope appears to be the Nonparametric approach with averages close to zero in all three simulations and low standard deviation. The OLS Positive Errors Model does not perform as well in the case of 400 observations simulation, nor does the Stochastic Frontier Model and the standard Two-sided Error System under the half-normal and uniform simulations. Measures of economies of scope are suspect using any of the methods when there are no zero output observations in the data sample.

Multi-product Economies of Scale

An accurate estimation of MPSE requires both a close approximation of the true frontier and marginal costs. It is possible to have a very good approximation of MPSE but be off on economies of scope and PSEs due to the necessary estimation of incremental costs for scope and the PSEs.

The nonparametric approach appears to be the most robust estimator of MPSE (Figure 3.7). It has an average difference closest to zero in all three simulations and the lowest standard deviation in both the half-normal case and 400 observation cases (Table 3.8). Its mean absolute deviation is also lowest except compared to the OLS Positive Errors model under the uniform distribution. The standard deviation was only slightly higher for the nonparametric approach compared to the OLS Positive Errors model in the uniform case with a standard deviation of 0.05 for the Nonparametric model and 0.04 for the OLS Positive Errors model (Table 3.8). All average differences except OLS Positive Errors in the uniform case were negative implying that MPSE was, for the most part, over estimated by the models.

Of the four modeling methods in all three simulations, the two-sided error system had the largest average differences from zero and the highest standard deviations (Table 3.8). No observations were correctly estimated for MPSE (Figure 3.7) in any of the three simulations. The standard two-sided system approach never approaches the zero difference.

The Stochastic Frontier method results were mixed. While it was out performed by the nonparametric approach in all simulations, it was close to the "true MPSE' in the case of the 400 observations. However, in the uniform distribution simulation, it did not perform well with an average difference of -0.21 and standard deviation of 0.26 (Table 3.8).

Product-Specific Economies of Scale

The estimation of the PSEs for both y_1 and y_2 for the half-normal and uniform simulations yielded similar results for all three parametric type estimations (Table 3.9). The parametric approaches appear to slightly outperform the nonparametric approach in the estimation of PSE_1 (Figure 3.8 Panel A) in the half-normal simulation but performed similarly in the estimation of PSE_2 (Figure 3.8 Panel B) under the same distribution in terms of absolute distance from zero. For the uniform simulation, the PSE_1 and PSE_2 estimates from the Nonparametric Model were similar to both the Stochastic Frontier Method and the two-sided error systems with the OLS Positive Errors Model being the closest to zero under the uniform simulation (Table 3.9).

Under the half normal and the uniform simulations, the two-sided error system and the stochastic frontier underestimated PSE's for y_1 and y_2. OLS Positive Errors under estimated PSEs under both distributions except for the half-normal PSE_1. In the 400 observation simulation, OLS overestimated both PSEs where that was not the case for the Nonparametric Model and OLS Positive Errors Model.

The average difference and standard deviation for the PSEs from the Stochastic Frontier Method in the 400 observation simulation are off significantly (Table 3.9). Of the parametric methods, it appears that two-sided error system performed best when there were no single output firms having the lowest standard deviations and averages fairly close to zero, especially for PSE_2 (Figure 3.9 Panel B).

In the 400 observations simulation, while the standard deviation was higher for the nonparametric method than OLS and OLS Positive Errors, the average for PSE_1 was closest to zero using the nonparametric method and closer than OLS Positive Errors and the Stochastic

Frontier Method for PSE_2 (Table 3.9). None of the methods accurately predict the PSEs when there were no zero observations (Figures 3.8 and 3.9, panel C).

The challenge for each model in the 400 observations simulation is that there are no firms producing only a single output. This requires each method to extrapolate estimates out of sample for the purpose of calculating incremental costs. If the smallest firms are not efficient, a linear projection will be inaccurate depending on the amount of inefficiency of the smaller firms.

Implications of the Results

Results suggest the two-sided error system is least accurate for estimating a frontier function and associated cost measures. This method lacks consistency with the economic definition of a cost function. This is apparent in that it does not, in any simulation, robustly estimate the MPSE or the economies of scope.

The stochastic frontier method appears susceptible to incorrect distributional assumptions on the one sided error as it estimates the frontier much closer to the "true" frontier under a half-normal distribution rather than the uniform distribution. Results also suggest that the stochastic frontier method has difficulty extrapolating when there are no zero output firms as shown by its inability to accurately estimate economies of scope or PSEs for the 400 observations simulation. However, in the case of a normal distribution it accurately estimates the frontier and, with the existence of zero output observations, accurately estimates economies of scope and PSEs.

The OLS positive errors model appears to accurately project the cost frontier regardless of the distributional assumption and whether there are no single output firms. However, like the stochastic frontier method, the OLS positive errors method has difficulty extrapolating when there are no single output firms. Thus, under no single output cases, the economies of scope estimations from the positive errors model may be incorrect, as may be PSE estimates.

The nonparametric method in all three simulations is fairly robust in estimating the "true" cost frontier and associated economic measures. It is also the model most capable of handling data with no single output firms as shown by its proximity to zero in estimating economies of scope and PSEs. It does not appear to be particularly susceptible to distributional assumptions.

It is important to remember that all of the parametric methods used the correct function form (normalized quadratic). These results may be different should the data not be consistent with that functional form. Functional form and statistical assumptions are not necessary in the case of the nonparametric method, thus, the results will likely be more robust. Therefore, if a researcher is unsure of model specification or the data generation process, the nonparametric approach may be a good alternative to parametric estimation.

Conclusions

Four methods for estimating a cost frontier and associated economic measures were examined under three different simulations including a half-normal distribution, uniform distribution, and a data set with no single output firms. The first method examined was a traditional two-sided error system regression with costs residing above and below the fitted curve. The second was the stochastic frontier method initially proposed by Aigner, Lovell, and Schmidt where the error term ensures all observations lie on or above the cost frontier. The third method was an OLS regression where the error term was restricted to take on positive values only ensuring that all observations lie on or above the cost frontier. Finally, a nonparametric DEA method proposed by Färe et. al. using a series of linear segments was used to trace out a cost frontier. For each simulation, cost efficiency, economies of scope, multi-product scale economies, and product specific scale economies were calculated and compared to the known values from the “true” cost frontier.

The results show that the three frontier estimators are capable of estimating the “true” frontier in some simulations however; the stochastic frontier method was not as robust as the nonparametric method or the OLS Positive Errors Model. This result was also observed in the calculation of multi-product scale economies where all three frontier functions estimated the 400 observations data set and the half-normal data set close, whereas the Stochastic Frontier Model was not. The OLS method could not estimate a frontier and corresponding cost efficiency and was the furthest from the “true” calculation of multi-product scale economies indicating it was not close in estimating marginal cost.

The Stochastic Frontier Model appears to be less robust for estimating the “true” measures that require calculating incremental costs such as economies of scope and product-specific scale economies. Though the two-sided error model was less accurate in obtaining the “true” estimates in the half-normal and uniform simulations, the Stochastic Frontier method was less accurate in estimating scope economies or product specific economies of scale when there were no single output firms.

Overall, the nonparametric approach estimated the frontiers and associated economic measures close to the “true” values considering no special assumptions or specifications were required in its estimation. It’s estimation of the frontier was about as close, or closer to the “true” values as any of the methods examined and its calculations of MPSE and economies of scope were the closest in several of the scenarios presented. The nonparametric approach did not significantly fail to estimate PSEs compared to any other method. Therefore, it appears that the DEA method is robust for estimating scale and scope measures.

Chapter 3 References

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

Atkinson, Scott E. and Robert Halvorsen. “Parametric Efficiency Tests, Economies of Scale, and Input Demand in U.S. Electric Power Generation” *International Economic Review.* 25(4)(October 1984):647-662.

Afriat, S.N. “Efficiency Estimation of Production Functions.” *International Economic Review*, 13, 3(1972): 568-598.

Banker, Rajiv D. and Ajay Maindiratta. “Nonparametric Analysis of Technical and Allocative Efficiencies in Production.” *Econometrica*, 46,69 (November 1988): 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

Battese, G. and T. Coelli. “Prediction of Firm-level Technical Efficiencies with a Generalized Frontier Production Function and Panel Data,” *Journal of Econometrics*, 38(1988). 387-399.

Battese, G., and T Coelli, 1992 “ Frontier Production Functions, Technical Efficiency and Panel Data: With Application to Paddy Farmers in India,” *Journal of Productivity Analysis*, 3(1992): 153-169.

Charnes, A., W.W. Cooper, and E. Rhodes. “ Measuring the Efficiency of Decision Making Units.” *European Journal of Operational Research*, 9(1978): 181-185.

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

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

Chavas, Jean-Paul, Bradfor Barham, Jeremy Foltz, and Kwansoo Kim. “ Analysis and Decomposition of Scope Economies: R&D at US Research Universities.*” Applied Economics*, 44(2012): 1387-1404.

Coelli, T. (1991)“Maximum Likelihood Estimation of Stochastic Frontier Production Functions with Time Varying Technical Efficiency Using the Computer Program FRONTIER Version 2.0” Department of Economics, University of New England, Armidale, Australia

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

Diewert, W.E. and T.J. Wales. "A Normalized Quadratic Semiflexible Functional Form" *Journal of Econometrics*, 37, 3(1988): 327-342

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

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

Farrell, M.J. and M. Fieldhouse. "Estimating Efficient Production Under Increasing Returns to Scale." *Journal of the Royal Statistical Society*, 125, 2(1962): 252-267.

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

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

Greene, William. (2005) "Efficiency of Public Spending in Developing Countries: A Stochastic Frontier Approach." Stern School of Business, New York University.

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

Mas-Colell, Andreu, Michael D. Whinston, and Jerry R. Green. 1995. Microeconomic Theory. Oxford University Press. New York, NY.

Meeusen, W. and J. Van den Broeck. "Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error", *International Economic Review*, 18(1977): 435-444.

Odeh, Oluwarotimi O, Allen M. Featherstone, and Jason S. Bergtold. "Reliability of Statistical Software" *American Journal of Agricultural Economics*, 92(5): 1472-1489.

Parman, Bryon J., Allen M. Featherstone, and Brian K. Coffey "A Nonparametric Approach to Multi-product and Product Specific Economies of Scale" Working Paper (2013)

Samuelson, P., 1938. Foundations of Economic Analysis, Harvard University Press, Cambridge

Shephard, R.W. 1953. Cost and Production Functions, Princeton University Press, Volume 1, Issue 2, Princeton NJ.

Chapter 3 Tables

Table 3.1 Assumed 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}	0.6
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 3.2 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 3.3 Summary statistics for efficiency calculations from generated data including half-normal and uniform distributions.

Economic Measure	Average	Standard Deviation	Minimum	Maximum
------Half-normal Distribution------				
Multi-product Scale Economies	0.931	0.108	0.772	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	1.000
------400 Observations------				
Multi-product Scale Economies	0.918	0.082	0.773	1.989
Cost Efficiency	0.751	0.159	0.129	1.000
Scope	0.085	0.028	0.062	0.514
Product-specific Scale Economies for y1	0.808	0.047	0.678	0.957
Product-specific Scale Economies for y2	0.848	0.041	0.733	0.996

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

Table 3.4 Parameter estimates and standard errors for three simulations of each parametric model

	N= 500 Half-Normal Distribution			N=500 Uniform Distribution			N=400 No Zero Outputs		
	Two-sided Errors	One-sided Errors	Stochastic Frontier	Two-sided Errors	One-sided Errors	Stochastic Frontier	Two-sided Errors	One-sided Errors	Stochastic Frontier
A_1	28.83	56.05	32.00	29.77	56.05	28.90	60.92	76.82	302.28
	(4.67)	(3.61)	(27.32)	(2.05)	(1.88)	(3.74)	(23.12)	(5.24)	(12.03)
A_2	79.33	88.39	46.05	80.18	88.39	78.24	52.38	54.74	-221.34
	(3.83)	(2.75)	(21.10)	(1.62)	(1.43)	(4.41)	(17.58)	(4.40)	(10.46)
A_{11}	0.49	-9.73	2.91	0.47	-9.73	3.04	1.17	24.21	-45.54
	(0.05)	(2.42)	(19.55)	(0.04)	(1.26)	(1.85)	(0.39)	(3.51)	(22.64)
A_{12}	0.67	-3.83	-7.58	0.56	-3.83	1.31	1.91	-16.74	-88.03
	(0.19)	(1.39)	(11.78)	(0.08)	(0.72)	(1.57)	(0.75)	(1.78)	(21.90)
A_{21}	0.54	-5.00	1.69	0.76	-5.00	4.45	0.12	-34.24	65.82
	(0.69)	(1.71)	(14.72)	(0.03)	(0.89)	(1.59)	(0.29)	(2.71)	(18.24)
A_{22}	-1.16	0.44	5.61	-0.39	0.44	1.66	-1.67	-0.99	78.41
	(0.15)	(1.13)	(9.84)	(0.07)	(0.58)	(1.26)	(0.59)	(1.50)	(18.03)
B_0	684.95	360.20	-2011.12	401.01	360.20	460.56	689.81	-194.91	-4270.54
	(60.23)	(100.29)	(42.22)	(26.66)	(52.15)	(1.34)	(80.79)	(42.64)	(1.82)
B_1	26.18	-267.67	1833.84	20.20	-267.67	12.97	25.65	103.79	3225.37
	(26.17)	(104.00)	(107.89)	(0.75)	(54.07)	(1.17)	(1.74)	(46.62)	(8.11)
B_2	70.13	-200.80	825.29	57.29	-200.80	48.74	68.61	56.44	2109.47
	(5.25)	(66.17)	(104.11)	(2.34)	(34.40)	(1.71)	(4.74)	(26.04)	(6.91)
B_{11}	0.34	46.59	-834.08	-0.09	46.59	-65.45	0.12	-210.43	-529.61
	(0.18)	(89.08)	(74.46)	(0.08)	(46.36)	(1.00)	(0.17)	(34.17)	(8.11)
B_{12}	0.83	143.01	-297.13	0.20	143.01	-28.37	0.22	195.80	-1076.81
	(0.51)	(34.06)	(62.21)	(0.23)	(20.21)	(1.05)	(0.50)	(14.26)	(10.09)
B_{22}	2.85	27.48	-135.50	0.64	27.48	-8.34	0.25	2.74	-283.21
	(1.88)	(26.88)	(114.14)	(0.87)	(14.90)	(1.17)	(1.80)	(10.34)	(13.61)
C_{11}	2.29	1.64	2.20	1.76	1.64	2.05	-0.35	-1.41	22.91
	(0.15)	(0.18)	(1.25)	(0.06)	(0.09)	(0.50)	(1.67)	(0.64)	(13.05)
C_{12}	-0.78	-0.25	0.10	-0.48	-0.25	-0.75	1.40	0.013	-15.66
	(0.53)	(0.07)	(0.35)	(0.02)	(0.04)	(0.19)	(1.25)	(0.53)	(8.36)
C_{22}	3.13	2.41	3.16	2.66	2.41	2.38	1.32	6.40	14.66
	(0.68)	(0.12)	(0.78)	(0.04)	(0.06)	(0.32)	(0.97)	(0.43)	(5.63)

Table 3.5 Eigenvalues for "B" (prices) and "C" (outputs) matrices for each model and simulation

Half-Normal		Uniform		No Zero Outputs	
B	C	B	C	B	C
------*Two-sided Error System*------					
3.09	3.50	0.70	2.80	0.41	2.10
0.09	1.80	-0.10	1.50	-0.40	-1.14
X	√	**X**	√	**X**	**X**
------*Stochastic Frontier*-------					
-37.0	3.20	3.10	2.90	677	34.8
-931	-2.20	-76.0	1.40	-1489	2.60
√	**X**	**X**	√	**X**	√
------*OLS Positive Errors*------					
180	2.40	180	2.40	118	6.00
-106	1.50	-106	1.50	-325	-1.50
X	√	**X**	√	**X**	**X**

Note: The known cost function is concave in prices (B matrix) and convex in outputs(C matrix). For concavity, the matrix must yield negative eigenvalues and for convexity the matrix must yield positive eigenvalues. A "√" implies correct curvature while "X" implies a curvature violation.

Table 3.6 Statistics for simulated cost efficiency differences for the OLS positive errors, stochastic frontier, and nonparametric estimations

	Average	Standard Deviation	Minimum	Maximum	Mean Absolute Deviation
		------Half-normal Distribution------			
OLS Positive Errors	-0.020	0.039	-0.277	0.063	0.023
Stochastic Frontier	-0.015	0.043	-0.304	0.155	0.024
Nonparametric	-0.025	0.041	-0.530	-0.003	0.026
		------Uniform Distribution------			
OLS Positive Errors	-0.011	0.013	-0.062	0.122	0.013
Stochastic Frontier	-0.198	0.118	-1.478	-0.058	0.198
Nonparametric	-0.004	0.007	-0.079	0.000	0.004
		------400 Observations------			
OLS Positive Errors	-0.017	0.023	-0.173	0.015	0.017
Stochastic Frontier	-0.028	0.049	-0.351	0.139	0.039
Nonparametric	-0.022	0.032	-0.386	-0.003	0.022

Table 3.7 Statistics for economies of scope differences from all four methods from all three data sets.

