

Semiparametric stochastic frontier estimation using the Generalized Additive Models

ABSTRACT

This article builds a semiparametric stochastic frontier function using generalized additive models that accounts for random noise in the data. We estimated the parameters of the model by applying the generalized spline-smoothing approach to measure technical efficiency scores of Wisconsin dairy producers for the period 1993 to 1998. Results showed that the sample producers did not use resources efficiently, as the estimated mean technical efficiency score was found to be 0.778. Unlike precedent studies, we found no correlation between the estimated technical efficiency scores and four farm-specific characteristics, such as operation type, milk system, barn type, and milk frequency.

Keywords: Generalized additive models; spline-smoothing approach; semiparametric stochastic frontiers; technical efficiency.

JEL Classifications: C13, C14, Q12

1. INTRODUCTION

Literature shows the earliest empirical studies of efficiency measurement at the micro level dates back to the 1960s when Chennareddy [1] measure farmers' allocative efficiency scores defined as the ability of enterprises to use factors of production in optimal proportions, given input prices. Later, Lau and Yotopoulos [2] develop a dual profit function to measure both allocative and technical efficiency scores in which the latter reflects the enterprises' ability to obtain maximum output from a given set of inputs. However, these studies followed the prominent theoretical results of Farrell [3] whose research lays out the foundation of bounded functions or frontiers. Generally, there are three classes of frontier models, i.e., parametric, nonparametric, and semiparametric, none of which is superior to one another. Parametric frontier models have been criticized for the inherent assumptions on the distribution of the one-sided random error term indicating technical inefficiency [4,5] and the specification of functional forms [6]. Nonparametric frontiers are tedious in terms of model specification and estimation, but do not have the shortcomings of parametric frontier models [7]. Semiparametric frontier models, which are defined as a combination of both parametric and nonparametric models, have not been widely used in empirical studies [8].

The major contribution of this paper to the literature is two-fold. First, we built a semiparametric stochastic frontier production function using generalized additive models (GAMs) that were proposed by Hastie and Tibshirani [9]. Second, we applied the GAM theory to a sample of panel data collected from the Wisconsin dairy farms from 1993 to 1998 to estimate technical efficiency scores of dairy producers. This period is of particular importance from a policy perspective because Wisconsin dairy farms produced one-fifth of the total milk production in the United States during 1990s [10]. We estimated the parameters of the model using a nonparametric technique, known as generalized spline-smoothing approach, which was initially proposed by Wahba [11]. Furthermore, we utilized the corrected ordinary least squares (COLS) method of Richmond [12] to estimate technical efficiency scores of dairy farms in the sample data.

The rest of the paper is organized as follows. Section 2 briefly explains GAMs and the spline-smoothing approach. Section 3 presents the findings of empirical analysis and a discussion of the results. Section 4 concludes the paper and provides directions for further research.

2. METHODOLOGY

2.1. Generalized Additive Models (GAMs)

Generalized additive models (GAMs), proposed by Hastie and Tibshirani [9], are the nonparametric extension of generalized linear models (GLMs). In a GLM the relationship between the expected value of response and each of the predictors is linear and additive. A GAM is derived from a GLM by maintaining the additivity assumption and relaxing the linearity premise. A simple description of GAMs follows. Suppose a set of predetermined and/or random variable matrix of predictors $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)'$ measures the mean variation of a random response Y with n observations. We can write any multiple regression models as equation (1):

$$\begin{aligned} Y_i &= \alpha + x_{i1}\beta_1 + x_{i2}\beta_2 + \dots + x_{ik}\beta_k + \varepsilon_i \quad \text{or} \\ Y_i &= m(\mathbf{x}_i) + \varepsilon_i, \quad i = 1, \dots, n \end{aligned} \quad (1)$$

