Are Efficient Farms and Inefficient Farms Heterogeneous?

Youngjune Kim
PhD student
Department of Agricultural Economics
Kansas State University
yjkim@ksu.edu

Bowen Chen
PhD student
Department of Agricultural Economics
Kansas State University
bwchen@ksu.edu

Allen M. Featherstone
Professor
Department of Agricultural Economics
Kansas State University
afeather@ksu.edu

Dustin L. Pendell
Associate Professor
Department of Agricultural Economics
Kansas State University
dpendell@ksu.edu

Key words: Heterogeneity, Super DEA, Quantile regression, Nonparametric efficiency

Selected Paper prepared for presentation for the 2017, Southern Agricultural Economic Association, Mobile, AL, February 4-7.

1. Introduction

Cow-calf producers in the Kansas experienced one of the largest economic returns in the last 50 years in 2014 (Pendell, Kim, and Herbel). A drought and strengthening beef demand enable it possible to record the highest average return in 2014. Despite the highest average return in 2014 over last 50 years, the amount of profitability was significantly variable across producers. The difference in average annual returns over variable costs between low-profit and high- profit cow-calf producers is almost $670 per cow in 2009~2014 (Pendell, Kim, and Herbel). An important question that arises from the variability is that what characteristics determine these differences in profitability. Featherstone et al. (1997) noted that production efficiency is the main factor to increases the competitiveness of the beef sector. They found that feed, labor and capital cost are relatively important in explaining efficiency. In addition, herd size and degree of specialization were negatively correlated with efficiency.

The production frontier approach has been widely used to measure efficiency in agricultural economics (Andreu and Grunewald, 2006). Data Envelopment Analysis (DEA) is one of the techniques developed for efficiency measurement.

The typical steps of DEA analysis are: first, measuring the efficiency values by non-parametric technique, and second, examining the sources of inefficiency by utilizing censored models. Empirical works closely follows the efficiency measuring to examine the sources of inefficiency by using the censored models.

However, the underlying assumption behind them is that both efficient farms and inefficient farms are homogenous from a standpoint of efficiency, even though each may have a different strategy to increase efficiency. To the best of our knowledge, the consideration of heterogeneity in DEA analysis at both stages is a rare event in the literature.

A limited number of studies have attempted to explain the heterogeneity. Samoilenko and Osei-Bryson (2007) raise the concern of diversity in the set of Decision Making Units (DMUs) and propose a three-step methodology allowing for the presence of the heterogeneity of the sample. The three steps are: cluster analysis, phase DEA analysis and decision tree utilization. Actually, heterogeneity is an artificial issue in their case because the data collected are from 25 countries, which is too broad. Similarly, Wang et al (2013) considers heterogeneity of production technology among provinces in China. The preliminary research on the spatial and technological heterogeneity increases the discriminatory power of DEA to some extent, but is still not convincing enough due to the fundamental assumption of homogeneity for DEA analysis. The ignorance of heterogeneity is also an intensively debated issue with regards to censored models; the models used at the second stage of DEA analysis. Neglecting across individual heterogeneity can lead to inconsistent or inefficient estimators, and to misleading predictions (Chesher, 1984). Much effort has been made to the diagnosis of the misspecification, and the main approach of which is on the basis of the information matrix. Blundell and Meghir (2002) apply the matrix to test independence and normality of the Tobit model in the case of household demand for clothing and to married women's labor supply. Similar work is done by Reynolds and Shonkwiler (1991) and Blundell and Meghir (2002). However, the admittance of heterogeneity at the second stage is a dangerous attempt because it would necessarily leave the efficiency values obtained at the first stage under scrutiny.

Under the standard DEA, all DMUs have a unity score along an efficient frontier. Since technical efficiency cannot be greater than one, studies that use both DEA and Tobit fail to capture the marginal effects of input variables on the efficiency improvement for each efficient group and inefficient group, separately, which leaves the heterogeneity in the

marginal effects hard to be examined.