	Average	Standard Deviation	Minimum	Maximum	Mean Absolute Deviation
		------*Half-normal Distribution*------			
Two-sided Error System	-0.300	0.057	-0.474	-0.194	0.300
OLS Positive Errors	-0.082	0.067	-0.291	0.093	0.088
Stochastic Frontier	-0.101	0.056	-0.266	0.052	0.103
Nonparametric	-0.089	0.030	-0.249	0.029	0.089
		------*Uniform Distribution*------			
Two-sided Error System	-0.300	0.058	-0.489	-0.191	0.300
OLS Positive Errors	0.010	0.023	-0.044	0.190	0.018
Stochastic Frontier	-0.158	0.041	-0.312	-0.091	0.158
Nonparametric	-0.019	0.079	-0.904	0.017	0.020
		------*400 Observations*------			
Two-sided Error System	-0.148	0.115	-0.437	0.152	0.175
OLS Positive Errors	-0.187	0.053	-0.376	-0.048	0.187
Stochastic Frontier	-2.324	0.607	-4.109	-0.142	2.324
Nonparametric	0.070	0.036	0.025	0.514	0.070

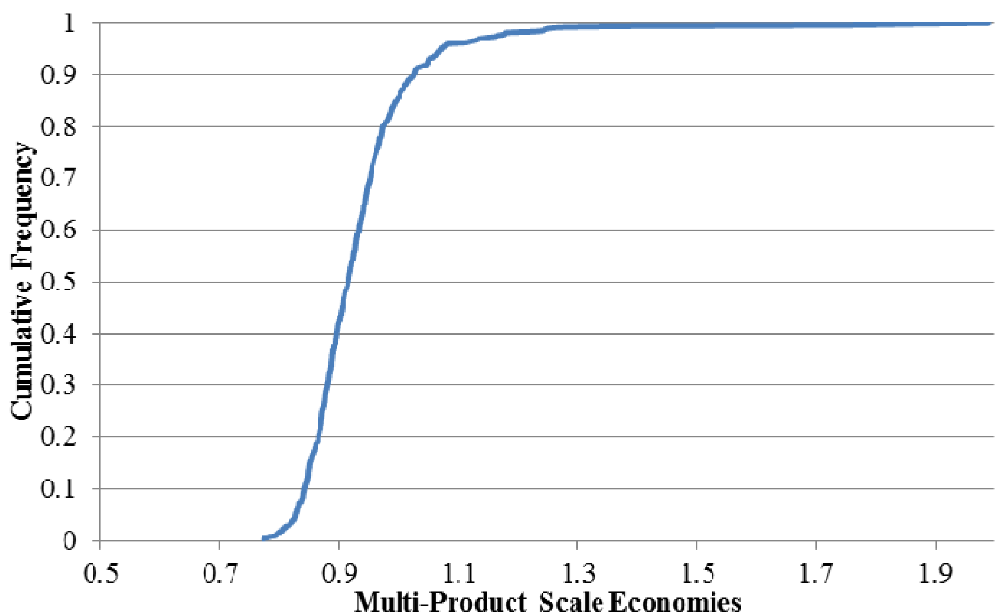
Table 3.8 Statistics for Multi-product Scale Economies differences from all four methods from all three data sets.

	Average	Standard Deviation	Minimum	Maximum	Mean Absolute Deviation
		------Half-normal Distribution------			
Two-sided Error System	-0.443	0.361	-2.917	-0.067	0.443
OLS Positive Errors	-0.257	0.658	-5.995	0.104	0.272
Stochastic Frontier	-0.084	0.183	-1.577	0.107	0.107
Nonparametric	-0.002	0.096	-0.678	0.739	0.049
		------Uniform Distribution------			
Two-sided Error System	-0.482	0.609	-7.857	-0.068	0.482
OLS Positive Errors	0.023	0.044	-0.124	0.515	0.027
Stochastic Frontier	-0.210	0.258	-3.384	-0.030	0.210
Nonparametric	-0.012	0.054	-0.277	0.114	0.029
		------400 Observations------			
Two-sided Error System	-0.371	0.300	-4.642	-0.065	0.371
OLS Positive Errors	-0.087	0.131	-1.331	0.105	0.103
Stochastic Frontier	-0.074	0.152	-0.727	0.177	0.118
Nonparametric	-0.009	0.137	-0.821	0.743	0.058

Table 3.9 Statistics for Product-specific Scale Economies differences for y_1 and y_2 from all four methods and all three data sets

		Average	Standard Deviation	Minimum	Maximum	Mean Absolute Deviation
		------Half-normal Distribution------				
	Two-sided Error System	0.099	0.024	0.047	0.147	0.099
y1	OLS Positive Errors	-0.056	0.039	-0.130	0.140	0.064
	Stochastic Frontier	0.111	0.018	0.081	0.176	0.111
	Nonparametric	0.128	0.202	-0.259	0.573	0.202
	Two-sided Error System	0.053	0.011	0.022	0.081	0.053
y2	OLS Positive Errors	0.075	0.016	0.037	0.148	0.075
	Stochastic Frontier	0.060	0.006	0.004	0.072	0.060
	Nonparametric	0.027	0.099	-0.344	0.303	0.085
		------Uniform Distribution------				
	Two-sided Error System	0.098	0.025	0.024	0.147	0.098
y1	OLS Positive Errors	0.012	0.013	-0.012	0.049	0.013
	Stochastic Frontier	0.056	0.018	0.012	0.093	0.056
	Nonparametric	0.020	0.123	-0.241	0.299	0.097
	Two-sided Error System	0.052	0.011	0.002	0.075	0.052
y2	OLS Positive Errors	0.005	0.002	0.000	0.011	0.005
	Stochastic Frontier	0.004	0.004	-0.008	0.012	0.005
	Nonparametric	-0.004	0.053	-0.186	0.126	0.039
		------400 Observations------				
	Two-sided Error System	-0.218	0.054	-0.368	-0.048	0.218
y1	OLS Positive Errors	-0.421	0.296	-4.110	-0.066	0.421
	Stochastic Frontier	1.696	32.94	-619.9	120.5	6.480
	Nonparametric	0.187	0.310	-0.311	0.957	0.266
	Two-sided Error System	-0.036	0.015	-0.086	0.000	0.036
y2	OLS Positive Errors	0.302	0.028	0.042	0.373	0.302
	Stochastic Frontier	-9.224	248.9	-4952	444.3	18.15
	Nonparametric	0.132	0.209	-0.248	0.937	0.183

Chapter 3 Figures



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

Figure 3.1 Frontier Multi-Product Scale Economies Cumulative Frequency for Simulated Data.

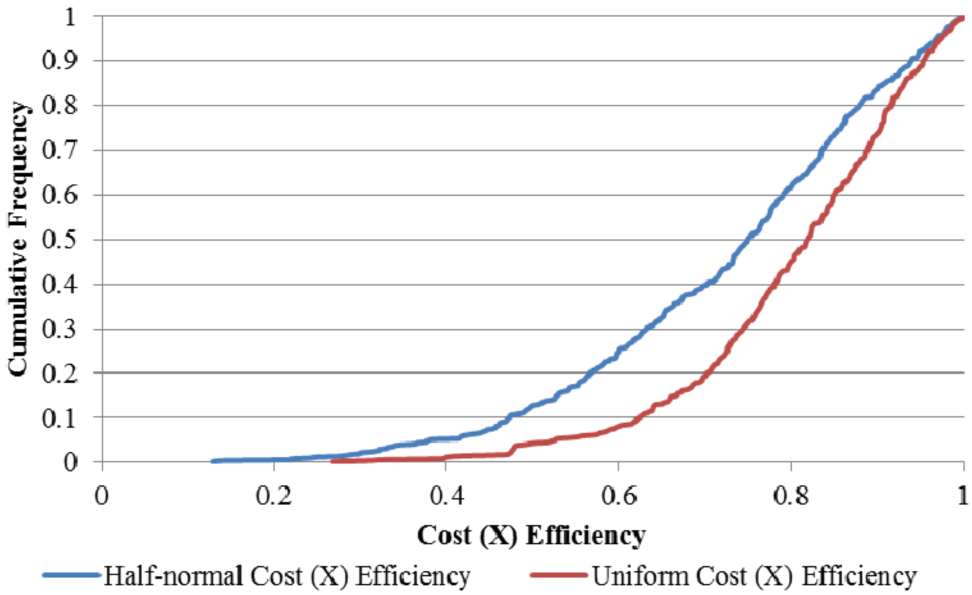


Figure 3.2 Frontier Cost Efficiencies Cumulative Frequency for both Half-normal and Uniform Distributions.

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

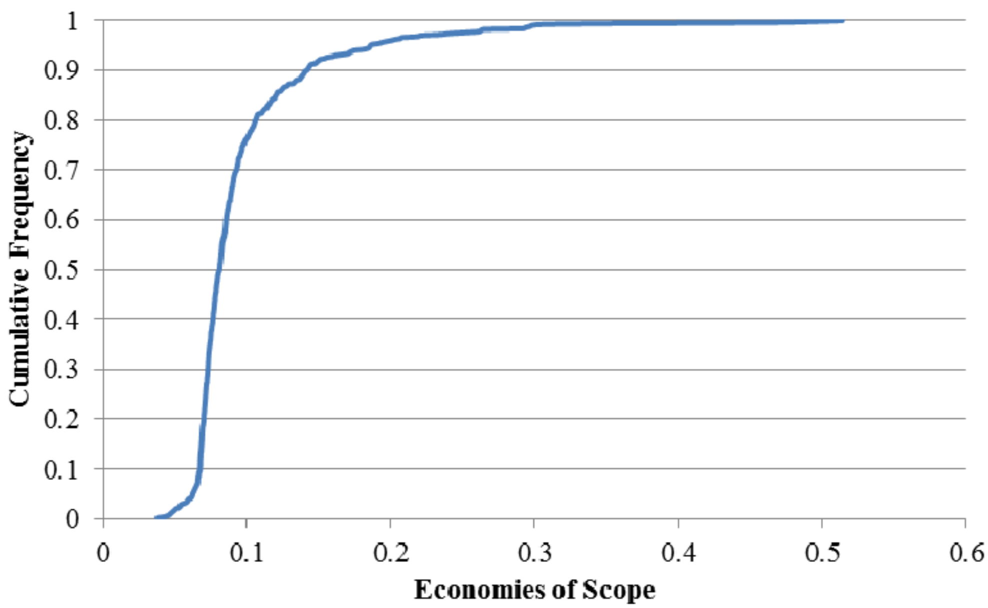


Figure 3.3 Frontier Economies of Scope Cumulative Frequency

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

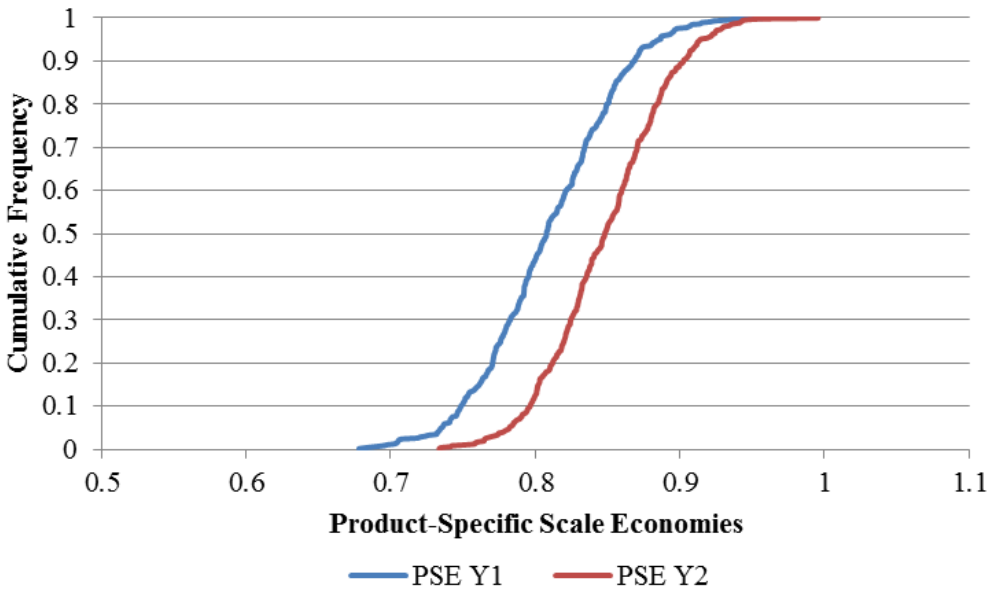


Figure 3.4 Frontier Product-Specific Scale Economies

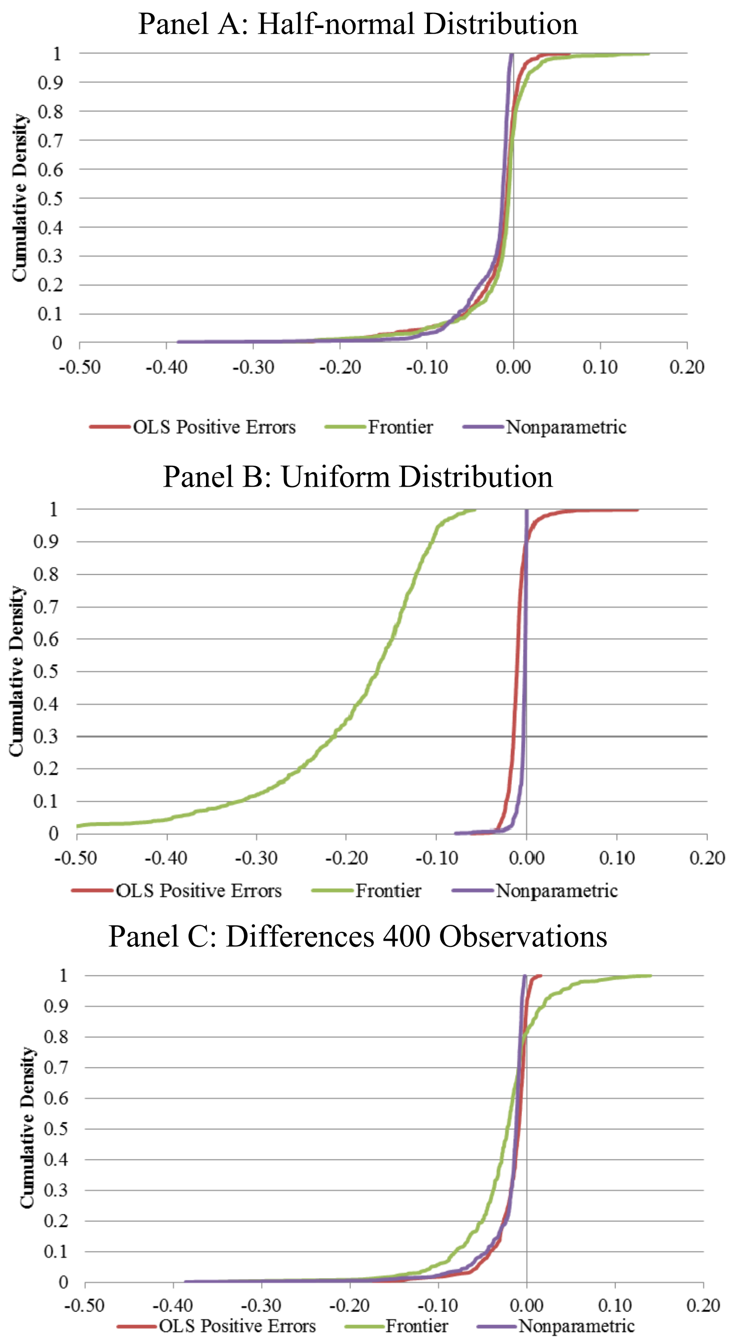


Figure 3.5 Differences between frontier cost efficiency and estimated cost efficiency for the nonparametric, frontier, and OLS positive errors models

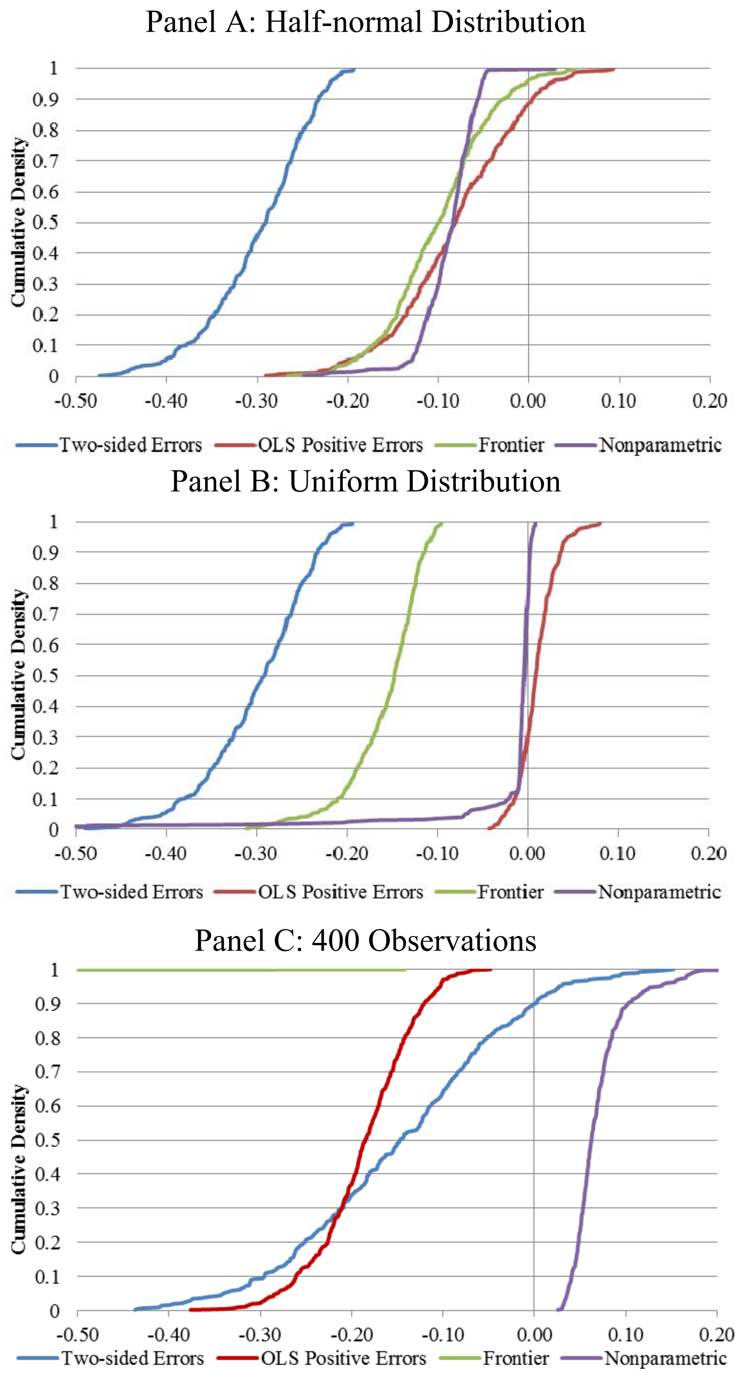


Figure 3.6 Differences between frontier Economies of Scope and estimated Economies of Scope from Two-sided Errors, OLS Positive Errors, Frontier, and Nonparametric models.