in which $m(\mathbf{x}_i) = \alpha + \mathbf{x}_i'\boldsymbol{\beta}$, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_k)'$, $E(\varepsilon_i) = 0$, and $\text{Var}(\varepsilon_i) = \sigma^2$. The pioneer assumption in equation (1) is that the relationship between the expected value of Y_i and each of k -covariate elements of $\mathbf{x}_i$ is linear and additive. To derive a GAM, we should use one of the two following methods that relax the linearity assumption and maintain the additivity supposition. The first approach is to use surface smoothers such as kernel functions, which are nonparametric estimates of the regression model. Kernel functions preclude practitioners from using more than two predictors in the model. This major drawback of kernel functions is known as the curse of dimensionality problem [13]. The second approach is to specify GAMs as equation (2):

$$\begin{aligned} E[Y_i | \mathbf{x}_i = (x_{i1}, \dots, x_{ik})] &= G \left[\alpha + \sum_{j=1}^k f_j(x_{ij}) \right] = G[f(\mathbf{x}_i)] \\ \text{or} \end{aligned} \quad (2)$$

$$f(\mathbf{x}_i) = \alpha + \sum_{j=1}^k f_j(x_{ij}),$$

in which the distribution of $\mathbf{Y}_i$ follows an exponential family similar to GLMs, α is a constant, the f_j s are arbitrary univariate smooth functions; one for each predictor, and $G(\cdot)$ is a fixed link function. To avoid having free constant in each of the functions f_j and for the purpose of identification, Hastie and Tibshirani [9] assume that $E[f_j(x_{ij})] = 0$ that implies $E[f(\mathbf{x}_i)] = \alpha$ in the range of $1 \leq i \leq n$ and $1 \leq j \leq k$. Thus, in GAMs the conditional mean response depends upon a sum of individual univariate functions each contains one predictor from the covariate matrix. When the linearity assumption in GAMs is relaxed, then the effects of the independent variables in equation (2) may be nonlinear because the univariate smooth functions f_j are now arbitrary [14]. In this paper, we used the generalized spline-smoothing method [11] to estimate the parameters of equation (2), which is explained in the next section.

2.2. Spline Smoothing

The most usual method of estimating m from sample data in equation (1) is to minimize the residual sum of squares over all observations in relation to the given functional form. However, there is no guarantee that a linear relationship between the dependent and independent variables exists. One way to find the source of such failure is to use a Taylor-expansion series as

$$m(x) = m(x_0) + m'(x_0)(x - x_0) + o(|x - x_0|^2) \quad (3)$$

in which m , an unknown function, is at least twice differentiable and there is a point x close to some fixed point x_0 for which m follows a linear model with an intercept $m(x_0) - m'(x_0)x_0$ and slope $m'(x_0)$. Using the Taylor-expansion series may generate two extreme scenarios [11]:

- m is assumed to be linear implying that the slope remains invariant and the residual term $o(|x - x_0|^2)$ is small, which is an unrealistic assumption. This scenario provides a useful summary of the sample data and presents a comprehensive description of its features despite the fact that it uses too little information available in the data.
- m is assumed to have variant slopes implying that at each point x different slopes exist that connect every two responses by lines associated with their own individual slopes. Although this scenario uses too much information available in the sample data it does not provide a useful summary of the data, and as a result, it fails to show a satisfied description of basic trends in the sample observations.

To circumvent this problem, consider m'' that shows the rate of change in the slope of a function m . Since m'' varies from one point to another, equation (4) represents the summation of the entire changes in slopes of the fitted regression

$$\Phi(m) = \int_{x_1}^x m''(x)^2 dx \quad (4)$$

that can be minimized over all functions provided that they will be double-differentiated. Such minimization yields equation (5)

$$RSS(m) + \tau \Phi(m), \quad \tau \geq 0, \quad (5)$$

in which τ , known as the smoothing parameter (or the span degree), can vary between zero and infinity. If it approaches zero, a regression with flexible slopes is yielded, whereas if it approaches infinity a linear regression model is derived. Eubank [15] states that if n in equation (4) is greater than or equal two, then equation (6) provides a unique estimator, $\mathcal{G}_\tau$, known as the spline-smoothing estimator that minimizes equation (5). The spline-smoothing estimators are linear in the sense that one can find constants such as $g_i(x)$ for each estimation point x such that

$$\mathcal{G}_\tau(x) = \sum_{i=1}^n g_i(x) y_i \quad i = 1, 2, \dots, n \quad (6)$$