Recent developments in DEA methods, so called “Super DEA”, which produces the ranking information of the efficient DMUs, facilitates study on the heterogeneity. Anderson and Peterson (1993) introduce the methodology calling it a modified version of DEA based upon comparison of efficient DMUs relative to a reference technology spanned by all other units. One of the important property associated with the super DEA is that the efficiency value of a unit will be assigned as 1 or higher if and only if the unit is efficient. This makes it possible to discriminate the performance among efficient DMUs.

The objective of this study is to examine the sources of inefficiency with particular consideration on the heterogeneity at the farm level. To accomplish this, we exploit a methodology called super DEA. For efficient farms, the technical efficiency scores can be greater than one, by dropping the constraint that bounds the score of evaluated farm. We can eventually make it possible to relax the underlying homogeneity assumption. The quantile regression approach enables us to derive different marginal effect of the explanatory variables for each the efficient group and non-efficient group.

2. Methodology

This study employs a two-stage methodology to investigate the heterogeneity of efficient and inefficient farms. In the first stage, a nonparametric approach called Super DEA, which produces the ranking information in the efficient DMUs, facilitates the study on the heterogeneity. In the second stage, a quantile regression is used to identify the sources of efficiency for efficient farms and inefficient farms, respectively.

The Super DEA score for the inefficient unit is the same as the standard DEA score. In

order to measure the efficient units, the efficient units are allowed to have an efficiency score greater than one by dropping the constraint that bounds the score of evaluated unit.

Figure 1 is an example of a single output, and consuming two goods X_1 and X_2. For Standard DEA models, efficient units lie on the efficient frontier i.e. A, B and C and other DMUs produce the same output with a lower input combination. Unit D is dominated by the other three DMU's and produces the same amount with a higher input combination. The inefficiency of unit D can be measured by its radial distance to the frontier along ray extending from origin to D and intersecting the AB segment of frontier.

If B was excluded from the frontier, in order to calculate ranking of the efficient set of DMUs, a new frontier would be considered using only units A and C. The super-efficient score for unit B is obtainable by calculating its distance to the new frontier. This allow the scores for efficient units to exceed unity, 100%.