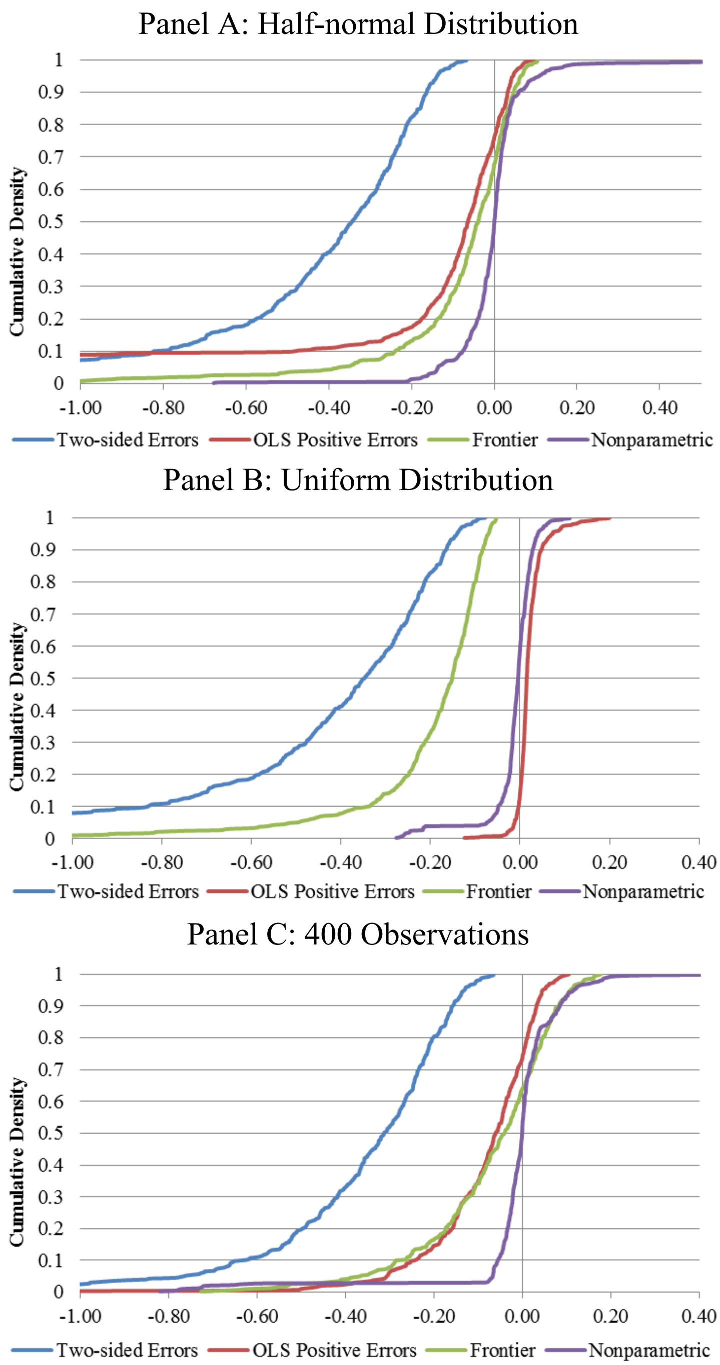


Figure 3.7 Differences between frontier Multi-product Scale Economies and estimated Multi-product Scale Economies from the Two-sided Errors, OLS Positive Errors, Frontier, and Nonparametric models.

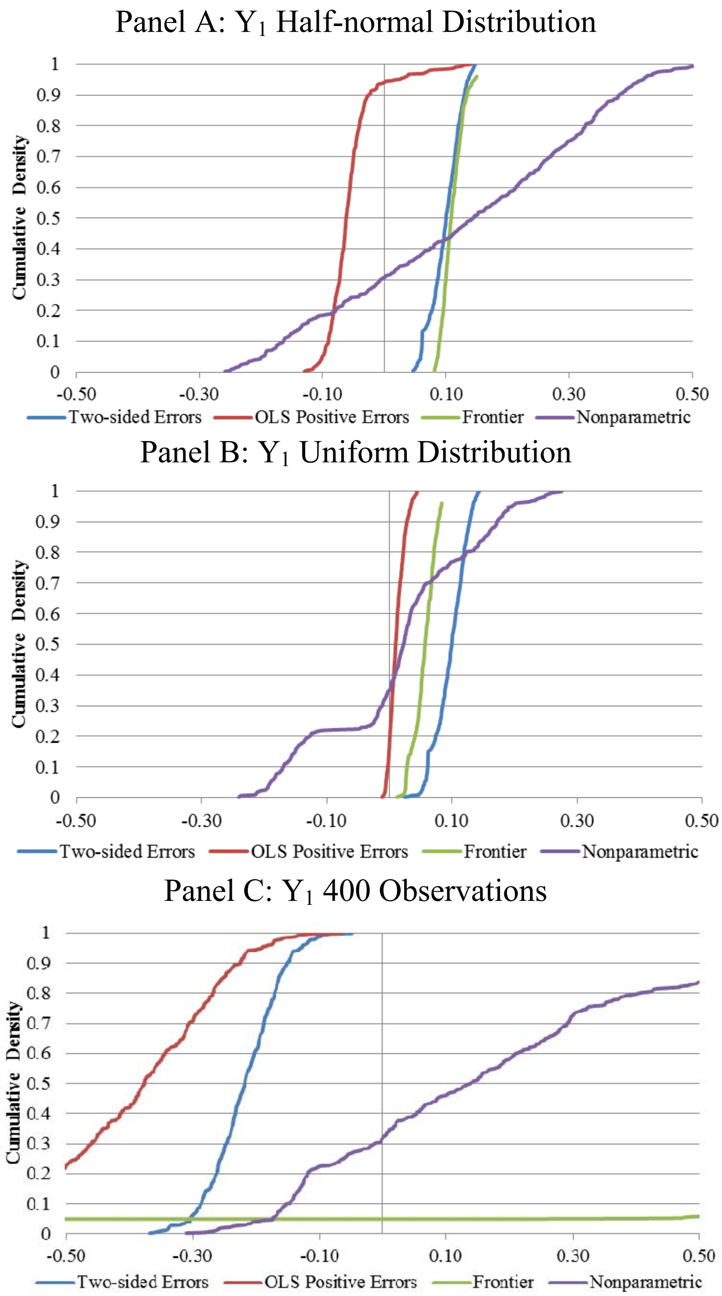


Figure 3.8 Differences between frontier Product-specific Scale Economies for Y_1 and estimated Product-specific Scale Economies for Y_1 from the Two-sided Errors, OLS Positive Errors, Frontier, and Nonparametric models.

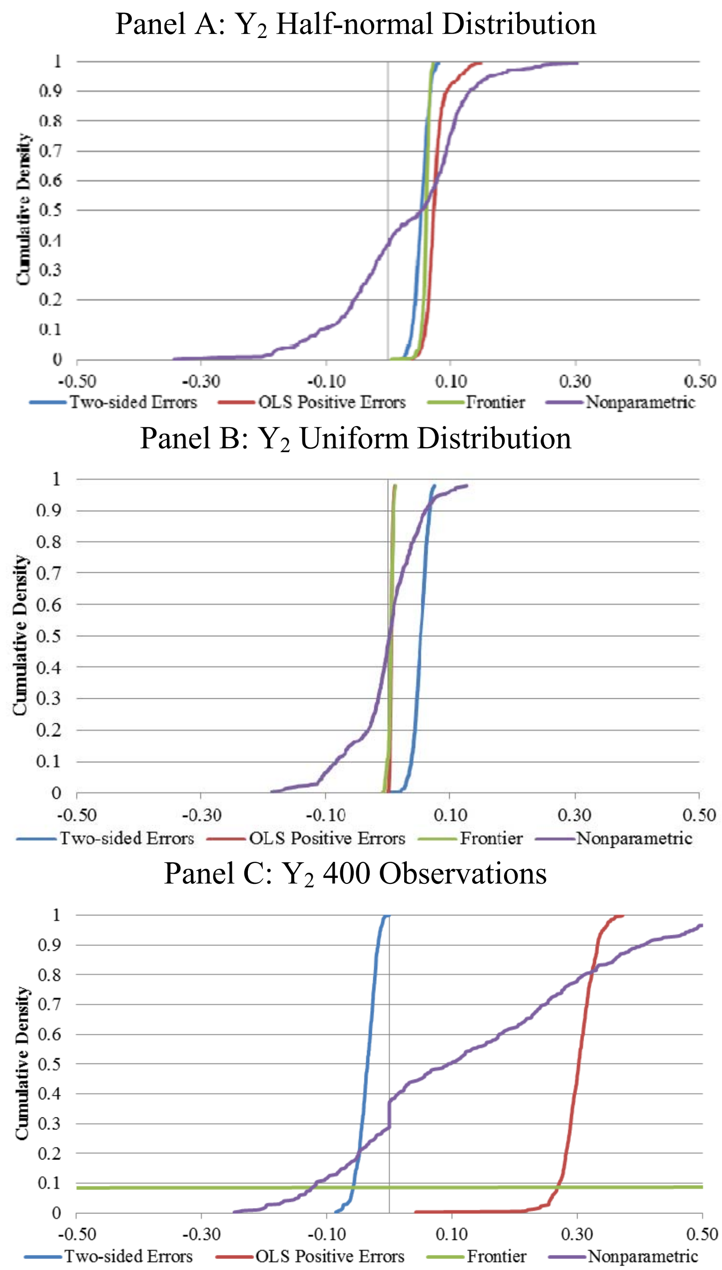


Figure 3.9 Differences between frontier Product-specific Scale Economies for Y_2 and estimated Product-specific Scale Economies for Y_2 from the Two-sided Errors, OLS Positive Errors, Frontier, and Nonparametric models.

Chapter 4 - A Nonparametric Approach to Multi-product and Product-specific Scale Economies, Economies of Scope, and Cost Efficiency for Kansas Farms

Introduction

Reducing costs through understanding economies of scale and economies of scope is fundamental in producer theory. Estimation of cost functions has allowed the measurement of economies of scale and scope using duality theory developed by Samuelson (1938) and Shephard (1953), and later work from Baumol et. al. (1982). The full employment of land, labor, and capital make this particularly important in agriculture where resource expansion is costly. Knowing where potential cost savings exist, and for which farms, provides economists and producers with valuable information as they make investment decisions. However, typically scale and scope measures are not calculated from cost frontiers.

In instances where frontiers have been estimated parametrically (Atkinson and Halversen 1984), a common method used is the stochastic frontier developed by Aigner et al (1977). This method has since been modified to estimate cost frontiers (Coelli and Battese 1994). Using the stochastic frontier cost method, Mafoua and Hossian (2001) examined economies of scale and economies of scope cost savings with a multi-product analysis of corn and soybeans using a panel data set.

An alternative method for frontier estimation uses a series of linear segments to envelope the data (Farrell 1957). From Farrell's original model, the nonparametric method evolved until Färe, Grassokopf and Lovell (1985) formally established a linear program for cost frontier estimation. Chavas and Aliber (1993) used the non-parametric cost frontier method to estimate economies of scope by measuring cost savings from multiple outputs instead of producing each output individually using the data envelope analysis (DEA).

The DEA approach to cost frontier estimation has some attractive advantages over parametric methods. Particularly advantageous is that there is not a need to specify a potentially technologically restrictive functional form. The nonparametric approach is consistent with economic theory by ensuring curvature of the cost function is not violated during the estimation process. Lusk et al. examined the relative variability needed in the estimation of dual cost functions to recover the underlying technology. They found that the relative variability necessary to accurately estimate a dual cost function parametrically requires more than 20 years of data based on observed data. Thus, some estimates of scope and scale may be fragile due to the inability to trace out that underlining production process.

Economies of scale estimations from nonparametric methods have been limited to measuring scale efficiency by estimating the model assuming constant returns to scale, and comparing with it with variable returns to scale (Cooper et. al. 2007). The results of the estimates are compared using the ratio of the two cost estimations yielding a measure known as scale efficiency. Paul et. al. (2004) noted however that scale efficiency is not the same as the multi-product scale economies explained by Baumol et. al. and cannot be interpreted as such. Also, there are no measures for product-specific economies of scale reported from nonparametric frontier estimations. Thus, those using nonparametric methods to estimate cost frontiers have been forced to parametrically estimate traditional scale measures (Paul et. al. 2004, Kumar, Sunil, and Gulati 2008).

Recently Parman et al. have shown that the DEA approach can estimate cost efficiency, economies of scale, and economies of scope relatively close to the “true” values of these economic measures relative to other parametric methods. Their approach reduces the need to conduct estimations using multiple methods, and provides scale measures consistent with Baumol et. al.

The primary objective of this research is to estimate economies of scale using the nonparametric approach through estimations of multi-product and product-specific scale economies and cost efficiency for Kansas farms. Product-specific scale economies are evaluated to determine if farm size is related to cost savings for individual products. From the cost frontier, it is possible to determine what type and size of farms make up the frontier and how far other farms are from the most efficient producers (cost efficiency). Using the nonparametric methods of Parman et al. for estimating scale allows the trade-off between cost efficiency and multi-product economies of scale to be examined to determine those farms that will reduce costs more by increasing output versus becoming cost efficient.

The second objective of this study evaluates estimating a panel of Kansas farm’s economic measures by year. Specifically, this objective addresses if scale and scope remain consistent across years as the cost frontier shifts due to technology improvement and/or weather variability. This has important implications for understanding how the cost function behaves over time.

Methods

Following Parman et. al., to estimate the frontier, economies of scope, and scale economies, the minimum cost (C_i) of producing the farm output mix is determined using DEA.

Costs are minimized for a given set of input prices (w_i) and outputs (y_i) with the choice being the optimal input bundle (x_i^*).

$$\begin{aligned}
&\min C_i = w_i' x_i^* \\
&s.t \\
&Xz \le x_i^* \\
&y'z \ge y_i \\
&z_1 + z_2 + \ldots + z_n = 1 \\
&z_i \in \mathbb{R}^+
\end{aligned} \tag{4.1}$$

where there are "n" farms. The vector Z represents the weight of a particular farm with the sum of Z_i*'s* equal to 1 for variable returns to scale. 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 ith farm. The cost of producing all outputs except one ($C_{i,all-p}$) where p is the dropped output is determined by dropping the p^{th} output constraint. The marginal costs ($MC_{i,p}$) are obtained from the shadow prices on the output constraint (equation 4.1). Using the cost and output measures obtained from the previous program, economies of scope, multi-product economies of scale, cost efficiency and product-specific economies can be calculated.

Cost efficiency (*CE*) identifies a farm's proximity to the cost frontier for a given input/output bundle. It is the quotient of the estimated frontier cost (equation 4.1) and the actual total cost (*ATC*) the farm incurred while producing their output bundle. This measure must be greater than 0 but less than or equal to 1.

$$CE_i = \left[\frac{C_i}{ATC_i} \right] \tag{4.2}$$

The calculation of multi-product economies of scale (MPSE) uses the total cost of producing all outputs ($C_{i,all}$), the marginal costs defined above, and the output levels produced. The MPSE is the change in total cost for a proportional change in the production of all outputs. For each output constraint (equation 4.1), the $MC_{i,p}$ is determined by the shadow price on the p^th^ constraint.

$$MPSE_i = \left[\frac{C_{i,all}}{\sum_p MC_{i,p} Y_{i,p}} \right] \tag{4.3}$$

Product specific economies of scale (PSE) require the calculation of the incremental costs ($IC_{i,p}$) that are the cost of producing all outputs minus the sum of the costs of all individual outputs except output (p).

$$IC_{i,p} = C_i - \sum_j C_{i,j \neq p} \forall j \tag{4.4}$$

Average incremental costs ($AIC_{i,p}$) are determined by dividing incremental costs by individual output:

$$AIC_{i,p} = \frac{IC_{i,p}}{y_{i,p}} \tag{4.5}$$

Using the average incremental cost and the marginal cost calculation above, PSEs are calculated by:

$$PSE_{i,p} = \frac{AIC_{i,p}}{MC_{i,p}} \tag{4.6}$$

The calculation of scope economies (SC_i) identifies the potential for cost savings through product diversification.

$$SC_i = \left[\frac{(\sum_p C_{i,p}) - C_{i,all}}{C_i} \right] \tag{4.7}$$

where $C_{i,p}$ is the cost of producing output p for farm *i*, and $C_{i,all}$ is the cost of joint production of all outputs for farm *i*.

Estimating the frontier nonparametrically using a data set with no single output farms reveals difficulty estimating the incremental costs by forcing one of the output constraints to zero (Equation 4.1). Thus, the only alternative is to drop one of the constraints. However, when an output constraint is dropped, the program may allow some of the output for the dropped constraint to be produced resulting in an overstatement in the cost of that one output ($C_{i,p}$) that will cause an over statement in economies of scope (equation 4.7) and an understatement in product specific scale economies (equation 4.6).

The additional product-specific production costs from an output being produced when it should be zero must be removed. The cost of producing y_1 only ($C_{i,1}$) assumes that only (y_1^1) is being produced. However, DEA allows some $y_{i,2}^1$ to be produced in this situation overstating the cost of producing y_1 only ($C_{i,1}$). To remove the additional cost, the percentage of $y_{i,1}^1$ is multiplied by the cost of producing y_1 only, yielding an adjusted cost ($C^a_{i,1}$). This new adjusted cost is then used in the calculation of incremental costs and associated economic measures (equation 4.8).

$$C^a_{i,1} = C_{i,1} \left(\frac{y^1_{i,1}}{y^1_{i,1} + y^2_{i,2}} \right) \tag{4.8}$$

Data

The data for this study contains 241 Kansas Farm Management Association farms (KFMA) for the years 2002-2011. Input quantities are aggregated into categories including seed, fertilizer, chemicals, feed, fuel, labor, land, and machinery. Associated prices for each input are indexed by year using the NASS[8] website or information from *Agricultural Outlook*. The land price is the Kansas cash rental rate from *Kansas Farm Facts*.