Although using spline-smoothing estimators resolves the problem of fitting regression with variant slopes, it has one major drawback caused from a lack of theory and appropriate algorithm. These estimators are data specific and therefore sensitive to the choice of span degree. To fix the problem, Stone [16] proposes the cross validation (CV) method, which is briefly explained in the next section.

2.3. Model Specification

Equation (7) specifies a multiple variable production function

$$Y_{it} = f(X_{it}) + \varepsilon_{it}, \quad i = 1, 2, \dots, n_t, \text{ and } t = 1, 2, \dots, T. \quad \text{or} \quad (7)$$

$$f(X_{it}) = \alpha + \sum_{j=1}^k f_{ji}(X_{jit}) + \varepsilon_{it}$$

in which Y_{it} is output, f represents an unknown functional form that must be estimated using a nonparametric method, such as the spline smoothing technique, X_{it} is a multidimensional series of explanatory variables with real values, i.e., $X_{it} \in \mathfrak{R}^k$, and ε_{it} represents random error terms that are *i.i.d.* with zero mean and common distribution $\mathfrak{F}$. Moreover, it is assumed that f , X_{it} , and ε_{it} are independent and the identification condition described in the methodology section still holds. For the simplicity, we drop the subscripts i and t and specify the model for

only two independent variables. However, the extension of the model to include more than two predictors is straightforward. Equation (8) specifies a simple GAM with two predictors

$$E[Y | x_1, x_2] = f(x_1, x_2) = \alpha + f_1(x_1) + f_2(x_2) \quad (8)$$

Furthermore, we may notice that

$$f_1(x_1) = \int f(x_1, x_2) g(x_2) dx_2 = E[f(x_1, X_2)] \quad (9)$$

Chen et al. [17] show that if $E[f_2(x_2)]$ is assumed to be equal zero, then

$\int f_2(x_2) g(x_2) dx_2 = 0$. Therefore, $f_1(x_1)$ is estimated by $\hat{f}_1(x_1) = T^{-1} \sum_{t=1}^T \hat{f}(x_1, x_{t2})$ where $\hat{f}(x_1, x_{t2})$ is the nonparametric estimator of $f(x_1, x_2)$. Friedman and Stuetzle [18] show that if f is assumed to be a twice-differentiable smooth function, then using the backfitting algorithm will generate reliable semiparametric estimators. A brief explanation of the backfitting algorithm is as follows. Consider equation (8), this algorithm initially estimates $\hat{f}_1(x_1)$ and then, while fixing the fitted function $\hat{f}_1(x_1)$, projects the mean dependent variable on x_2 by smoothing the residual $Y - \hat{\alpha} - \hat{f}_1(x_1)$, which leads to the estimation $\hat{f}_2(x_2)$. The next step is to improve the estimation $\hat{f}_1(x_1)$ by smoothing the residual $Y - \hat{\alpha} - \hat{f}_2(x_2)$ on x_1 , which, in turn, enhances the estimators used to smooth the residual $Y - \hat{\alpha} - \hat{f}_1(x_1)$ on x_2 in the second step. This procedure continues until reliable and efficient estimators are achieved. The spline smoothing technique is used to obtain the initial estimate of $f_1(x_1)$ in the iterative smoothing procedure [11].