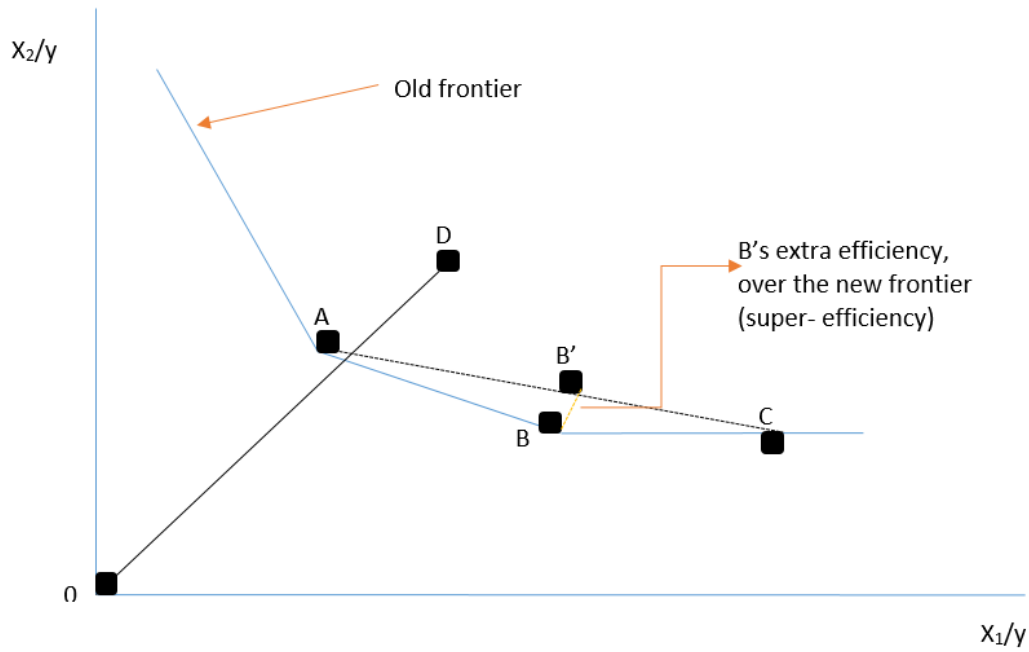

Figure 1: Standard and Super-Efficient DEA Models

The slope parameters of the explanatory variables on efficiency for the efficient farms and inefficient farms may be different from each other. Therefore, instead of estimating the OLS regression that is based on the conditional mean, estimating different estimations for each efficient group and inefficient group via quantile regression (Koenker and Bassett, 1978) will allow us to find out different marginal effect of the explanatory variables for each group.

The quantile regression estimator, $\widehat{\beta_\tau}$, can be derived by minimizing the following linear programming problem;

(1) $$Q(\beta_\tau) = \sum_{i:\, y_i \geq x_i{}'\beta}^{N} \tau |y_i - x_i{}'\beta_\tau| + \sum_{i:\, y_i < x_i{}'\beta}^{N} (1-\tau)|y_i - x_i{}'\beta_\tau|$$

where τ denotes quartiles, y_i is a dependent variable and x_i is a vector of explanatory variabl es. If τ = 0.5, this problem becomes minimizing the sum of the absolute deviations and is the same as the median regression.

3. Data

The data used in this study are from the Kansas Farm Management Association (KFMA). The empirical analysis was applied to 374 farms that kept beef cows during 2014. All observations with at least one zero inputs were dropped from the sample since they are measured in natural logarithm. Following Featherstone et al. (1997), six inputs were used: feed, labor, capital, utilities, veterinary, and miscellaneous to derive the efficiency ranking estimates. In addition to input variables, five farm characteristics variables were used to examine the relationship between farm characteristics and the efficiency measures: age, number of beef cows, percentage of income from beef cows, leverage, and percentage of acres owned. We use accrual income as the output variable. Table 1 reports the mean and standard deviation of the input and

output variables.

Table 1. Summary statistics

Variable	Unit	Mean	Std. Dev
Beef Accrual Livestock Income	$	323,017.1	474,388.2
Age of Operator	Years	58.2	10.8
Number of Beef Cows	No.	112.7	128.7
Percentage of Income from Beef Cows	%	33.3	23.8
Leverage	%	28.3	21.3
Percentage of Acres Owned	%	36.0	28.6
Feed	$	91,917.5	236,282.8
Labor	$	37,966.4	56750.5
Utilities	$	10,612.5	12,230.8
Capital	$	28,352.8	41,490.4
Veterinary	$	13,821.5	33,347.1
Miscellaneous	$	6,350.0	7,166.2

4. Results

Table 2 summarizes the estimated super technical efficiency measures. We assumed the technology exhibits variable returns to scale in this study. Technical efficiency ranges from 0.07 to 8.57, with mean of 0.87 and standard deviation of 0.87. The high mean indicates that the farms in the Kansas produce efficiently on average. The standard deviation is also high, which can be attributed to the efficient farms.

We see that the majority of farms are operating inefficiently, since only 24% of farms are with efficiency value greater than one. In addition, only 4% of farms are very inefficient (below 0.2). Figure 2 shows the cumulative distribution function of the efficiency values. The median is at the left side of the mean, meaning that the distribution of efficiency is left skewed.