Outputs are aggregated into two categories including crops and livestock using output prices from NASS. Accrual revenue is divided by corresponding prices to obtain output quantities. Table 4.1 reports descriptive statistics for production quantity indices while Table 4.2 shows the price indices for both inputs and outputs for all ten years. The DEA model is estimated for the 241 farms for each year individually.

Estimating the cost frontier each year may cause some farms that operate on or close to the frontier in some years to be off the frontier in others due to the randomness of weather, rate of technology adoption, or other unforeseen phenomenon as the frontier shifts from year to year. This is important if the model includes data from an area where a drought occurs in isolated regions, and does not affect all farms universally.

Using the traditional USDA sales classes, of the 2,410 total observations, 92 fell into the category of gross revenues less than $100k, approximately 4% of the total while, the $100k-$250k categories includes 481 observations or nearly 20% of farms. The largest category is the $250k-$500k in annual gross revenues group that accounts for nearly 35% of farms or 837 observations. The $500k-$1m category is the next largest with 705 observations or 29% .Farms with gross revenues greater than 1 million had 295 observations or 12% of the total.

[8] http://www.nass.usda.gov/Statistics_by_Subject/index.php

Results

Two types of analysis were completed; One estimating all years together, 2,410 individual observations, and yielded 2,363 marginal cost estimations for crops. Therefore crop-specific scale economy calculations are reported for 2,363 observations. Though all observations produced crops, crop marginal cost estimates that were non-unique for farms on the frontier were also dropped[9]. Livestock-specific scale economies are reported for 1,749 observations which is significantly less than crops because many of the observations do not produce livestock. The calculations for multi-product scale economies include 1,671 observations, the number of observations that yielded marginal cost estimates for both crops and livestock. Economies of scope were calculated for 1,694 total observations. Cost efficiency is calculated for all 2,410 observations.

From the analysis that estimated each year individually, there were 2,271 observations yielding unique marginal costs for crops and 1,714 for livestock with 1,630 observations having marginal cost estimations for both. Thus, there are 1,630 estimates of each individual year's multi-product scale economies. Economies of scope were calculated from 1,684 total observations. The disparity in the number of observations of each economic measure between the single frontier and annual estimations arises because there are ten frontiers in the 2nd analysis and one in the 1st analysis affecting the number of non-unique marginal cost and incremental cost estimations.

After dropping the observations with non-unique marginal costs or zero output observations for livestock, the number of observations in the calculation of multi-product scale

[9] Also, some farms had a marginal cost calculation equal to zero for crops and livestock if they are small and highly inefficient not fully utilizing current resource allocation

economies (and economies of scope approximately) for the combined years estimation are: 31 with gross revenues less than $100k (2%), 380 with gross revenues between $100k and $250k (24%), 543 had gross revenues between $250k and $500k (36%), 491 with gross revenues between $500k and $1m (31%), and 226 with gross revenues above $1m (14%). Observations estimating each year individually are as follows: 28 with gross revenues less than $100k (1.7%), 298 with gross revenues between $100k and $250k (18%), 613 with gross revenues between $250k and $500k (38%), 489 with gross revenues between $500k and $1m (30%), and 226 with gross revenues above $1m (%12). Table 4.3 presents the summary statistics for the economic cost measures for both the annual and combined estimates. Table 4.4 shows the summary statistics for each year for the annual analysis.

F-tests were conducted for each economic measure to determine if the economic measures estimated annually were statistically different from the measures estimated with a single frontier. This was done by creating dummy variables for each year and regressing them on each economic measure. For all the economic measures including MPSE, cost efficiency, economies of scope, and the PSEs, the tests revealed that at a significance level of 5% these measures were statistically different (Table 4.5).

Cost Efficiency

The cost efficiency calculation for each farm represents its current distance from the frontier. A cost efficiency of 1 is on the frontier while those further from 1 are less cost efficient. From the single frontier analysis, average cost efficiency levels were highest for farms greater than $1m (0.55) and for farms less than $100k (0.55). Farms with gross revenues between $500k and $1m had an average cost efficiency of 0.48 while the categories $100k to $250k and $250k to $500k had averages of 0.43 and 0.42 respectively (Table 4.6). The standard deviation is relatively high for farms less than \$100k (0.20) compared to the other for revenue categories which had standard deviations of less than 0.15 (Table 4.6).

Estimation of each year yielded a higher overall average cost efficiency (Table 4.3) and higher average cost efficiencies for each gross revenue category (Table 4.6). This implies that farms are closer to each year's frontier on average than an overall frontier which is to be expected if the frontier is shifting. Each gross revenue category however retained its respective rank for overall average cost efficiency, i.e. farms with greater than \$1m in gross revenues had the highest average cost efficiency while farms in the \$100k to \$250k range had the lowest (Table 4.6). Examination of the annual cost efficiency averages (Table 4.4) reveals that average cost efficiencies have been lower in recent years than between the years 2003 to 2008.

Figure 4.1 shows the cumulative density or the amount of observations below a given cost efficiency level for the size categories. The slope of each curve indicates the variation observed for each group where a steeper slope represents less variability. Figure 4.1 shows an obvious flatter cumulative density for farms with gross revenues less than \$100k indicating a large disparity for cost efficiency levels in this revenue group which is true for both the annual estimations and the single frontier. However, in the single frontier estimation, the cumulative density for cost efficiency of farms less than \$100k in gross revenues crosses the curve for farms greater than \$1m in gross revenues at a cumulative density of 0.7. For the annual estimations this does not occur indicating that the largest farms are strictly closer to the frontier than any smaller revenue category. This implies that in the year with the lowest total cost, there were relatively many small farms close to the frontier however, in each year on average, the largest farms are closer to the frontier. The results for cost efficiency remain similar for the annual estimation and the single frontier for the rest of the revenue categories.

Multi-product Economies of Scale

Multi-product scale economies represent potential cost saving by reducing average per unit cost through spreading it over larger quantities. Because MPSE is calculated as total cost divided by the sum of the products of marginal costs and their associated output levels, an MPSE greater than 1 implies that increasing production uniformly across outputs will reduce average costs resulting in economies of scale. For MPSEs to be greater than one, the existence of economies of scope, and/or product-specific economies of scale (Fernandez-Cornejo et al 1992) are required (Baumol et al.). If the MPSE equals 1, then the farm is at constant returns to scale. However, if the MPSE is less than 1 for a given farm, then that farm can reduce average cost by proportionately reducing outputs since that farm lies in the diseconomies of scale region.

Single frontier estimation revealed that MPSE for each gross revenue category is highest for the smallest farm revenue category and gets progressively smaller for larger farms (Table 4.7). MPSE averages ranged from 2.7 (farms less than $100k) to 0.9 (revenues $500k-$1m and farms greater than $1m). Farms with sales of $100k to $250k had an average MPSE at 1.7 and farms between $250k and $500k were closer to unity at 1.1.

The overall average estimated annually was similar to the single frontier at 1.171 compared to 1.142 respectively (Table 4.3). The MPSEs are also smaller for each gross revenue category overall estimated yearly relative to the single frontier estimates while retaining the same relative rank of each category (Table 4.7). Yearly average MPSE estimates show farms, on average, remaining close to constant returns to scale each year (Table 4.4).

Figure 4.2 Panels A and B present the distribution for the single frontier analysis and the multiple frontier analysis respectively. The results are similar except that MPSE is lower when estimated annually as reflected by MPSE density curves closer to one. Standard deviation is relatively low farms in the \$500k to \$1m and farms greater than \$1m as illustrated by the nearly vertical cumulative density curves. Within these two groups, MPSE is relatively constant among large farms. For the smallest two categories, the density curves are relatively flatter, especially for the smallest farm category showing a disparity.

Economies of Scope

Economies of scope represent cost savings through the production of crops and livestock. This savings may be due to the use of resources required for the production of both products such as equipment or storage resources. An economy of scope calculation greater than 0 implies cost savings are realized though multi-product operations. Results show greater difference for economies of scope than for cost efficiency between the farm revenue categories (Table 4.8). From the data estimated for the single frontier, the highest average level of cost savings from economies of scope is for farms between \$100k and \$250k with average economies of scope of 30% (Table 4.8). Large farms including farms with revenues over \$1m and farms between \$500k and \$1m had relatively low economies of scope figures of 13% and 12% respectively. Economies of scope for the smallest category were also high (26%). Using annual frontiers, the measurement of economies of scope is less than those estimated from a single frontier. Annual averages for the scope measures range from 0.06 in 2004 to 0.17 in 2002 (Table 4.4).

Standard deviations for the economies of scope calculations were below 0.10 for the single frontier but higher for the two smallest gross revenue categories for the annual estimations (Table 4.8). Figure 4.3 Panel A shows that the cumulative density for farms in the \$100k to \$250k category is relatively flatter indicating more overall disparity among economies of scope calculations for this revenue group. The largest gross farm revenue category (greater than $1m) had a relatively low average and relatively high standard deviation at 0.08.

One key difference between the two estimations of scope (annual frontiers versus a single frontier) was that annual estimations yielded negative economies of scope for some of the observations in the larger gross revenue categories. While nearly all of the observations of cost savings from scope estimated annually were lower than the simultaneously estimated data set, none of the simultaneous estimates yielded negative scope economies (diseconomies of scope).

Product-specific Economies of Scale

If a product-specific economies of scale (PSE) measure is greater than 1, it implies that there exists potential cost savings from increasing that output, and a PSE less than 1 implies cost savings by reducing that output. The overall average product-specific economies of scale measure for livestock (LSE) is higher than the product-specific economies of scale measure for crops (CSE) at 0.83 and 0.77 respectively (Table 4.3) using the single frontier. However the reverse is true for the PSE estimations from annual analysis, though the difference is relatively small (0.01). All farms operate either at constant returns to scale for CSE and LSE or in the region of diseconomies of scale for crops and livestock.

For CSE under a single frontier, the smallest farm revenue group (less than $100k) was the closest to constant returns to scale on average at 0.85 where the furthest group was the $500k to $1m with an average CSE of 0.74 (Table 4.9). There was not much difference in average CSE within the four groups with gross revenues greater than $100k. From annual frontiers, the revenue group of less than $100k was also highest but closer to constant returns to scale than in the previous estimation at 0.97. The greater than $1m sales group had a PSE of 0.83. Yearly overall averages for CSE (Table 4.4) are between 0.8 and 0.9 during the ten year sample showing relatively little variation in overall crop-specific economies of scale from year to year.

The cumulative density curves (Figure 4 Panel A) for CSE for the four largest gross revenue categories estimated simultaneously overlap with small differences in slope indicating the relative variation within groups is also small. However, the CSE density curve for the group containing farms with revenues less than $100k is relatively flat for 50% of the farms and steep for the other 50%. This indicates that many farms in the less than $100k category are operating at a low CSE while others are at or close to constant returns to scale for crops with a single frontier. In Figure 4 Panel B the CSE for the smallest gross revenue categories is not as flat illustrating a tighter distribution with an annual frontier.

Single frontier estimates reveal that the averages between groups for LSE were highest among smaller revenue grossing farms with the three smallest categories all having an average LSE higher than 0.84 (Table 4.10). The two largest revenue grossing categories had nearly identical LSE averages at approximately 0.80. Annual frontier analysis shows that the smallest revenue group (less than $100k) is close to constant returns to scale on average with the other revenue groups averaging between 0.84 and 0.87. Annual averages (Table 4.4) for LSE yield results similar to CSE in that the lowest LSE estimate occurs in 2002 and the rest are approximately between 0.8 and 0.9 indicating relative stability for livestock-specific scale economy estimates from year to year.

For the single frontier data set, the standard deviations for LSE were higher than for CSE indicating more variability in the product-specific scale economies for livestock than crops (Table 4.10). The highest standard deviation was for the gross revenue category less than $100k

(0.23) and lowest was for the greater than \$1m category (0.18). However, the density curves for all five categories are similar, without the obvious differences between groups that the other economic measures show (Figure 4.5 Panel A). The annual frontier analysis shows similar results for standard deviations among revenue categories with the exception of farms with revenues less than \$100k.

Implications

Differences between Annual Frontier and Single Frontier Analysis

The statistical test used to determine if the means from the model estimating single frontier was different from those estimating the frontier annually indicated statistical differences at the 5% level. However, the results show that the means are not that economically different. Overall average MPSE was around constant returns and did not vary much from year to year (Table 4.5). Crop-specific and livestock-specific scale differences from both estimations were similar in mean and relative rank with relatively little variation from year to year.

The largest difference between economic measure estimates occurred with respect to cost efficiency and economies of scope. In the case of cost efficiency, the difference in overall average from estimating a single frontier versus annual frontiers occurs due to the cost frontier shifting from year to year. Estimating a single frontier assumes the frontier does not shift and thus movement of farms closer to, and further from the frontier is due to efficiency. Estimating the frontier annually allows the frontier to shift and average cost efficiency to remain constant assuming farms are not changing their relative efficiency.

Allowing the frontier to shift from year to year will also affect calculations of economies of scope. Economies of scope are based on estimations of the intercept and when estimated for a single frontier will not change. Thus, increase in the cost of producing each output individually will appear from a single frontier as higher cost savings from joint production rather than changes in the intercepts as the frontier shifts.

It appears that annual frontiers suggest that CSEs and LSEs are closer to one and the scope is closer to zero than the single frontier results. While there is some variation in the annual economic measures, they are relatively stable from year to year.

Implications for KFMA Farms

Despite the differences between the estimation of cost efficiency and scope between the annual and single frontier estimations, the implications are the same in that larger farms, and the smallest category, are typically closer to the frontier and economies of scope diminish as farms grow larger. Further, economies of scale exist for small farms and tend to be exhausted for farms with sales greater than $250k.

For the smallest farm category (less than $100k), the estimates for cost efficiency and MPSE suggest that these farms have a greater incentive to increase in size rather than move closer to the frontier. Estimated annually, the cost efficiency for this group is 0.66 and the MPSE is 1.99. This shows costs can be reduced by on-average 50% by increasing in size and 34% by becoming more efficient. Economies of size are clearly important for these farms.

For the $100k to $250k group, the implications are also similar in that the benefits are nearly equal in becoming more efficient versus increasing output. From the annual frontier estimates, the overall average cost efficiency is 0.56 indicating that they can save 44% becoming more efficient. Potential cost savings from scale are around 41% indicating a closeness between the two.

For largest three gross revenue categories, the results show cost savings from reaching the frontier is more important than adjusting farm size. All three categories are near constant returns to scale, or slightly in the diseconomies of scale region. The average cost efficiencies range from 0.57 to 0.75 indicating that there is room for cost savings by becoming more efficient.

Economies of scope are more important. Multi-product smaller farms realize greater cost savings through joint production than larger farms. At some point, the advantage of joint production is exhausted. Farms with less than $250k in gross revenues tend to experience greater cost savings with joint livestock and crop production. As farm sales increase however, the incentive to grow larger due to additional cost savings from scale, and savings from joint production diminish.

Product-specific scale economies from annual frontiers are between 0.75 and 0.95 for crop-specific economies of scale and livestock-specific economies of scale. These measures do not vary as much based on farm size as the other measures. Perhaps the conclusion is that the individual enterprises are more size neutral. When arranged in a multi-product farm, multi-product scale measures differ due to level of scope economies. Multi-product farms reap the benefits of joint production (scope) and are not as far from constant returns to scale for livestock or crops specifically. However, the large potential for cost savings illustrated by small farms typically having high MPSE suggests the importance of economies of scope for these operations.

Conclusions

The objectives of this research were to determine the level of cost savings from cost efficiency, economies of scale and economies of scope based on farm size for Kansas farms.

This research also evaluated the difference between estimating the frontier yearly versus a single frontier.

The results suggest that there exists a larger incentive for small farms to expand and exploit cost savings through multi-product scale economies. Scale economies are larger than potential saving from becoming more efficient farms with sales less than \$100k. As farms move past the \$100k in sales range, the potential cost savings from efficiency is about the same as from adjusting size. After sales reach \$250k, most of the economies of size are exhausted and cost differences occur due to inefficiency (not being on or close to the frontier).

Estimating the measures as a single multi-year frontier yielded results that were statistically different than from estimating annual frontiers. The measures of economies of scope were lower when estimated annually and the PSEs are higher. Interestingly, while those measures were statistically different, the variability in measures from year to year was not large and the measures of multi-product scale economies were nearly the same. For example, multi-product scale economies for Kansas Farms were between 0.97 and 1.17 for the ten year period and cost efficiency measures were between 0.55 and 0.67 for the 2002 to 2011 time period. This indicates that while the cost frontier may shift from year to year, its shape remains relatively consistent.

Chapter 4 References

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

Atkinson, Scott E. and Robert Halvorsen. "Parametric Efficiency Tests, Economies of Scale, and Input Demand in U.S. Electric Power Generation" *International Economic Review.* 25(4)(October 1984):647-662.

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

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

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

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

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

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

Fernandez-Cornejo, J., C.M. Gempesaw II, J.G. Elterich, and S.E. Stefanou. "Dynamic Measures of Scope and Scale Economies: An Application to German Agriculture" *American Journal of Agricultural Economics*, 74(1992): 329-342.

Kumar, Sunil, and Richita Gulati. "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)(2008):33-69.

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

Mafoua, Edouard K. "Scope and Scale Economies for Multi-Product Farms: Firm-level Panel Data Analysis" Selected Paper AAEA August 2001.