As mentioned in the previous section, the use of spline-smoothing method is associated with the inconsistency and sensitivity of the estimates to the choice of span degree that can be fixed by the CV method. Given equation (9), Kneip and Simar [13] show that the CV method minimizes $\sum_{-(i,t)} [Y_{it} - \hat{\alpha} - \hat{f}_{-(it)}(x_{it})]^2$ by estimating $\hat{f}_{-(it)}$ using a two-step procedure, known as

“the leaving out the observation (y_{it}, x_{it}) .” In the first step, for a fixed firm i , $i = 1, 2, \dots, n_t$ and in every sequence of time period t , $t = 1, 2, \dots, T$ one pair of the sample data, i.e., the i -th and t -th observations, is put aside and the mean response function f , defined in equation (7), is re-estimated based on the $n - 1$ remaining observations. In the second step, the algorithm is repeated and continued to estimate f until convergence.

Finally, as it is mentioned in Section 1, we measure technical efficiency scores of the Wisconsin dairy producers by using the COLS method. As Greene [19] states the COLS estimators are consistent and unbiased similar to the least squares estimators. Besides, by using the COLS method there is no need to make *a priori* assumption on the distribution of the random error terms before estimating the parameters of the model [20]. Notwithstanding its shortcomings, Greene [19] asserts that the COLS method is a popular approach in the analysis of panel data in estimating technical efficiency scores [21-23].

3. EMPIRICAL ANALYSIS

3.1. Data Description

This study uses an unbalanced panel data from the Wisconsin dairy farms database. The Centre for Dairy Research in the Department of Agricultural, Resource, and Managerial Economics at the University of Wisconsin-Madison collects information from individual dairy farms in each year. In total, we use 510 sample observations collected from 223 dairy farms during the period 1993 to 1998. To estimate technical efficiency scores of Wisconsin dairy producers, six groups of farm-specific independent variables, including land, labor, and farm accrual expenses on grain, livestock, machinery, and miscellaneous as well as two socio-economic variables, such as age, and education in the form of years of schooling are used. The dependent variable is defined as the total fluid milk production of 3.7 per cent fat content per year measured in pounds. We measure land in hectare as annual total (owned and rented) tillable area. Lack of data prevents us from separating pasture and planted areas. We define labor as annual total equivalent worker unit (hereafter, ewu), which contains both hired and family labor as well as the working hours of the main operator where it is applicable. We add all the payments made to rent machinery and farm equipment, purchase oil and parts, and depreciation together and consider them as the total farm-specific machinery accrual expenses. Livestock expenses are computed by adding the amount of money spent on cattle lease, breeding, livestock replacement, veterinary, and milk marketing. We calculate the grain production accrual expenditures by the amount of money spent on purchasing grain and roughage, fertilizer, seed, and spray to grow on-farm grain and hay. We also include accrual expenses related to non-farm production processes as miscellaneous expenditures. These expenses include electricity bills, insurance invoices, interest payments, and building depreciation. We deflate the last four groups of variables using the appropriate Producer Price Index (*PPI*), and convert all the variables (except socioeconomic variables) to logarithmic forms to mitigate possible heteroskedasticity in the model. Finally, we divide all the variables by the number of milk cows to avoid farm size effects.

Table 1 presents a summary of the statistical descriptions of the farm-specific variables. The sample mean of milk production per cows in Wisconsin is 18,784 pounds with a standard deviation of 3,244 during the period of study. The average tillable area in the sample data is 4.39 hectares and the mean labor used is 0.035 ewu per each milking cow. The amount of money spent on purchasing grain and roughage (\$US 7.79 per milking cows) is the highest accrual expenses in the sample data. Table 1 also shows that each dairy farm in the sample data, on average, has 93 milking cows during the period of the study. Finally, the average age of the primal farm operator is 43 years and the average years of schooling is approximately 14 years.

[Insert Table 1 here]

3.2. Estimation Results

We, first, estimate the mean response function f defined in equation (2) by using R (version 2.10.1), and then measure the individual technical efficiency scores using the COLS approach. In particular, we consider a nonparametric relationship between the dependent variable and the six groups of farm-specific independent variables to avoid any misspecification in the model. We also assume a parametric relationship between the dependent variable and the socioeconomic variables on the ground that dairy farm operators accumulate more experience each year as they get older and participate in formal and informal school training.