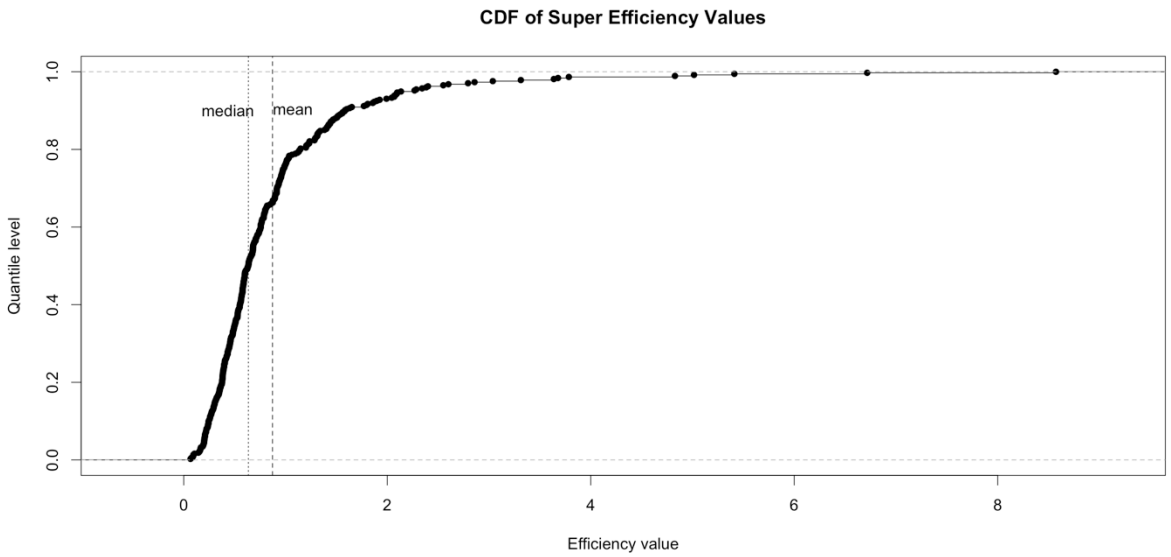

Figure 2. Cumulative density function (CDF) of super technical efficiency values

Table 2. Distribution of efficiency estimates

	Variable Returns to Scale	
Distribution	Number	Percentage
0 to 0.1	5	1.3
0.1 to 0.2	12	3.2
0.2 to 0.3	35	9.4
0.3 to 0.4	40	10.7
0.4 to 0.5	37	9.9
0.5 to 0.6	48	12.8
0.6 to 0.7	33	8.8
0.7 to 0.8	30	8.0
0.8 to 0.9	15	4.0
0.9 to 1	29	7.8
Greater than 1	90	24.1
Total	374	100

The estimation results of technical efficiency on input variables are shown in Table 3. We select the input variables following Featherstone et al (1997). The second column lists the OLS regression results. The labor, utility and capital have significant negative effects on the average of efficiency values. Amongst them, the utility has the most impact on the efficiency. Specifically, 1% decrease in the utility are expected to increase the efficiency by 0.3%. Note

that the inference can be applied to both efficient farms and inefficient farms. Now, we use the quantile regression on the median, rather than on the mean. We find that the regression results are quite similar to the OLS estimates. As we know, the OLS estimates reflect the impact of input variables on the mean of efficiency, while the median regression reflects the impact of input variables on the median, or the center, of efficiency. The minimal distinction between the OLS estimates and median regression estimates are due to that the mean and median of efficiency are close (mean is 0.87 and median is 0.64). Most importantly, both of them locates at the inefficient range. In other words, both the OLS and median regression reflect more causality relationship for inefficient farms. The consequence of this is that the causality for efficient farms are not recovered by the two regressions. If the causality for efficient farms is different to the causality for inefficient farms, the estimates would be biased by ignoring the heterogeneity.

