British Journal of Economics, Management & Trade
3(4): 405-418, 2013

SCIENCEDOMAIN ***international***
www.sciencedomain.org

SCIENCEDOMAIN

Semiparametric Stochastic Frontier Estimation Using Generalized Additive Models

Morteza Haghiri[1*]

[1]Memorial University – Grenfell Campus 20 University Drive, Corner Brook, Newfoundland and Labrador, A2H 5G4, Canada.

Author's contribution

The only author performed the whole research work. Author MH wrote the first draft of the paper. Author MH read and approved the final manuscript.

Research Article

Received 5th June 2013
Accepted 20th July 2013
Published 30th July 2013

ABSTRACT

This article specified a semiparametric stochastic frontier function using generalized additive models that accounts for random noise in the sample data. We estimated the parameters of the model by applying the generalized spline-smoothing approach to measure technical efficiency scores of Wisconsin dairy producers between 1993 and 1998. Results showed that the sample dairy 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.

1. INTRODUCTION

Literature shows the earliest empirical studies of efficiency measurement at the micro level dates back to the 1960s when [1] measured farmers' allocative efficiency which was defined as the firm's ability in using factors of production in optimal proportions, given input prices. Later, [2] developed a dual profit function to measure both allocative and technical efficiency scores in which the latter reflects a firm's ability to obtain the maximum amount of output

**Corresponding author: Email: mhaghiri@mun.ca;*

from a given set of inputs. These studies followed the prominent theoretical results of [3] whose research laid out the foundation of bounded functions (frontiers) in the literature. Frontiers are either deterministic or stochastic depending upon how output is presumably bounded from above, respectively, by a deterministic or stochastic production function. In the former models, any deviation from the frontier is due to technical inefficiency, whereas in the latter functions, deviations from the frontier can be imputed to both the statistical noise and technical inefficiency [4,5,6]. From theoretical perspective, 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 criticizing for the inherent assumptions on the distribution of the one-sided random error terms indicating technical inefficiency [7,8] and the specification of functional forms [9]. Nonparametric frontiers models are tedious in terms of model-specification and estimation, but these types of models do not have the aforementioned problems of the parametric frontier models [10]. Semiparametric frontier models, which are defined as a combination of both parametric and nonparametric models, have not been widely used in empirical studies [11,12,13]. Using the semiparametric stochastic frontier models is important because it would allow practitioners to have flexibility in specifying predictors of the model, which would not have been the case if parametric or nonparametric frontier functions were modeled solely. Although from theoretical perspective our model is built on precedent studies its major contribution to the literature is three-fold:

First, we built a semiparametric stochastic frontier production function within the framework of generalized additive models (GAMs) and estimated the parameters of the model using a nonparametric technique, known as generalized spline-smoothing approach [14].Despite of conducting several studies in measuring technical efficiency of dairy farms through both parametric and nonparametric frontier models, literature does not show any studies in this context in which a stochastic semiparametric frontier model is used. Therefore, this paper is the first that uses spline-smoothing method to estimate technical efficiency scores of a large set of dairy producers in the State of Wisconsin.

Second, we applied the GAM theory to a sample of unbalanced panel data collected from the Wisconsin dairy farms from 1993 to 1998 to estimate technical efficiency scores of dairy producers. This period of time was 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. We utilized the corrected ordinary least squares (COLS) method to estimate technical efficiency scores of dairy farms in the sample data [15].

Third, we wanted to know whether farm specific characteristics, such as operation type, milk system, barn type, and milk frequency would have any impact on the efficiency of dairy producers in the region. The knowledge of relationship between farm characteristics and efficiency scores helps policymakers understand better the industry at the time of making policy decisions. This is an important issue for the U.S. dairy industry because, unlike the Canadian counterpart, the U.S. industry is not being operated under the supply management system where the total industry output is controlled by the regulatory institutions. For the time being, the U.S. dairy industry (except California) is being operated under a combination of various dairy policies, including the dairy price support program, the pooled price discrimination program, the import barriers policy, the export subsidy program, and the tariff-rate quota policy, which has recently become the dominant U.S. dairy policy [5]. As a result, trading dairy products between Canada and the U.S. have been exposed to various sources of trade disputes [16]. It has been expected that different policies would affect the efficiency of dairy producers in the U.S. Such differences in trade policies have motivated us to conduct this research that aims to measure dairy firms' productive efficiency in the State of Wisconsin.

The rest of the paper is organized as follows. Section 2 briefly explains GAMs and the spline-smoothing approach. Section 3 reviews the recent studies of measuring technical efficiency in the U.S. dairy industry. Section 4 presents the findings of empirical analysis and a discussion of the results. Section 5 concludes the paper and provides directions for further research.

2. METHODOLOGY

2.1 Generalized Additive Models (GAMs)

Generalized additive models (GAMs), proposed by [17], are the nonparametric extension of generalized linear models (GLMs), introduced by [18]. 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 $\mathbf{X} = (\mathbf{x}_1,..,\mathbf{x}_n)'$ measures the mean variation of a random response Y with n observations. We can write any multiple regression models as equation (1):

$$\mathbf{Y}_i = \alpha + x_{i1}\boldsymbol{\beta}_1 + x_{i2}\boldsymbol{\beta}_2 + ... + x_{ik}\boldsymbol{\beta}_k + \varepsilon_i,$$

or

$$\mathbf{Y}_i = m(\mathbf{x}_i) + \varepsilon_i, \qquad i = 1,\ldots,n. \tag{1}$$

in which $m(\mathbf{x}_i) = \alpha + \mathbf{x}_i'\boldsymbol{\beta}$ and $\boldsymbol{\beta} = (\boldsymbol{\beta}_1,...,\boldsymbol{\beta}_k)$. In equation (1), $E(\varepsilon) \neq 0$ because as it is mentioned in the previous section, ε is defined as $\varepsilon = v - u$, where v shows the ordinary noise terms which can be either positive or negative and u represents technical inefficiency, which is non-negative. In other words, $E(\varepsilon) = E(v - u) = E(v) - E(u) = -E(u) \neq 0$. Thus, to make equation (1) more applicable, at first, we should transform the model as the following.