Table 2 presents estimated technical efficiency scores of the sample data that is classified by the group mean performance (hereafter, efficiency class interval). For each of the efficiency class intervals and the overall average estimated technical efficiency scores, we obtain lower and upper bounds of confidence limit by using a bootstrap technique, so-called the bias-corrected and accelerated percentile method with 1000 replications [24]. Results show that the average technical efficiency scores of the sample data are found to be 77.8 per cent with a confidence limit of (0.769, 0.787). This means that the same volume of milk production could have been theoretically achieved with approximately 22 per cent fewer input used if all the sample dairy-farm operators had operated at 100 per cent efficiency. The magnitude of the estimated mean technical efficiency scores is higher than that of Haghiri et al. [23] reported (67.5 per cent) for the sample data collected from the New York dairy farms. This finding is important to policymakers because the same U.S. policy implemented in both states during the period of study. **Table 2** also shows that the majority (84.8 per cent) of the sample dairy farms falls in the category of 66-85 per cent technical efficiency.

[Insert table 2 here]

To investigate variations among the estimated technical efficiency scores in the sample data, we conduct a two-factor ANOVA test with no replication. The computed F -test value (12.02) rejects the null hypothesis of no variation among the estimated technical efficiency scores during the period of the study at the 0.05 level of significance. This implies that the performance of the sample dairy producers varies significantly from year-to-year. Furthermore, we use the chi-square statistic test to examine the null hypothesis of no evidence of a relationship between the estimated technical efficiency scores and time trend (i.e., technological changes) in the sample data. The chi-square test allows us to find whether the estimated technical efficiency scores and time trend are independent of each other [25]. Since the calculated value of the chi-square (110.93) with 20 degrees of freedom exceeds the statistic critical value (37.56) at the 0.05 level of significance it is reasonable to conclude that technological changes are occurred in the sample data between 1993 and 1998.

To examine the distribution type of the estimated technical efficiency scores we conduct a univariate Shapiro-Wilk [26] test of normality for the sample data. The calculated p -value of the statistic (0.441) indicates that the null hypothesis of normal distribution of the estimated technical efficiency scores at the 0.05 level of significance is not rejected. In addition, we specify

a regression model to investigate factors that cause inefficiency in the sample data. The model uses the estimated technical efficiency scores as dependent variable and four farm-specific characteristics, including operation type, milk system, barn type, and milk frequency as independent variables. Results show that none of the four predictors is able to explain variations in the dependent variable with 95 per cent confidence. The estimated sample correlation coefficients for operation type (0.0326), barn type (0.0343), and milk frequency (0.0698) are all positive while it is negative for milk system (-0.0041). Finally, we find no correlation between farm size and the estimated technical efficiency scores. The estimated sample correlation coefficient (0.065) is positive, but it is not significant at the 0.05 level.

4. CONCLUSION

Dairy industry in U.S. operates under a combination of various policies that are placed by both the federal and state governments. The implementation of each of these policies will affect the performance of the U.S. dairy farmers. Performance of a dairy farmer, ultimately, induces to how much output a farm can obtain from a set of inputs, given the underlying technology (i.e., technical efficiency). If in a dairy farm, maximum level of output from a given set of inputs is obtained, the farm is called a frontier enterprise. Several methods have been introduced in the literature to estimate frontiers, yet none of them provides accurate results as the estimation of technical efficiency scores is data specific. As a result, policy decisions made based on the results of such methods will lead to misallocation of the scarce resources, which otherwise would have been allotted to proper sector. Notwithstanding, this does not preclude us from proposing new methodology to enhance the estimation of frontiers in which the inherent problem of precedent methods is rendered. We set up a stochastic semiparametric frontier model to estimate technical efficiency scores of a sample data collected from Wisconsin dairy producers for the period 1993 to 1998. The model is specified based on the generalized additive theory and estimated using the spline smoothing approach. Results show that the sample of Wisconsin dairy producers is, on average, 77.8 per cent technically efficient and the distribution of the estimated technical efficiency scores is normal. We suggest the following areas for further research. First, lack of the information related to the input and output prices for each of the dairy farms in the sample prevent us from measuring allocative and overall efficiency scores. Second, in this study labor is used in terms of physical quantity and not in terms of dollar values. This may raise the question that how sensitive the findings of the research would be if labor is measured in terms of dollar values.