Table 3. Effect of input factors on efficiency

	OLS	Median	Lower Quantile	Upper Quantile
Intercept	4.22*** (0.46)	4.00*** (0.33)	4.94*** (0.31)	4.03*** (0.77)
Feed	0.03 (0.04)	0.00 (0.03)	-0.01 (0.01)	0.11* (0.06)
Labor	-0.06*** (0.02)	-0.05*** (0.02)	-0.06*** (0.01)	-0.03 (0.05)
Utilities	-0.29*** (0.05)	-0.26*** (0.04)	-0.3*** (0.02)	-0.32*** (0.09)
Capital	-0.13*** (0.03)	-0.13 (0.03)	-0.13*** (0.01)	-0.16*** (0.05)
Veterinary	-0.01 (0.05)	-0.01*** (0.03)	-0.03** (0.01)	-0.06 (0.08)
Miscellaneous	-0.05 (0.06)	-0.04 (0.05)	-0.09** (0.04)	0.04 (0.11)

Notes: Single, double, and triple asterisk (*) denote coefficient is significant at the 10%, 5%, and 1% level, respectively.

[a] Coefficient estimates represent marginal effects. Standard errors are in parentheses.

To differentiate between the efficient and inefficient farms, we apply two more quantile

regressions, one with lower quantile level (38%) which locates at the central position of inefficient farms, and another one with upper quantile level (88%) which locates at the central position of efficient farms. As expected, the lower quantile regression results are close to the median or OLS estimates, while the upper quantile regression results are different, especially for the feed variable. The input in feed has no significant effect on efficiency for inefficient farms, while the has significant effects on efficiency for efficient farms. Unlike inefficient farms, an increase in feed is positively associated with an increase in efficiency. In other words, the efficiency can be improved if the efficient farms invest more on feed. It confirms with our hypothesis that efficient and inefficient farms are heterogeneous, in the way how the input variables affect the efficiency[1].

Table 4. Effect of Farm Characteristics on efficiency

	OLS	Median	Lower quantile	Upper quantile
Intercept	0.93*** (0.31)	0.60*** (0.17)	0.41*** (0.13)	1.78** (0.77)
Age of Operator	-0.01 (0.00)	-0.00 (0.00)	-0.00 (0.00)	-0.01 (0.01)
Number of Beef Cows	-0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	-0.00 (0.00)
Percentage of Income from Beef Cows	1.06*** (0.24)	0.77*** (0.13)	0.66*** (0.10)	1.83** (0.78)
Leverage	-0.36 (0.20)	-0.27** (0.11)	-0.18** (0.09)	-0.34 (0.63)
Percentage of Acres Owned	-0.01 (0.13)	-0.08 (0.09)	-0.03 (0.07)	0.07 (0.32)

Notes: Single, double, and triple asterisk (*) denote coefficient is significant at the 10%, 5%, and 1% level, respectively.

[a] Coefficient estimates represent marginal effects. Standard errors are in parentheses.

1 We run the Wald test on the hypothesis that the coefficients are equal between the lower and upper quantile regressions. The test results show that the coefficient for feed is significantly unequal at 95% significance level. The coefficients for other input variables are insignificant.

We run the OLS regression and quantile regressions on the farm characteristics. The regression results, as shown by Table 4, indicate that the percentage of income from beef cows have positive effects both on inefficient and efficient farms. Percentage of income from beef cows is an index indicating the degree of specialization. This might suggest that focusing efforts on beef cow is more important than diversification of enterprises that may lead themselves to be efficient. Though the upper and lower quantile regression return different results for the two groups of farms, the Wald test shows that the coefficient estimates are statistically equal.

5. Conclusion

This article contributes to the methodology of deriving different marginal effect on efficiency. Previous studies implicitly assume homogeneity across entire observation and derive one single marginal effects of covariates on efficiency. Since efficiency measures are censored at 1 in the standard DEA, it was impossible to divide efficient group and inefficient group and identify their own sources of efficiency. Super DEA allows the efficient units to have an efficiency score greater than 1 by dropping the constraint that bounds. It eventually leads us to divide whole sample into two groups, efficient farms and inefficient farms, and find their different strategies for enhancing efficiency by employing the quantile regression.