Parman, Bryon J., Allen M. Featherstone, Brian K. Coffey. "A Nonparametric Approach to Multi-product and Product-specific Economies of Scale" 2013. Forthcoming

Parman, Bryon J., Allen M. Featherstone, Vincent Amanor-Boadu. "A Comparison of Parametric and Nonparametric Estimation Methods of Production Frontiers and Multi-product Scale, Product-Specific Scale, and Scope Economies." 2013. Forthcoming

Paul, Catherine Morrison, Richard Nehring, David Banker, and Agapi Somwaru. "Scale Economies and Efficiency in U.S. Agriculture: Are Traditional Farms History?" *Journal of Productivity Analysis*, 22(2004): 185-2005.

Samuelson, P., 1938. Foundations of Economic Analysis, Harvard University Press, Cambridge

Shephard, R.W. 1953. Cost and Production Functions, Princeton University Press, Volume 1, Issue 2, Princeton NJ.

Chapter 4 Tables

Table 4.1 Summary statistics for Kansas Farm Management Farms of input and output quantity indices, 2002 to 2011.

	Mean	Standard Deviation	Minimum	Maximum
		----------*Inputs*----------		
Seed	188	199	1	2010
Fertilizer	298	286	0	3328
Chemicals	229	222	0	2457
Machinery	530	437	24	4163
Feed	401	1497	0	30454
Fuel	162	160	4	1906
Labor	192	306	0	3753
Land	2514	1628	127	11797
		----------*Outputs*----------		
Crops	2458	2182	30	50140
Livestock	1514	3386	0	26277

N=2410

Table 4.2 Price indices for farm inputs and outputs for each year 2002-2011.

Year / Product	Seed	Fertilizer	Chemicals	Machinery	Feed	Fuel	Labor	Rent	Crops	Livestock
2002	154	124	121	151	114	140	157	123	109	103
2003	158	140	121	162	121	165	160	126	120	116
2004	168	164	123	173	117	216	165	129	111	118
2005	182	176	128	182	124	239	171	141	134	116
2006	204	216	129	191	149	264	177	147	186	118
2007	259	392	139	209	194	344	183	165	259	117
2008	299	275	149	222	186	229	188	184	186	106
2009	310	252	144	230	180	284	189	190	177	123
2010	332	328	145	244	226	362	192	205	239	151
2011	359	333	153	257	260	360	199	212	246	160

Source: http://www.nass.usda.gov/Statistics_by_Subject/index.php

Table 4.3 Overall summary statistics for estimated cost measures for Kansas Farm Management Farms estimated from a single frontier and annually.

	N	Average	Standard Deviation	Minimum	Maximum
		----- Single Frontier------			
Cost Efficiency	2410	0.462	0.136	0.138	1.000
Multi-product Economies of Scale	1571	1.142	0.407	0.588	4.210
Economies of Scope	1571	0.175	0.093	0.003	0.553
Crop-specific Economies of Scale	2363	0.768	0.167	0.023	1.000
Livestock-specific Economies of Scale	1649	0.830	0.190	0.010	1.000
		------ Annual Frontiers------			
Cost Efficiency	2410	0.608	0.168	0.138	1.000
Multi-product Economies of Scale	1630	1.171	1.328	0.072	3.079
Economies of Scope	1630	0.110	0.101	-0.220	0.639
Crop-specific Economies of Scale	2271	0.862	0.182	0.105	1.000
Livestock-specific Economies of Scale	1714	0.854	0.183	0.016	1.000

Table 4.4 Annual averages for cost efficiency, MPSE, PSEs, and economies of scope for Kansas Farm Management Farms

Year	Cost Efficiency	Multi-product scale economies	Economies of scope	PSE Crops	PSE Livestock
2002	0.546	1.061	0.170	0.752	0.796
2003	0.639	1.066	0.085	0.940	0.906
2004	0.635	0.999	0.063	0.937	0.906
2005	0.668	0.992	0.074	0.914	0.869
2006	0.610	1.068	0.124	0.848	0.852
2007	0.606	1.060	0.112	0.916	0.810
2008	0.653	1.022	0.096	0.926	0.803
2009	0.596	0.967	0.098	0.832	0.810
2010	0.546	1.155	0.157	0.795	0.892
2011	0.586	1.170	0.108	0.866	0.898

Table 4.5 F-Test results evaluating statistical differences in cost frontiers.

Measure	F-Statistic	P-Value
Cost Efficiency	10.28	0.000
Multi Product Economies of Scale	1.98	0.046
Economies of Scope	36.20	0.000
Crop-specific Economies of Scale	11.97	0.000
Livestock-specific Economies of Scale	7.43	0.000

Table 4.6 Summary statistics for cost efficiency for Kansas Farm Management Farms estimated from a single frontier and annually.

Gross Revenues	N	Average	Standard Deviation	Minimum	Maximum
			----- *Single Frontier*------		
Less than $100k	92	0.549	0.204	0.198	1.000
$100k-$250k	481	0.432	0.125	0.151	1.000
$250k-$500k	837	0.421	0.110	0.138	1.000
$500k-$1m	705	0.483	0.130	0.218	1.000
Greater than $1m	295	0.552	0.146	0.280	1.000
			------ *Annual Frontiers*------		
Less than $100k	92	0.660	0.206	0.261	1.000
$100k-$250k	481	0.559	0.155	0.225	1.000
$250k-$500k	837	0.567	0.145	0.198	1.000
$500k-$1m	705	0.627	0.155	0.219	1.000
Greater than $1m	295	0.749	0.179	0.357	1.000

Table 4.7 Summary statistics for multi-product economies of scale for Kansas Farm Management Farms estimated from a single frontier and annually.

Gross Revenues	N	Average	Standard Deviation	Minimum	Maximum
		------ *Single Frontier*------			
Less than $100k	31	2.691	0.542	1.938	3.732
$100k-$250k	380	1.619	0.380	0.965	4.210
$250k-$500k	543	1.106	0.224	0.711	1.708
$500k-$1m	491	0.916	0.121	0.658	1.359
Greater than $1m	226	0.918	0.070	0.588	1.010
		------*Annual Frontiers*------			
Less than $100k	28	1.991	0.586	1.266	3.079
$100k-$250k	298	1.406	0.892	0.729	3.053
$250k-$500k	613	1.048	0.188	0.576	1.670
$500k-$1m	489	0.941	0.140	0.575	1.250
Greater than $1m	202	0.850	0.128	0.072	1.075

Table 4.8 Summary statistics for economies of scope from Kansas Farm Management Farms estimated from a single frontier and annually.

Gross Revenues	N	Average	Standard Deviation	Minimum	Maximum
		----- *Single Frontier*------			
Less than $100k	31	0.255	0.075	0.063	0.443
$100k-$250k	380	0.301	0.091	0.108	0.529
$250k-$500k	543	0.169	0.060	0.057	0.553
$500k-$1m	491	0.123	0.050	0.020	0.307
Greater than $1m	226	0.134	0.083	0.003	0.332
		------*Annual Frontiers*------			
Less than $100k	28	0.201	0.161	0.000	0.481
$100k-$250k	298	0.196	0.133	-0.011	0.639
$250k-$500k	613	0.116	0.072	-0.128	0.323
$500k-$1m	489	0.075	0.059	-0.122	0.218
Greater than $1m	202	0.037	0.092	-0.220	0.558

Table 4.9. Summary statistics for crop-specific economies of scale categorized by gross revenues estimated simultaneously and individually by year

Gross Revenues	N	Average	Standard Deviation	Minimum	Maximum
			----- *Single Frontier*------		
Less than $100k	77	0.854	0.233	0.076	1.000
$100k-$250k	476	0.780	0.155	0.023	1.000
$250k-$500k	834	0.775	0.175	0.028	1.000
$500k-$1m	703	0.743	0.162	0.120	1.000
Greater than $1m	273	0.764	0.141	0.282	1.000
			------*Annual Frontiers*------		
Less than $100k	53	0.974	0.095	0.387	1.000
$100k-$250k	465	0.873	0.191	0.192	1.000
$250k-$500k	821	0.902	0.145	0.149	1.000
$500k-$1m	676	0.847	0.145	0.282	1.000
Greater than $1m	256	.0826	0.147	0.105	1.000

Table 4.10 Summary statistics for livestock-specific economies of scale categorized by gross revenues estimated simultaneously and individually by year

Gross Revenues	N	Average	Standard Deviation	Minimum	Maximum
			----- *Single Frontier*------		
Less than $100k	31	0.850	0.232	0.082	1.000
$100k-$250k	380	0.877	0.192	0.046	1.000
$250k-$500k	543	0.842	0.182	0.010	1.000
$500k-$1m	491	0.799	0.190	0.031	1.000
Greater than $1m	226	0.807	0.179	0.029	1.000
			------*Annual Frontiers*------		
Less than $100k	45	0.969	0.140	0.094	1.000
$100k-$250k	310	0.846	0.230	0.020	1.000
$250k-$500k	625	0.836	0.181	0.016	1.000
$500k-$1m	512	0.862	0.167	0.048	1.000
Greater than $1m	222	0.873	0.142	0.095	1.000

Chapter 4 Figures

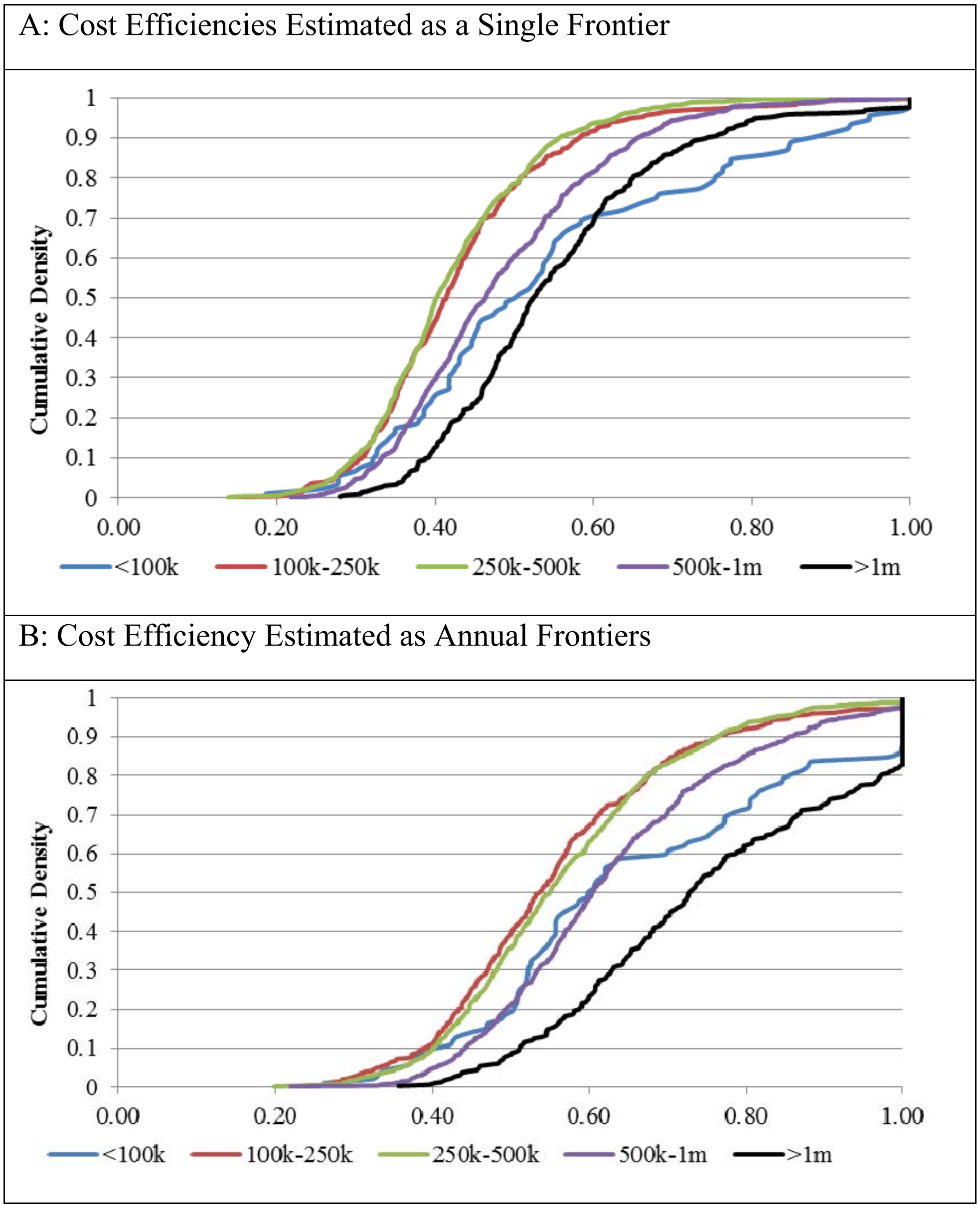


Figure 4.1 Cumulative Density of Cost Efficiency Estimates for Kansas Farms Categorized by Farm Gross Revenue

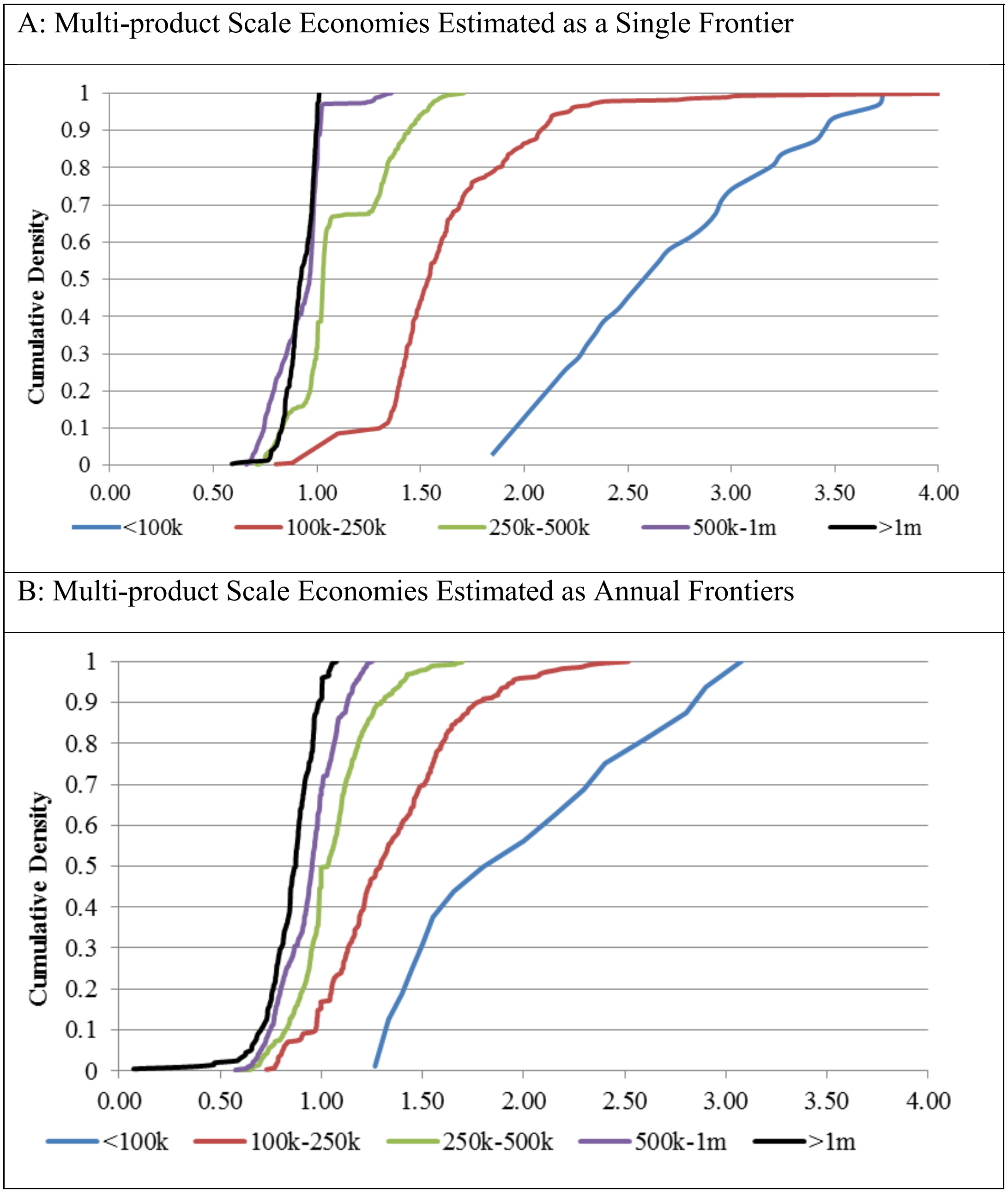


Figure 4.2 Cumulative Density of Multi-product Scale Economies Estimates for Kansas Farms Categorized by Farm Gross Revenue