REFERENCES

- [1] Chennareddy, V. (1967), "Production efficiency in south Indian agriculture," *Journal of Farm Economics*, Vol. 49, No.4, pp.816-820.
- [2] Lau, L. J. and P. A. Yotopoulos (1971), "A test for relative efficiency and applications to Indian agriculture," *American Economic Review*, Vol.61, No.1, pp.94-109.
- [3] Farrell, M. J. (1957), "The measurement of productive efficiency," *Journal of the Royal Statistics Society Series A*, Vol.120, No.3, pp.253-290.
- [4] Haghiri, M. (2013), "Applied nonparametric regression analysis: The choice of generalized additive models," *Review of Economics and Finance*, Vol. 3, No. 1, pp.25-34.

- [5] Kumbhakar, S. C. and E. G. Tsionas (2002), "Nonparametric stochastic frontier models," 2nd North American Productivity and Efficiency Conference, Union College, Schenectady, Albany, New York, June 20-22.
- [6] Anderson, D. P., Chaisantikulawat, T., Tan Khee, G. A., Kebbeh, M., Lin, N. and C. R. Shumway (1996), "Choice of functional form for agricultural production analysis," *Review of Agricultural Economics*, Vol.18, No.2, pp.223-231.
- [7] Haghiri, M. and A. Simchi (2010), "A nonparametric extension of generalized quadratic Box-Cox models in measuring technical efficiency," *Applied Economics*, Vol.42, No.1, pp.1-8.
- [8] Huang, L. and L. Fu (2002), "Semiparametric smooth coefficient models," *Journal of Business and Economic Statistics*, Vol.20, No.3, pp.412-422.
- [9] Hastie, T. J. and R. J. Tibshirani (1990), *Generalized Additive Models*, London: Chapman and Hall.
- [10] U.S. Department of Agriculture (USDA), Annual Report, various issues.
- [11] Wahba, G. (1990), *Spline Models for Observational Data*, 1st Edition, Philadelphia: CBMS-NSF Regional Conference Series SIAM (Society for Industrial and Applied Mathematics).
- [12] Richmond, J. (1974), "Estimating the efficiency of production," *International Economic Review*, Vol.15, No.2, pp.515-521.
- [13] Kneip, A. and L. Simar (1996), "A general framework for frontier estimation with panel data," *Journal of Productivity Analysis*, Vol.7, No.2/3, pp.187- 212.
- [14] Schimek, M. G. and B. A. Turlach (2000), Additive and Generalized Additive Models, in M. G. Schimek (Ed.) *Smoothing and Regression: Approaches, Computation, and Application*, New York: John-Wiley and Sons, pp.277-327.
- [15] Eubank, R. L. (2000), Spline regression, in M. G. Schimek (Ed.) *Smoothing and Regression: Approaches, Computation, and Application*, New Jersey: John Wiley and Sons, pp.1-18.
- [16] Stone, M. (1974), "Cross-validatory choice and assessment of statistical predictions," *Journal of the Royal Statistical Society Series B*, Vol. 36, No.2, pp.111-147.
- [17] Chen, R., Hardle, W., Linton, O. B. and E. Sevarance-Lossin (1996), "Nonparametric estimation of additive separable regression models," *Statistical Theory and Computational Aspects of Smoothing*, Physica Verlag, pp.247-265.
- [18] Friedman, J. H. and W. Stuetzle (1981), "Projection pursuit regression," *Journal of the American Statistical Association*, Vol.76, No.376, pp.817-823.
- [19] Greene, W. H. (2008), The econometric approach to efficiency analysis, in H. O. Fried, C. A. Knox Lovell and S. S. Schmidt (Eds.) *The Measurement of Productive Efficiency and Productivity Change*, London: Oxford University Press, pp.1-158.
- [20] Haghiri, M. (2003), "Stochastic nonparametric frontier analysis in measuring technical efficiency: a case study of the North American dairy industry," Unpublished, Ph.D. Dissertation, Department of Bioresource Policy, Business and Economics, University of Saskatchewan, Saskatoon, Saskatchewan, Canada.
- [21] Cornwell, C., Schmidt, P. and R. Sickles (1990), "Production frontiers with cross sectional and time series variation in efficiency levels," *Journal of Econometrics*, Vol. 46, No.1/2, pp. 185-200.
- [22] Evans D, A., Tandon, C. M. and J. Lauer (2000), "The comparative efficiency of national health systems in producing health: An analysis of 191 countries," GPE Discussion Paper, No. 29, EIP/GPE/EQC, World Health Organization, Geneva.