This analysis finds that an increase in feed affects efficiency in opposite ways for efficient farms and inefficient farms. Feed is positively correlated with efficiency while feed has no significant effect on entire group. This result convinces our hypothesis that efficient and inefficient farms are heterogeneous. In other words, if employing the standard DEA, it leads to have a biased coefficient for efficient farms.

Researchers may utilize these two stage methods in deriving efficiency measure and finding

marginal effects on efficiency in need of future research. If efficient and inefficient group may be heterogeneous, traditional way is likely to have a potential risk of having biased result. This study suggests that future research should utilize Super DEA and quantile regression, instead of employing DEA and Tobit.

References

Andersen, Per, and Niels Christian Petersen. 1993. "A Procedure for Ranking Efficient Units in Data Envelopment Analysis." *Management Science* 39 (10): 1261–64.

Andreu, Monica Lopez, and Orlen C. Grunewald. 2006. "Sources of Inefficiency in Kansas Farms." 2006 Annual Meeting, February 5-8, 2006, Orlando, Florida 35283. Southern Agricultural Economics Association.

Blundell, Richard, and Costas Meghir. 1987. "Bivariate Alternatives to the Tobit Model." *Journal of Econometrics* 34 (1): 179–200.

Chen, Y., 2005. "Measuring super-efficiency in DEA in the presence of infeasibility". *European Journal of Operational Research* 161: 545-551.

Chesher, Andrew. 1984. "Testing for Neglected Heterogeneity." *Econometrica* 52 (4): 865–72.

Coelli, Tim, Sanzidur Rahman, and Colin Thirtle. 2002. "Technical, Allocative, Cost and Scale Efficiencies in Bangladesh Rice Cultivation: A Non-Parametric Approach." *Journal of Agricultural Economics* 53 (3): 607–26.

Featherstone, Allen M., Michael R. Langemeier, and Mohammad Ismet. 1997. "A Nonparametric Analysis of Efficiency for a Sample of Kansas Beef Cow Farms." *Journal of Agricultural and Applied Economics* 29 (01): 175–84.

Helfand, Steven M., and Edward S. Levine. 2004. "Farm Size and the Determinants of Productive Efficiency in the Brazilian Center-West." *Agricultural Economics* 31 (2-3): 241–49.

Hoff, Ayoe. 2007. "Second Stage DEA: Comparison of Approaches for Modelling the DEA Score." *European Journal of Operational Research* 181 (1): 425–35.

Koenker,R.,and G. Bassett. 1978. Regression Quantiles. *Econometrica* 46(1):33–50.

McDonald, John. 2009. "Using Least Squares and Tobit in Second Stage DEA Efficiency Analyses." *European Journal of Operational Research* 197 (2): 792–98.

Pendell, D.L., Y. Kim, and K. Herbel. "Differences Between High-, Medium-, and LowProfit Cow-Calf Producers: An Analysis of 2010-2014 Kansas Farm Management Association

Cow-Calf Enterprise." Kansas State University, Department of Agricultural Economics (Publication: KSU-AgEcon-DP-YK-KH-2015.1). December 2015.

Reynolds, A., and J. S. Shonkwiler. 1991. "Testing and Correcting for Distributional Misspecifications in the Tobit Model: An Application of the Information Matrix Test." *Empirical Economics* 16 (3): 313–23.

Samoilenko, Sergey, and Kweku-Muata Osei-Bryson. 2008. "Increasing the Discriminatory Power of DEA in the Presence of the Sample Heterogeneity with Cluster Analysis and Decision Trees." *Expert Systems with Applications* 34 (2): 1568–81.

Wang, Qunwei, Zengyao Zhao, Peng Zhou, and Dequn Zhou. 2013. "Energy Efficiency and Production Technology Heterogeneity in China: A Meta-Frontier DEA Approach." *Economic Modelling* 35 (September): 283–89.

Xue, Mei and Harker T., Patrick. 2002. "Note: Ranking DMUs with Infeasible Super-Efficiency DEA Models". *Management Science* 48(5):705-710.