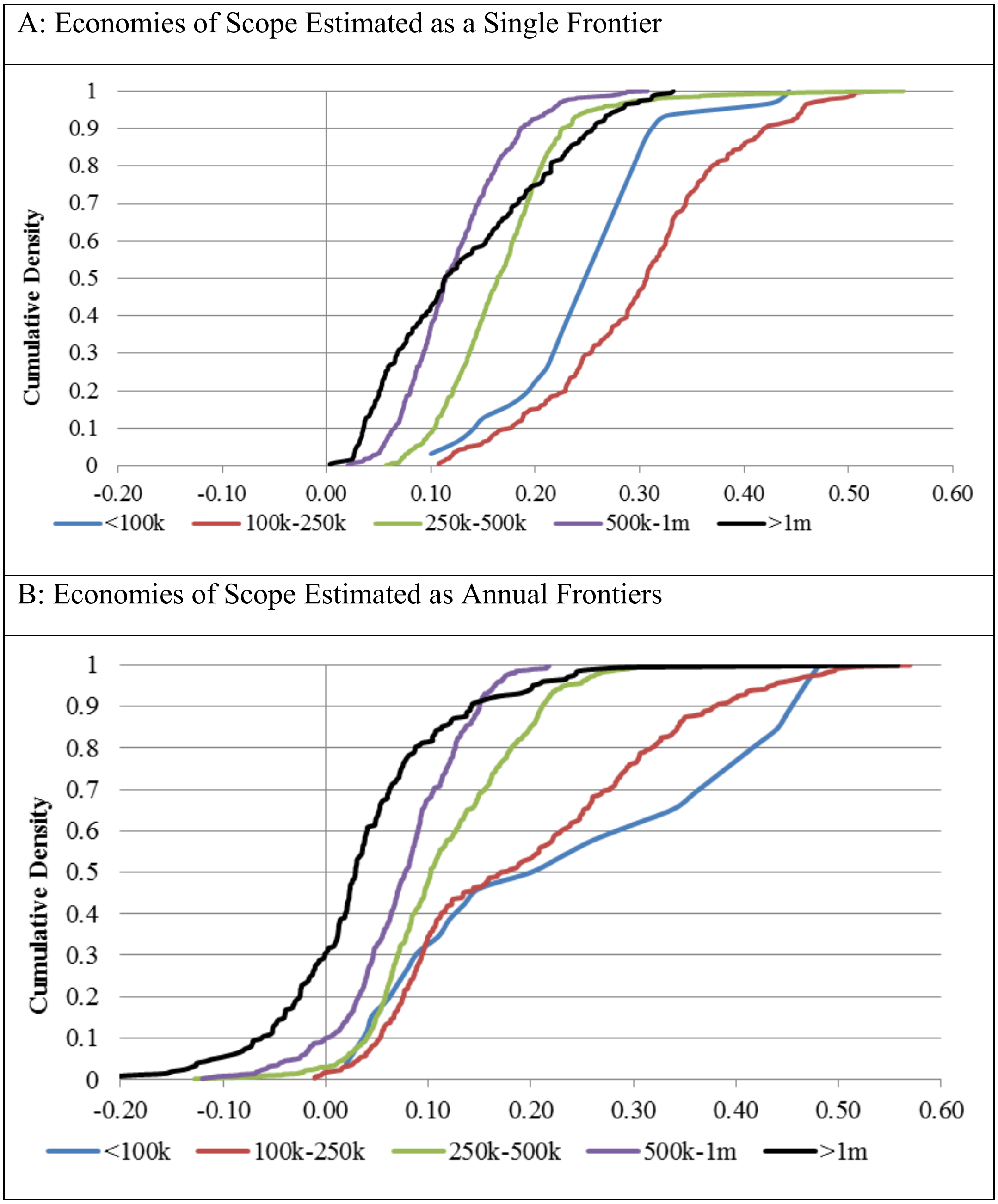
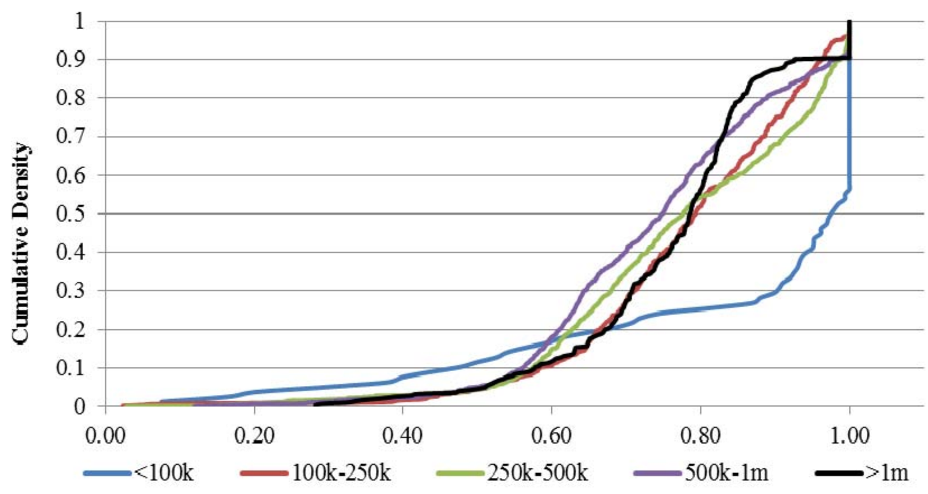


Figure 4.3 Cumulative Density of Economies of Scope Estimates for Kansas Farms Categorized by Farm Gross Revenue

A: Crop-Specific Economies of Scale Estimated as a Single Frontier



B: Crop-specific Economies of Scale Estimated as Annual Frontiers

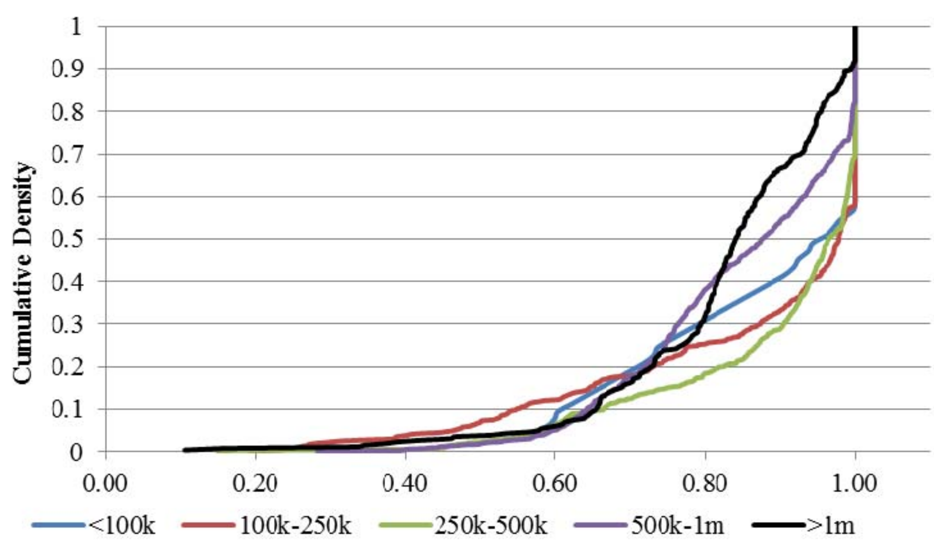
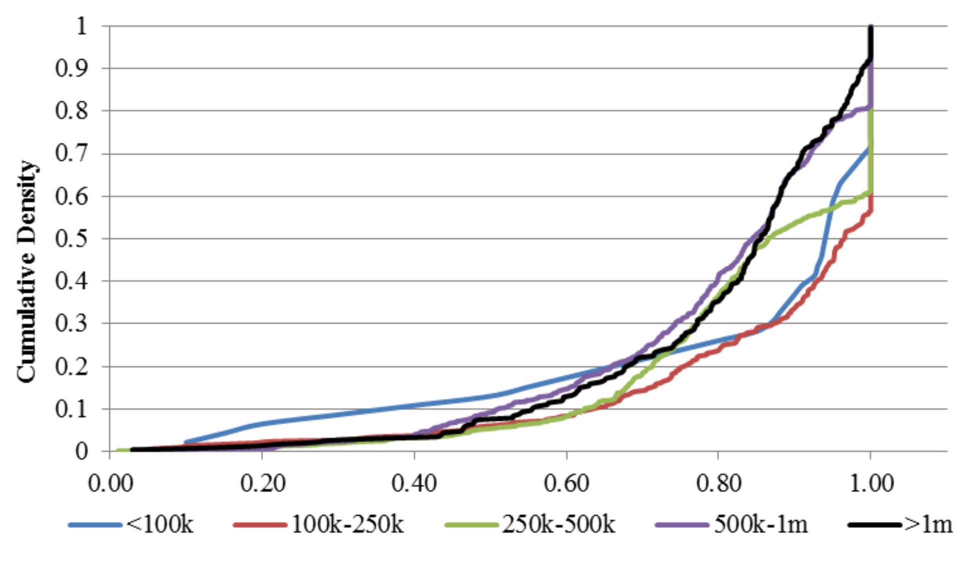


Figure 4.4 Cumulative Density of Crop-specific Scale Economy Estimates for Kansas Farms Categorized by Farm Gross Revenue

A: Livestock-Specific Economies of Scale Estimated as a Single Frontier



B: Livestock-specific Economies of Scale Estimated as Annual Frontiers

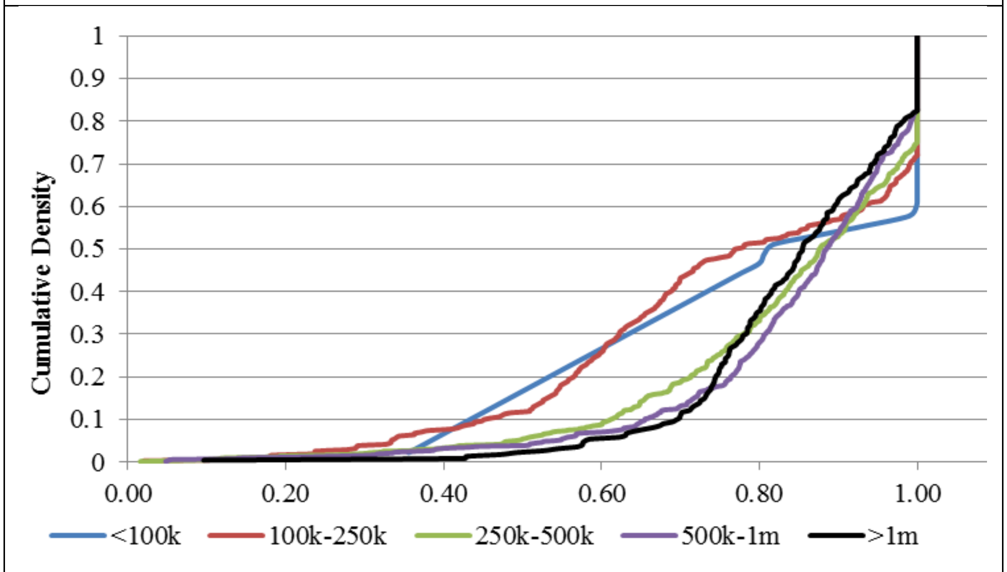


Figure 4.5 Cumulative Density of Livestock-specific Scale Economy Estimates for Kansas Farms Categorized by Farm Gross Revenue

Chapter 5 - Conclusions

This research presents a method for measuring multi-product and product-specific economies using data envelopment analysis, a comparison of this method with alternative methods, and an empirical example of its use. The objective of this research is to provide a theoretically consistent method of frontier estimation that is able to recover economic information, while eliminating the need to use multiple methods. The additional nonparametric measures were developed using the DEA cost frontier method proposed by Färe et. al. where the marginal costs and incremental costs can be estimated, and used, to calculate multi-product and product-specific scale economy measures consistent methods proposed by Baumol et. al. The tests conducted, and overall results, indicate that the nonparametric cost frontier can be used to estimate these economic cost measures. The empirical application of these methods on Kansas farms shows its efficacy in practice, and provides some useful information for economists, producers, and policy makers with respect to cost savings potential for Kansas farms.

In Chapter 2, an approach was formalized to calculate multi-product and product-specific economies of scale from DEA. It was then compared to data from a "true" frontier cost function using two different error (inefficiency) distributions, and two "true" cost functions. Chapter 2 also compared calculating scope economies and incremental costs by dropping an output constraint with constraining the appropriate output to equal zero.

When measuring observations with cost inefficiency, the nonparametric approach using either distributional assumption was able to estimate multi-product scale economies, product-specific scale economies, cost efficiencies, and economies of scope. The mean differences between the nonparametric estimates and the "true" frontier were close to zero with low standard deviations. While the PSE estimates are close to the PSEs of the "true" frontier function in the half-normal case, the deviations from the nonparametric approach for the PSE calculations using a uniform distribution illustrate the importance of having observations from efficient firms producing only a single output. In areas where there are few, or no, single output observations or where observations are not on the "true" frontier, the incremental costs for these observations may deviate from the "true" frontier function. Areas where the data are clustered yield more precise estimates than areas where observations are sparse.

Chapter 3 compares the nonparametric approach with three parametric cost frontier estimation techniques. The parametric methods include a two-sided error system, the stochastic frontier estimation, and an OLS estimation restricting the errors to be positive. Cost frontiers were estimated using the four estimation methods from a half-normal and uniform cost inefficiency distribution. Along with the two different distributions, a data set with no single output firms was used to evaluate each method's ability to extrapolate incremental cost measures out-of-sample. The estimates calculated included multi-product and product-specific scale economies, economies of scope, and cost efficiency. These measures were compared to the "true" values form the simulated cost frontiers. Cost efficiency is reported for only three of the methods except the two-sided error system because it does not estimate a frontier.

The two-sided error system was the furthest from the "true" values for multi-product scale economies in all three cases. For the product-specific scale economy estimates, the two-sided error system performed similar to the other parametric methods using the half-normal and half uniform data simulations but was closest to the "true" values for data with no single output observations. For economies of scope, the two-sided error system was the furthest from the "true" values except for the simulation with no single output firms.

The OLS positive errors model estimated cost efficiency closest to the true measures when compared to the other parametric methods for each of the three data simulations. For multi-product scale economies and product specific scale economies, the OLS positive errors results were mixed in that it was closer than the other methods under the uniform distributional assumption, but further under the half-normal distributional assumption and the single output firm case. Results were also mixed for economies of scope calculations from the OLS positive errors model where it was closer than other parametric methods under the uniform and half-normal distributions, but further than the two-sided error system for the case with no single output firms observed.

The stochastic frontier method estimated multi-product scale economies closer to the "true" measures compared to the two-sided error system but further than the OLS positive errors model. Cost efficiency for the stochastic frontier was similar to the OLS positive errors model for all three data sets, as were product-specific scale economies estimates under the half-normal and uniform distributional assumptions. However, economies of scope calculations were inaccurate for all three data sets. Product-specific scale economies using the data set with no observed single output firm observations were also inaccurate for this method.

Overall, the nonparametric approach estimated the frontiers and associated economic measures relatively close to the "true" values. The estimated economic measures were as close or closer to the "true" values than any of the methods examined. The MPSE and economies of scope measures were the most accurate of the scenarios examined. In the case of the PSE's, the nonparametric approach was not more inaccurate compared to the other methods.

Chapter 4 used the nonparametric approach developed in Chapter 2 to estimate a cost frontier for Kansas farms. The data contained farm level data from 241 farms for the years 2002-2011. The data were estimated in two fashions nonparametrically. First, a single frontier was analyzed for the 2,410 individual data points. For the second analysis, each year estimated its own cost frontier. There was a statistical difference in the means for the cost measures estimated annually compared to a single frontier. Thus, the frontier shifts from year to year. With respect to cost efficiency, differences in the means occurred due to the cost frontier shifting. While mean calculations of scale and scope were statistically significant annually, the differences were economically small so that economic interpretations were not affected. For economies of scope, the differences between the years are consistent from year to year. The means for the product-specific scale measures from both estimates were relatively close over time.

An important result of Chapter 4 is that it was possible to estimate cost efficiency, economies of scale, and economies of scope measures from a single year's data. In these estimations, relative prices for inputs and outputs did not change within each year such that there was no relative price variability among farms. Parametric methods for these calculations using the dual cost approach has been shown in previous literature to require 20 years' worth of data to yield enough price variability to estimate the same cost measures (Lusk et. al.).

The estimations suggest that there exists an economic incentive for small farms to expand up to about $250k in sales. Savings from exploiting economies of scale are greater for the smallest farms (less than $100k in gross revenues) than cost savings through efficiency improvements aimed at moving the farm closer to the frontier. Cost savings from scale versus cost efficiency are the same magnitude for farms with sales between $100k and $250k in sales. However, economies of scale are exhausted when gross revenues reach approximately $500k. At average gross revenues higher than $500k, many farms are operating in a region of slight diseconomies of scale and cost savings can no longer come from scaling.

Results were mixed with respect to product-specific scale economies for crops relative to livestock and overall farm size. Frontier estimates of crop-specific and livestock-specific scale economies are close to constant returns to scale. The crop-specific and livestock-specific estimates are between 0.75 and 0.95 indicating that on average farms within this data set are operating near constant returns to scale or slightly in the diseconomies of scale region of product-specific economies of scale.

This dissertation also highlights areas for future research. The results suggest a thorough examination of product-specific economies of scale is warranted to provide more accurate estimates. Using the data from chapter 4, it may also be wise to estimate a cost frontier for alternative regions in Kansas due to highly variable precipitation rates.

To summarize, this dissertation operationalized the calculation of multiple economies of scale measures and product-specific economies of scale measures for DEA methods. This has not previously been reported in the literature. It then tests the methods compared to previous methods and finds that the measures are no worse than parametric methods. Finally, the methods are applied to Kansas farm-level data. Because of the nature of DEA analysis, annual measures of scope and scale can be measured. Current parametric methods based on duality often are unable to estimate these annual measures due to an insufficient amount of relative price variability.

Chapter 5 References

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

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

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

Appendix A - Software Codes Used in Chapter 2

A.1 Data Generation Using Shazam Software Package

Below is the software code used to generate the data for Chapters 2 and 3. A.1. generates 400 observations which are half normally distributed and calculates the associate economic measures including cost efficiency, multi-product scale economies, product-specific scale economies, and economies of scope. To generate uniform data set, change line 25 to read "genr e=UNI(0,900)".

To make the full 500 observations for the half-normal distribution, 50 observations are created producing y_1 only and y_2 only. For the 50 observations producing y_1 only, in the half-normal case, change line 3 to read "gen1 nreps=1" and line 75 to read "genr y2=0*(c12*p1-c11*p2+aa2*w1+bb2*w2+cc2)/(-D)". For the 50 observations with y_2 only, change line 3 as shown above and change line 74 to read "genr y1=0*(c22*p1-c12*p2+aa1*w1+bb1*w2+cc1)/D". Then combine the 3 data sets.