- [23] Haghiri, M., Nolan, J. F. and K. C. Tran (2004), “Assessing the impact of economic liberalization across countries: a comparison of dairy industry efficiency in Canada and the USA, *Applied Economics*, Vol.36, No.11, pp.1233-1243.
- [24] Briggs, A. H., Mooney, C. Z. and D. E. Wonderling (1999), “Constructing confidence intervals for cost-effectiveness ratios: An evaluation of parametric and nonparametric techniques using Monte Carlo simulation,” *Statistics in Medicine*, Vol. 18, No.23, pp.3245-3262.
- [25] Levine, D. M., Berenson, M. L., Krehbiel, T. C. and D. F. Stephen (2011), *Statistics for Manager using MS Excel*, 6th ed., New Jersey: Prentice Hall.
- [26] Shapiro, S. S. and M. B. Wilk (1965), “An analysis of variance test for normality (complete samples),” *Biometrika*, Vol. 52, No.3/4, pp.591-611.

Table 1. Statistical Description of the Variables (per Cows)

Variable	Mean	S. D.	Min.	Max.
Fluid milk production	18,784.1	3,244.0	6,812.7	31,472.5
Land	4.39	1.99	0.005	14.95
Labor	0.035	0.010	0.010	0.077
Machinery accrual expenses	3.87	2.24	0.006	14.79
Livestock accrual expenses	3.26	2.12	0.012	15.22
Grain accrual expenses	7.79	4.47	0.086	80.96
Miscellaneous accrual expenses	4.85	2.32	0.227	4.51
Age	43.0	9.0	23.0	71.0
Education	14.0	2.0	8.0	24.0
Cows	93.0	96.0	22.0	754.0

Source: Sample data. Output is annual total fluid milk production of 3.7 per cent fat content (pound); land is annual total tillable area (Hectare); labor is annual total equivalent worker unit; and real accrual expenses of machinery, livestock, grain, and miscellaneous are in U.S. dollars. Education is based on years of schooling.

Table 2. Technical Efficiency of Wisconsin Dairy Producers

Efficiency Class Interval	No. of Farms	Percent	Mean
≤ 0.65	7	3.14	0.627 (0.580 0.641)
0.66 - 0.75	81	36.32	0.719 (0.712 0.724)
0.76 – 0.85	108	48.43	0.802 (0.797 0.807)
0.86 – 0.95	24	10.76	0.891 (0.881 0.902)
> 0.95	3	1.35	0.975 (0.955 1.000)
Total	223	100.00	0.778 (0.769 0.787)

Source: Sample observations.

The numbers in parentheses are the confidence limits obtained by bootstrapping.