For the full 500 observations with a uniform distribution, change line 25 for uniform distribution shown above and repeat steps for y_1 only and y_2 only.

```
1.  par 60000
2.  size 10000
3.  gen1 nreps=8
4.  50, 100, 250, 500
5.  gen1 nobs=50
6.  gen1 tot=nreps*nobs
7.  set maxcol=tot
8.  sample 1 tot
9.  *coef of variation on prices
10. .1, .2, .3, .4
11. gen1 coefp=.11
12. *coef of variation on inputs - i.e. measurement error
```

```
.001, .005, .01, .02
gen1 coefi= 0.02
.001, .005, .01, .02
gen1 coefmep=0.02
*Generate original data input price(w), output price(p),
*get output quantity y using w and p, then use w and y to calculate cost
set ranfix
genr w11=9+NOR(9*coefp)
genr w22=18+NOR(18*coefp)
genr w33=7+NOR(7*coefp)
genr p11=325+NOR(100*coefp)
genr p22=800+NOR(100*coefp)
genr e=NOR(0,1000)
stat w11 w22 w33 p11 p22/pcov
*normalize input and output prices using third input price w33
genr w1=w11/w33
genr w2=w22/w33
genr w3=w33/w33
genr p1=P11/w33
genr p2=p22/w33
*parameters in cost function with two inputs and two outputs
*second order derivative of input prices in cost function is b11 b12 b22
*second order derivative of output prices in cost function is c11 c12 c22
gen1 b0=20
gen1 b1=10
gen1 b2=35
gen1 a1=30
gen1 a2=80
gen1 b11o=.3
gen1 b12o=.5
gen1 b22o=.7
gen1 c11o=1.2
gen1 c12o=-0.2
gen1 c22o=1.5
gen1 a11=.5
gen1 a12=1
gen1 a21=.6
gen1 a22=.5
gen1 zo=0.0
*parameters related to input price w, pr is transform matrix of p, negative semi-definite
matrix
matrix p=((b11o|b12o)'|(zo|b22o)')
matrix pr=p'
matrix pp=-p*pr
matrix b11=pp(1,1)
matrix b12=pp(1,2)
```

```
58. matrix b22=pp(2,2)
59. *parameters related to output quantity y, positive semi-definite matrix
60. matrix h=((c11o|c12o)'|(zo|c22o)')
61. matrix hr=h'
62. matrix hh=h*hr
63. matrix c11=hh(1,1)
64. matrix c12=hh(1,2)
65. matrix c22=hh(2,2)
66. *generate output quantity given input price w and output price p
67. matrix D=(c22*c11-c12*c12)
68. gen1 aa1=a12*c12-a11*c22
69. gen1 bb1=a22*c12-a21*c22
70. gen1 cc1=a2*c12-a1*c22
71. gen1 aa2=a12*c11-a11*c12
72. gen1 bb2=a22*c11-a21*c12
73. gen1 cc2=a2*c11-a1*c12
74. genr y1=(c22*p1-c12*p2+aa1*w1+bb1*w2+cc1)/D
75. genr y2=(c12*p1-c11*p2+aa2*w1+bb2*w2+cc2)/(-D)
76. stat y1 y2
77. *generate cost using input price w and output quantities (quadratic function with two
    inputs and outputs)
78. genr
    cs=b0+b1*w1+b2*w2+a1*y1+a2*y2+0.5*b11*w1*w1+b12*w1*w2+0.5*b22*w2*w2+
    0.5*c11*y1*y1+c12*y1*y2+ &
79. 0.5*c22*y2*y2+a11*w1*y1+a12*w1*y2+a21*w2*y1+a22*w2*y2 + abs(e)

80. *costs without the error
81. genr TotalCost=
    b0+b1*w1+b2*w2+a1*y1+a2*y2+0.5*b11*w1*w1+b12*w1*w2+0.5*b22*w2*w2+0.5*
    c11*y1*y1+c12*y1*y2+ &
82. 0.5*c22*y2*y2+a11*w1*y1+a12*w1*y2+a21*w2*y1+a22*w2*y2
83. *costs without the error_Unnormalized
84. genr Cost_UN=
    (b0+b1*w1+b2*w2+a1*y1+a2*y2+0.5*b11*w1*w1+b12*w1*w2+0.5*b22*w2*w2+0.5
    *c11*y1*y1+c12*y1*y2+
    0.5*c22*y2*y2+a11*w1*y1+a12*w1*y2+a21*w2*y1+a22*w2*y2)*w33
85. *Incremental Costs
86. genr ic_i_y1= a1*y1+.5*c11*y1*y1+c12*y1*y2+a11*w1*y1+a21*w2*y1
87. genr ic_i_y2= a2*y2+.5*c22*y2*y2+c12*y1*y2+a12*w1*y2+a22*w2*y2
88. *Marginal Costs
89. genr mc_y1= a1+c11*y1+c12*y2+a11*w1+a21*w2
90. genr mc_y2= a2+c22*y2+c12*y1+a12*w1+a22*w2
91. *calculate input demands using input price and output quantities
92. genr x1=b1+b11*w1+b12*w2+a11*y1+a12*y2
93. genr x2=b2+b12*w1+b22*w2+a21*y1+a22*y2
94. genr x3=(TotalCost-x1*w1-x2*w2)
```

```
*Costs of producing single outputs
genr cy2= TotalCost-ic_i_y1
genr cy1= TotalCost-ic_i_y2
*Scope
genr scope=(cy2+cy1-TotalCost)/TotalCost
*xeff
genr xeff=TotalCost/cs
*xerrors
genr x1err=x1/xeff
genr x2err=x2/xeff
genr x3err=x3/xeff
*PSE MSE
genr mse=TotalCost/(mc_y1*y1+mc_y2*y2)
genr pse_y1=ic_i_y1/(y1*mc_y1)
genr pse_y2=ic_i_y2/(y2*mc_y2)
write ("file name and destination") variable 1 variable 2 variable 3…….
```

A.2 Nonparametric Estimation Using General Algebraic Modeling Software (GAMS)

The code shown below was read in using text documents compiled from the data sets generated from the above Monte Carlo program. For the first paper this program was run a total of 5 times. The first run's results yielded the model "No Inefficiency" using a data set with all the x variables on the frontier run a shown below. Changes to the distribution came from the input data, not the code itself. The model shown has the output forced to zero for the incremental cost calculations by including the constraints shown in lines 56 and 58 in lines 88 and 105. For the model where the output constraint is dropped, delete the constraints from lines 88 and 105.

Nonparametric GAMS Programming Code

```
1. *CODE TO CALCULATE NONPARAMETRIC
2. *PRODUCT SPECIFIC SCALE EFFICIENCY
3. *$OFFSYMXREF OFFSYMLIST OFFUELXREF  OFFUELLIST
4. *$OFFSYMLIST OFFSYMXREF
5. *$OFFLISTING
6. *OPTION LIMCOL=0;
7. *OPTION LIMROW=0;
```

```
OPTION DECIMALS=7;
SETS
P OUTPUTS /Y1, Y2/
N INPUTS /X1,X2,X3/
K OBSERVATIONS /F1*F500/;
TABLE I(K,N) INPUT LEVELS
$INCLUDE "file inputs";
TABLE IP(K,N) INPUT PRICES
$INCLUDE "file prices";
TABLE Y(K,P) OUTPUT LEVELS
$INCLUDE "file output levels";
*PARAMETERS IN PLACE TO PREPARE FOR THE LOOP
PARAMETER R2(N)
/x1 1
x2  1
x3  1/;
PARAMETER R3(P)
/Y1 1
Y2 1/;
POSITIVE VARIABLES
Z(K) INTENSITY MEASURE
XI(N) OPTIMAL INPUT LEVEL;
VARIABLES
CA COST OBJ FUNCTION VALUE FOR ALL OUTPUTS
CY1 COST OBJ FUNCTION VALUE FOR Y1 ONLY
CY2 COST OBJ FUNCTION VALUE FOR Y2 ONLY;
EQUATIONS
OBJA OBJECTIVE FUNCTION FOR ALL OUTPUTS
OBJY1 OBJECTIVE FUNCTION FOR Y1 ONLY
OBJY2 OBJECTIVE FUNCTION FOR Y2 ONLY
CON1 INPUT CONSTRAINT
ALLOUT OUTPUT CONSTRAINT INCLUDING BOTH Y1 AND Y2
Y1OUT OUTPUT CONSTRAINT INCLUDING ONLY Y1
Y2OUT OUTPUT CONSTRAINT INCLUDING ONLY Y2
Y1OUTZERO OUTPUT CONSTRAINT FORCING Y1 OUTPUT TO BE ZERO
Y2OUTZERO OUTPUT CONSTRAINT FORCING Y2 OUTPUT TO BE ZERO
CON3 Z CONSTRAINT;
OBJA.. CA=E=SUM(N,XI(N)*R2(N));
OBJY1.. CY1=E=SUM(N,XI(N)*R2(N));
OBJY2.. CY2=E=SUM(N,XI(N)*R2(N));
CON1(N).. SUM(K, I(K,N)*Z(K))=L=XI(N);
*CONSTRAINT FOR ALL OUTPUTS
ALLOUT(P).. SUM(K, Y(K,P)*Z(K))-R3(P)=G=0;
*CONSTRAINT EXCLUDING Y2
Y1OUT("Y1").. SUM(K, Y(K,"Y1")*Z(K))-R3("Y1")=G=0;
*CONSTRAINT EXCLUDING Y1
```

54. Y2OUT("Y2").. SUM(K, Y(K,"Y2")*Z(K))-R3("Y2")=G=0;
55. *CONSTRAINT FORCING Y1 TO BE ZERO
56. Y1OUTZERO("Y1").. SUM(K, Y(K,"Y1")*Z(K))=e=0;
57. *CONSTRAINT FORCING Y2 TO BE ZERO
58. Y2OUTZERO("Y2").. SUM(K, Y(K,"Y2")*Z(K))=e=0;
59. *SET THIS CONSTRAINT LESS THAN OR EQUAL TO 1 FOR CONSTANT RETURNS TO SCALE
60. *SET THIS CONSTRAINT EQUAL TO 1 FOR VARIABLE RETURNS TO SCALE
61. CON3.. SUM(K,Z(K))=E=1;
62. ***************MODEL FOR ALL OUTPUTS*********************
63. MODEL MODALL /OBJA, CON1, ALLOUT, CON3/;
64. MODALL.WORKSPACE=7;
65. SETS
66. ITER1 /I1*I500/;
67. PARAMETER ACOSTS(ITER1) TOTAL COST OF PRODUCING ALL OUTPUTS;
68. PARAMETER MCOSTA (ITER1,P) MARGINAL COSTS OF OUTPUTS;
69. PARAMETER ZY2A (ITER1) LIVESTOCK OUTPUT AS CALCULATED WITH INTENSITY FACTORS;
70. PARAMETER ZY1A (ITER1) CROP OUTPUT AS CALCULATED WITH INTENSITY FACTORS;
71. PARAMETER ZVALSA (K,ITER1) Z VALUES OF EACH ITERATION;
72. TABLE VAL1(ITER1,N) INPUT LEVELS
73. $INCLUDE "file inputs";
74. TABLE VAL2(ITER1,N) INPUT PRICES
75. $INCLUDE "file input prices";
76. TABLE VAL3(ITER1,P) OUTPUT LEVELS
77. $INCLUDE "file output levels";
78. LOOP(ITER1,
79. R3(P)=VAL3(ITER1,P);
80. R2(N)=VAL2(ITER1,N);
81. SOLVE MODALL USING NLP MINIMIZING CA;
82. ACOSTS(ITER1)=CA.L;
83. MCOSTA(ITER1,P)=ALLOUT.M(P);
84. ZY1A (ITER1)=SUM(K,Z.L(K)*Y(K,"Y1"));
85. ZY2A (ITER1)=SUM(K,Z.L(K)*Y(K,"Y2"));
86. ZVALSA (K,ITER1)=Z.L(K););
87. ***************MODEL FOR ONLY Y2 ONLY*********************
88. MODEL MODY2 /OBJY2, CON1, Y2OUT, Y1OUTZERO, CON3/;
89. MODY2.WORKSPACE=1.5;
90. PARAMETER Y2COSTS(ITER1) TOTAL COST OF PRODUCING ALL OUTPUTS;
91. PARAMETER MCOSTY2 (ITER1,P) MARGINAL COSTS OF OUTPUTS;
92. PARAMETER ZY2_Y2 (ITER1) Y2 OUTPUT AS CALCULATED WITH INTENSITY FACTORS;
93. PARAMETER Zy1_Y2 (ITER1) Y1 OUTPUT AS CALCULATED WITH INTENSITY FACTORS;
94. PARAMETER ZVALSY2 (K,ITER1) Z VALUES OF EACH ITERATION;

```
LOOP(ITER1,
R3(P)=VAL3(ITER1,P);
R2(N)=VAL2(ITER1,N);
SOLVE MODY2 USING NLP MINIMIZING CY2;
Y2COSTS(ITER1)=CY2.L;
MCOSTY2(ITER1,P)=Y2OUT.M(P);
ZY1_Y2 (ITER1)=SUM(K,Z.L(K)*Y(K,"Y1"));
ZY2_Y2 (ITER1)=SUM(K,Z.L(K)*Y(K,"Y2"));
ZVALSY2 (K,ITER1)=Z.L(K););
***********MODEL FOR ONLY Y1*********************
MODEL MODY1 /OBJY1, CON1, Y1OUT,Y2OUTZERO,  CON3/;
MODY1.WORKSPACE=1.5;
PARAMETER Y1COSTS (ITER1) TOTAL COST OR PRODUCING ONLY Y1;
PARAMETER MCOSTY1 (ITER1,P) MARGINAL COSTS OF OUTPUTS;
PARAMETER ZY2_Y1 (ITER1) Y2 OUTPUT AS CALCULATED WITH INTENSITY FACTORS;
PARAMETER ZY1_Y1 (ITER1) Y1 OUTPUT AS CALCULATED WITH INTENSITY FACTORS;
PARAMETER ZVALSY1 (K,ITER1) Z VALUES OF EACH ITERATION;
LOOP(ITER1,
R3(P)=VAL3(ITER1,P);
R2(N)=VAL2(ITER1,N);
SOLVE MODY1 USING NLP MINIMIZING CY1;
Y1COSTS(ITER1)=CY1.L;
MCOSTY1(ITER1,P)=Y1OUT.M(P);
ZY1_Y1 (ITER1)=SUM(K,Z.L(K)*Y(K,"Y1"));
ZY2_Y1 (ITER1)=SUM(K,Z.L(K)*Y(K,"Y2"));
ZVALSY1 (K,ITER1)=Z.L(K););
*WRITE OUTPUT INTO SPACE DELIMITED FILES*
********PRODUCTION INTENSITY MEASURES********
*Z VALUES FROM MODEL FOR ALL OUTPUTS*
FILE ALLZ /"file destination"/;
ALLZ.PW=5000;
ALLZ.ND=4;
ALLZ.PC=4;
PUT ALLZ;
PUT 'PRODUCTION INTENSITY FACTORS FROM MODEL FOR ALL OUTPUTS'//;
PUT ' '; LOOP(ITER1, PUT ITER1.TL);
LOOP(K,
PUT/K.TE(K);
LOOP(ITER1,PUT ZVALSA(K,ITER1)););
*Z VALUES FROM MODEL FOR ONLY Y1*
FILE Y1Z / file destination /;
Y1Z.PW=5000;
```

```
Y1Z.ND=4;
Y1Z.PC=4;
PUT Y1Z;
PUT 'PRODUCTION INTENSITY FACTORS FROM MODEL FOR ONLY Y1'//;
PUT ' '; LOOP(ITER1, PUT ITER1.TL);
LOOP(K,
PUT/K.TE(K);
LOOP(ITER1,PUT ZVALSY1(K,ITER1)););
*Z VALUES FROM MODEL FOR ONLY Y2*
FILE Y2Z / file destination /;
Y2Z.PW=5000;
Y2Z.ND=4;
Y2Z.PC=4;
PUT Y2Z;
PUT 'PRODUCTION INTENSITY FACTORS FROM MODEL FOR ONLY Y2'//;
PUT ' '; LOOP(ITER1, PUT ITER1.TL);
LOOP(K,
PUT/K.TE(K);
LOOP(ITER1,PUT ZVALSY2(K,ITER1)););
**********CREATE FILE FOR TOTAL COST, MARGINAL COSTS, INCREMENTAL COST RESULTS**********
*TOTAL AND MARGINAL COSTS
FILE RESULTS / file destination /;
RESULTS.PW=5000;
RESULTS.ND=4;
RESULTS.PC=4;
PUT RESULTS;
PUT 'COSTS'//;
PUT ' ';LOOP(P, PUT P.TL);
PUT 'TOTAL COST' 'Y2COST' 'Y1COST' 'ZY2A' 'ZY1A' 'ZY2_Y2' 'ZY1_Y2' 'ZY2_Y1' 'ZY1_Y1';
LOOP(ITER1,
PUT/ITER1.TE(ITER1);
LOOP(P,PUT MCOSTA(ITER1,P)); PUT ACOSTS(ITER1) Y2COSTS(ITER1)
Y1COSTS(ITER1) ZY2A(ITER1) ZY1A(ITER1) ZY2_Y2(ITER1) ZY1_Y2(ITER1)
ZY2_Y1(ITER1) ZY1_Y1(ITER1));
Display MCOSTA
```

Appendix B - Software Codes Used in Chapter 3

B.1 Two-Sided Error Model

This routine is specified for the SHAZAM software platform. This code reads in prices and quantities and fits a traditional two-sided error regression curve to the data generated as shown in Appendix A. The economic measures are calculated within the code below and read out into an excel document.

```
1.  file 13 “file path”
2.  Read (13)  Firm pin1  pin2  pin3  pout1  pout2  out1  out2  in1  in2  in3;
3.  stat pin1  pin2  pin3  pout1  pout2  out1  out2  in1  in2  in3;
4.  genr b1=in1
5.  genr b2=in2
6.  genr b3=in3
7.  genr y1=out1
8.  genr y2=out2
9.  genr c1=pin1
10. genr c2=pin2
11. genr c3=pin3
12. genr d1=pout1
13. genr d2=pout2
14. genr w1=c1/c3
15. genr w2=c2/c3
16. genr x1=b1
17. genr x2=b2
18. genr x3=b3
19. genr cost=c1*b1+c2*b2+c3*b3
20. genr costn=cost/c3
21. stat w1/ mean= mw1
22. stat w2/ mean= mw2
23. stat y1/ mean= my1
24. stat y2/ mean= my2
25. nl 3 / ncoef = 15 iter = 2000 piter = 100 genrvar conv = .0000001
26. eq COSTN = A0 + A1*W1 + A2*W2  &
27. + AY1*Y1 + AY2*Y2 + 0.5*W11*W1*W1 + W12*W1*W2 + .5*W22*W2*W2  +
    .5*Y11*Y1*Y1 &
28. + Y12*Y1*Y2 + .5*Y22*Y2*Y2 + YW11*W1*Y1 + YW12*Y1*W2  &
29. + YW21*W1*Y2 + YW22*Y2*W2
```

```
30. eq X1=A1+W11*W1+W12*W2+YW11*Y1+YW21*Y2
31. eq X2=A2+W12*W1+W22*W2+YW12*Y1+YW22*Y2
32. ************************************************************
33. ************  set y1 = 0 and then y2 = 0 in total cost**************
34. *****************************************************
35. *Total Cost
36. genr ttlcst = A0 + A1*W1 + A2*W2  &
37. + AY1*Y1 + AY2*Y2 + 0.5*W11*W1*W1 + W12*W1*W2 + .5*W22*W2*W2  +
    .5*Y11*Y1*Y1 &
38. + Y12*Y1*Y2 + .5*Y22*Y2*Y2 + YW11*W1*Y1 + YW12*Y1*W2  &
39. + YW21*W1*Y2 + YW22*Y2*W2
40. *Cost function for Y1 (y2=0)
41. genr costy1 =A0 + A1*W1 + A2*W2  &
42. + AY1*Y1  + 0.5*W11*W1*W1 + W12*W1*W2  &
43. + .5*W22*W2*W2 + .5*Y11*Y1*Y1 + YW12*Y1*W2 + YW11*W1*Y1
44. *Cost function for Y2 (y1=0)
45. genr costy2 = A0 + A1*W1 +A2*W2  &
46. + AY2*Y2 + 0.5*W11*W1*W1 + W12*W1*W2 &
47. + .5*W22*W2*W2 + .5*Y22*Y2*Y2 + YW21*W1*Y2 + YW22*Y2*W2
48. print ttlcst
49. print costy1
50. print costy2
51. Estimate dC/dYi = MCYI
52. genr mcy1 =  AY1 + Y11*Y1 + Y12*Y2 &
53. + YW11*W1 + YW12*W2
54. genr mcy2 =  AY2 + Y12*Y1 + Y22*Y2 &
55. + YW21*W1 + YW22*W2
56. matrix P = MCY1|MCY2
57. *Multiproduct Scale Economies
58. genr MPSE = ttlcst/(y1*mcy1 + y2*mcy2)
59. genr scope = (COSTy1 + COSTy2 - ttlcst)/ttlcst
60. *generate IC1
61. genr IC1 = ttlcst -  (A0 + A1*W1 +A2*W2  &
62. + AY2*Y2 + 0.5*W11*W1*W1 + W12*W1*W2 &
63. + .5*W22*W2*W2 + .5*Y22*Y2*Y2 + YW21*W1*Y2 + YW22*Y2*W2)
64. *generate IC2
65. genr IC2 = ttlcst - (A0 + A1*W1 + A2*W2  &
66. + AY1*Y1  + 0.5*W11*W1*W1 + W12*W1*W2  &
67. + .5*W22*W2*W2 + .5*Y11*Y1*Y1 + YW11*W1*Y1 + YW12*Y1*W2 )
68. genr PSEY1 = IC1/(y1*mcy1)
69. genr PSEY2 = IC2/(y2*mcy2)
70. print IC1 IC2 PSEY1 PSEY2
71. write (file) ttlcst mcy1 mcy2 IC1 IC2
72. write (file) MPSE PSEY1 PSEY2 scope
73. stop
```

B.2 OLS Positive Errors Model

The following estimation was run using the GAMS software package. Prior to running the following code, the independent variable are multiplied out for each observation in the form of a normalized quadratic function. The total cost for each firm is copied and pasted directly into this code between lines 14 and 15 while the RHS variables are read in from a table (Lines 9-12). Calculations for cost efficiency, multi-product scale economies, product-specific scale economies, and economies are done in excel using the parameters estimated from this model.

```
1.  Regression
2.  *$OFFSYMXREF OFFSYMLIST OFFUELXREF  OFFUELLIST
3.  *$OFFSYMLIST OFFSYMXREF
4.  *$OFFLISTING
5.  *OPTION LIMCOL=0;
6.  *OPTION LIMROW=0;
7.  OPTION DECIMALS=7;
8.  SETS
9.  D DEPENDENTS /B0,B1,B2,B3,B4,B5,B6,B7,B8,B9,B10,B11,B12,B13,B14/
10. i OBSERVATIONS /F1*F500/;
11. TABLE X(i,D) DEPENDENTS
12. $INCLUDE "File Location "
13. parameter y(i)
14. / F1 " Total Cost for each firm"……..
15. /
16. variables
17. sse       objective function
18. gamma(D)   parameter estimates
19. e(i)      deviations;
20. positive variable e;
21. equations
22. obj
23. dev(i)  ;
24. obj.. sse =e= sum(i,power(e(i),2)); ;
25. dev(i).. e(i) =e= y(i)/1 - sum(D,gamma(D)*X(i,D));
26. model reg /obj,dev/;
27. options nlp=conopt;
28. options lp=minos;
29. solve reg using nlp minimizing sse;
```

B.3 Stochastic Frontier Model

The stochastic frontier model was estimated using FRONT V4.1 written by Tim Coelli and downloaded from (http://www.uq.edu.au/economics/cepa/frontier.php). The code shown below is a text document read into the program which specifies the models attributes, data source, and output destination. In using this specification, a normalized quadratic cost frontier is estimated for the case with 500 observations under the uniform distribution, a constant, and 14 coefficients. For this we read in the text document as follows:

```
1            1=ERROR COMPONENTS MODEL, 2=TE EFFECTS MODEL
Uniform-dta.txt      DATA FILE NAME
Uniform-out.txt      OUTPUT FILE NAME
2            1=PRODUCTION FUNCTION, 2=COST FUNCTION
n            LOGGED DEPENDENT VARIABLE (Y/N)
500            NUMBER OF CROSS-SECTIONS
1            NUMBER OF TIME PERIODS
500            NUMBER OF OBSERVATIONS IN TOTAL
14            NUMBER OF REGRESSOR VARIABLES (Xs)
n            MU (Y/N) [OR DELTA0 (Y/N) IF USING TE EFFECTS MODEL]
n            ETA (Y/N) [OR NUMBER OF TE EFFECTS REGRESSORS (Zs)]
n            STARTING VALUES (Y/N)
IF YES THEN    BETA0
BETA1 TO
BETAK
SIGMA SQUARED
GAMMA
MU            [OR DELTA0
ETA              DELTA1 TO
DELTAP]
```

When estimating different data sets, the input file and output file names are changed and for the 400 observations case, the number of cross-sections and observations is changed.

B.4 Nonparametric Model

See Appendix A for nonparametric estimation code instructions.

Appendix C - Software Codes Used in Chapter 4

Below is the program to estimate the nonparametric cost frontier for the KFMA data set using GAMS. Since there are no zero output firms, the incremental costs are calculated by dropping the appropriate output constraints rather than forcing it to equal zero. There are 8 inputs, 8 input prices, and 2 outputs.

C.1 Software Code for Cost Frontier Estimation Using KFMA Data

```
*CODE TO CALCULATE NONPARAMETRIC
*PRODUCT SPECIFIC SCALE EFFICIENCY
$OFFSYMXREF OFFSYMLIST OFFUELXREF  OFFUELLIST
$OFFSYMLIST OFFSYMXREF
$OFFLISTING
OPTION LIMCOL=0;
OPTION LIMROW=0;
OPTION DECIMALS=7;
SETS
P OUTPUTS /YCROP, YLIVE/
N INPUTS /yseed,yfert,ychem,yfeed,yfuel,ylab,yland,ymach/
K OBSERVATIONS /F1*F2410/;
TABLE I(K,N) INPUT LEVELS
$INCLUDE "file";
TABLE IP(K,N) INPUT PRICES
$INCLUDE "file";
TABLE Y(K,P) OUTPUT LEVELS
$INCLUDE "file";
*PARAMETERS IN PLACE TO PREPARE FOR THE LOOP
PARAMETER R2(N)
/yseed 1
yfert  1
ychem  1
yfeed  1
yfuel  1
ylab   1
yland  1
ymach 1/;
PARAMETER R3(P)
/YCROP 1
YLIVE 1/;
POSITIVE VARIABLES
Z(K) INTENSITY MEASURE
```

```
XI(N) OPTIMAL INPUT LEVEL;
VARIABLES
CA COST OBJ FUNCTION VALUE FOR ALL OUTPUTS
CC COST OBJ FUNCTION VALUE FOR CROPS ONLY
CL COST OBJ FUNCTION VALUE FOR LIVESTOCK ONLY;
EQUATIONS
OBJA OBJECTIVE FUNCTION FOR ALL OUTPUTS
OBJC OBJECTIVE FUNCTION FOR CROPS ONLY
OBJL OBJECTIVE FUNCTION FOR LIVESTOCK ONLY
CON1 INPUT CONSTRAINT
ALLOUT OUTPUT CONSTRAINT INCLUDING BOTH CROPS AND LIVESTOCK
CROPOUT OUTPUT CONSTRAINT INCLUDING ONLY CROPS
LIVEOUT OUTPUT CONSTRAINT INCLUDING ONLY LIVESTOCK
CON3 Z CONSTRAINT;
OBJA.. CA=E=SUM(N,XI(N)*R2(N));
OBJC.. CC=E=SUM(N,XI(N)*R2(N));
OBJL.. CL=E=SUM(N,XI(N)*R2(N));
CON1(N).. SUM(K, I(K,N)*Z(K))=L=XI(N);
*CONSTRAINT FOR ALL OUTPUTS
ALLOUT(P).. SUM(K, Y(K,P)*Z(K))-R3(P)=g=0;
*CONSTRAINT EXCLUDING LIVESTOCK
CROPOUT("YCROP").. SUM(K, Y(K,"YCROP")*Z(K))-R3("YCROP")=g=0;
*CONSTRAINT EXCLUDING CROPS
LIVEOUT("YLIVE").. SUM(K, Y(K,"YLIVE")*Z(K))-R3("YLIVE")=g=0;
*SET THIS CONSTRAINT LESS THAN OR EQUAL TO 1 FOR CONSTANT RETURNS TO SCALE
*SET THIS CONSTRAINT EQUAL TO 1 FOR VARIABLE RETURNS TO SCALE
CON3.. SUM(K,Z(K))=E=1;
***************MODEL FOR ALL OUTPUTS*********************
MODEL MODALL /OBJA, CON1, ALLOUT, CON3/;
MODALL.WORKSPACE=2.5;
SETS
ITER1 /I1*I2410/;
PARAMETER ACOSTS(ITER1) TOTAL COST OF PRODUCING ALL OUTPUTS;
PARAMETER MCOSTA (ITER1,P) MARGINAL COSTS OF OUTPUTS;
PARAMETER ZLIVEA (ITER1) LIVESTOCK OUTPUT AS CALCULATED WITH INTENSITY FACTORS;
PARAMETER ZCROPA (ITER1) CROP OUTPUT AS CALCULATED WITH INTENSITY FACTORS;
PARAMETER ZVALSA (K,ITER1) Z VALUES OF EACH ITERATION;
TABLE VAL1(ITER1,N) INPUT LEVELS
$INCLUDE "file";
TABLE VAL2(ITER1,N) INPUT PRICES
$INCLUDE "file";
TABLE VAL3(ITER1,P) OUTPUT LEVELS
$INCLUDE "file";
```

```
LOOP(ITER1,
R3(P)=VAL3(ITER1,P);
R2(N)=VAL2(ITER1,N);
SOLVE MODALL USING NLP MINIMIZING CA;
ACOSTS(ITER1)=CA.L;
MCOSTA(ITER1,P)=ALLOUT.M(P);
ZCROPA (ITER1)=SUM(K,Z.L(K)*Y(K,"YCROP"));
ZLIVEA (ITER1)=SUM(K,Z.L(K)*Y(K,"YLIVE"));
ZVALSA (K,ITER1)=Z.L(K););
***************MODEL FOR ONLY LIVESTOCK*********************
MODEL MODLIVE /OBJL, CON1, LIVEOUT, CON3/;
MODLIVE.WORKSPACE=2.5;
PARAMETER LCOSTS(ITER1) TOTAL COST OF PRODUCING ALL OUTPUTS;
PARAMETER MCOSTL (ITER1,P) MARGINAL COSTS OF OUTPUTS;
PARAMETER ZLIVEL (ITER1) LIVESTOCK OUTPUT AS CALCULATED WITH INTENSITY FACTORS;
PARAMETER ZCROPL (ITER1) CROP OUTPUT AS CALCULATED WITH INTENSITY FACTORS;
PARAMETER ZVALSL (K,ITER1) Z VALUES OF EACH ITERATION;
LOOP(ITER1,
R3(P)=VAL3(ITER1,P);
R2(N)=VAL2(ITER1,N);
SOLVE MODLIVE USING NLP MINIMIZING CL;
LCOSTS(ITER1)=CL.L;
MCOSTL(ITER1,P)=LIVEOUT.M(P);
ZCROPL (ITER1)=SUM(K,Z.L(K)*Y(K,"YCROP"));
ZLIVEL (ITER1)=SUM(K,Z.L(K)*Y(K,"YLIVE"));
ZVALSL (K,ITER1)=Z.L(K););
************MODEL FOR ONLY CROPS********************
MODEL MODCROP /OBJC, CON1, CROPOUT, CON3/;
MODCROP.WORKSPACE=2.5;
PARAMETER CCOSTS (ITER1) TOTAL COST OR PRODUCING ONLY CROPS;
PARAMETER MCOSTC (ITER1,P) MARGINAL COSTS OF OUTPUTS;
PARAMETER ZLIVEC (ITER1) LIVESTOCK OUTPUT AS CALCULATED WITH INTENSITY FACTORS;
PARAMETER ZCROPC (ITER1) CROP OUTPUT AS CALCULATED WITH INTENSITY FACTORS;
PARAMETER ZVALSC (K,ITER1) Z VALUES OF EACH ITERATION;
LOOP(ITER1,
R3(P)=VAL3(ITER1,P);
R2(N)=VAL2(ITER1,N);
SOLVE MODCROP USING NLP MINIMIZING CC;
CCOSTS(ITER1)=CC.L;
MCOSTC(ITER1,P)=CROPOUT.M(P);
ZCROPC (ITER1)=SUM(K,Z.L(K)*Y(K,"YCROP"));
ZLIVEC (ITER1)=SUM(K,Z.L(K)*Y(K,"YLIVE"));
```

```
ZVALSC (K,ITER1)=Z.L(K););
*WRITE OUTPUT INTO SPACE DELIMITED FILES*
********PRODUCTION INTENSITY MEASURES********
*Z VALUES FROM MODEL FOR ALL OUTPUTS*
FILE ALLZ /file/;
ALLZ.PW=5000;
ALLZ.ND=4;
ALLZ.PC=4;
PUT ALLZ;
PUT 'PRODUCTION INTENSITY FACTORS FROM MODEL FOR ALL OUTPUTS'//;
PUT ' '; LOOP(ITER1, PUT ITER1.TL);
LOOP(K,
PUT/K.TE(K);
LOOP(ITER1,PUT ZVALSA(K,ITER1)););
*Z VALUES FROM MODEL FOR ONLY CROPS*
FILE CROPZ /file/;
CROPZ.PW=5000;
CROPZ.ND=4;
CROPZ.PC=4;
PUT CROPZ;
PUT 'PRODUCTION INTENSITY FACTORS FROM MODEL FOR ONLY CROPS'//;
PUT ' '; LOOP(ITER1, PUT ITER1.TL);
LOOP(K,
PUT/K.TE(K);
LOOP(ITER1,PUT ZVALSC(K,ITER1)););
*Z VALUES FROM MODEL FOR ONLY LIVESTOCK*
FILE LIVEZ /file/;
LIVEZ.PW=5000;
LIVEZ.ND=4;
LIVEZ.PC=4;
PUT LIVEZ;
PUT 'PRODUCTION INTENSITY FACTORS FROM MODEL FOR ONLY LIVESTOCK'//;
PUT ' '; LOOP(ITER1, PUT ITER1.TL);
LOOP(K,
PUT/K.TE(K);
LOOP(ITER1,PUT ZVALSL(K,ITER1)););
**********CREATE FILE FOR TOTAL COST, MARGINAL COSTS, INCREMENTAL COST RESULTS**********
*TOTAL AND MARGINAL COSTS
FILE RESULTS /file/;
RESULTS.PW=5000;
RESULTS.ND=4;
RESULTS.PC=4;
PUT RESULTS;
```

```
PUT 'COSTS'//;
PUT ' ';LOOP(P, PUT P.TL);
PUT 'TOTAL COST' 'LCOST' 'CCOST' 'ZLIVEA' 'ZCROPA' 'ZLIVEL' 'ZCROPL' 'ZLIVEC' 'ZCROPC';
LOOP(ITER1,
PUT/ITER1.TE(ITER1);
LOOP(P,PUT MCOSTA(ITER1,P)); PUT ACOSTS(ITER1) LCOSTS(ITER1)
CCOSTS(ITER1) ZLIVEA(ITER1) ZCROPA(ITER1) ZLIVEL(ITER1) ZCROPL(ITER1)
ZLIVEC(ITER1) ZCROPC(ITER1));
```