$$\begin{aligned} y &= m(x) + v - u, \\ &= m(x) + v - u + E(u) - E(u), \\ &= m(x) - E(u) + v - u + E(u). \end{aligned}$$

Let define $\mu = v - u + E(u)$, then $E(\mu) = 0$. Therefore, the model becomes $g(x) = m(x) - E(u)$. It is noteworthy to mention that the spline smoothing method can be used to estimate $m(x)$ and if u follows some known distribution we are able to estimate $E(u)$. The pioneer assumption in equation (1) is that the relationship between the expected value of $\boldsymbol{Y_i}$ and each of k-covariate elements of $\boldsymbol{x_i}$ is linear and additive. To derive a GAM, we should use one of the two following methods that relaxes 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. Unless the sample size is sufficiently large, 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 [19]. The second approach is to specify GAMs as equation (2):

$$E\left[\mathbf{Y}_i \mid \mathbf{x}_i = (x_{i1},...,x_{ik})\right] = G\left[\alpha + \sum_{j=1}^{k} f_j(x_{ij})\right] = G\left[f(\mathbf{x}_i)\right], \quad \text{or}$$

$$f(\mathbf{x}_i) = \alpha + \sum_{j=1}^{k} f_j(x_{ij}). \tag{2}$$

in which the distribution of $\boldsymbol{Y}_i$ 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* (.) is a fixed link function. To avoid having free constant in each of the functions f_j and for the purpose of identification, [17] assumed $E\left[f_j(x_{ij})\right] = 0$ which implies $E\left[f(\mathbf{x}_i)\right] = \alpha$ in the range of $1 \le i \le n$ and $1 \le j \le k$. Thus, in GAMs the conditional mean response depends upon a summation of individual univariate functions which each contains one predictor from the covariate matrix. When the linearity assumption in GAMs is relaxed, then the effects of independent variables in equation (2) may be nonlinear because the univariate smooth functions f_j are now arbitrary [20]. There are several methods of spline smoothing techniques, including multivariate adaptive regression spline (MARS), conic multivariate adaptive regression spline (CMARS), and robust conic multivariate adaptive regression spline (RCMARS). For instance, literature shows that MARS can be an alternative approach to GLMs and generalized partial linear models (GPLMs) in the presence of interaction relationships amongst predictors [21,12,22]. In this paper, we used the generalized spline-smoothing method, which was initially proposed by [14], to estimate the parameters of equation (2), which is briefly explained in the next section. We invite interested readers to find more about the aforementioned spline smoothing methods in [23,24,25,26].

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

$$m(x) = m(x_0) + m'(x_0)(x - x_0) + o\left(\left| x - x_0 \right|^2\right), \tag{3}$$

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 [14]:

- *m* is assumed to be linear implying that the slope remains invariant and the residual term $o\left| x - x_0 \right|^2$ 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, i.e., m. Since m'' varies from one point to another, equation (4) represents the summation of the entire changes in slopes of the fitted regression

$$\Phi(m) = \int_{x_1}^{x_n} m''(x)^2 \, \mathrm{d}x \,, \tag{4}$$

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

$$RSS(m) + \tau\,\Phi(m), \quad \tau \geq 0 . \tag{5}$$

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. [27] stated that if *n* in equation (4) is greater than or equal two, then equation (6) provides a unique estimator, ϑ_τ, 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

$$\vartheta_\tau(x) = \sum_{i=1}^{n} g_i(x)\, y_i \,, \qquad i = 1,2,\dots,n . \tag{6}$$

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 circumvent the problem, [28] proposed 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

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

or

$$f(X_{it}) = \alpha + \sum_{j=1}^{k} f_{ji}(X_{jit}) + \varepsilon_{it}. \tag{7}$$

which Y_{it} represents output, f is an unknown functional form that must be estimated using a nonparametric technique, such as the spline smoothing technique, X_{it} is a multidimensional series of explanatory variables with real values, i.e., $X_{it} \in \Re^k$, and ε_{it} represents random error terms that are *i.i.d.* with zero mean and common distribution $\Im$. Moreover, it is assumed that $f, X_{it}, \text{and}\, \varepsilon_{it}$ are independent and the identification condition described in the methodology section still holds. For the simplicity, we drop the subscripts $i \,\text{and}\, t$ and specify the model for only two independent variables. However, the extension of the model to include more than two independent variables is straightforward. Consider equation (8) that specifies a simple GAM with two predictors

$$E\left[Y \mid x_1, x_2\right] = f\ (x_1, x_2) = \alpha + f_1(x_1) + f_2(x_2), \tag{8}$$

Furthermore, we may notice that

$$f_1(x_1) = \int f\ (x_1, x_2)\, g\ (x)\, \mathrm{d}x_2 = E\ [f\ (x_1, X_2)], \tag{9}$$

Chen et al. [29] showed that if we assume $E\ [f_2\ (x_2)] = 0,$ then $\int f_2\left(x_2\right) g\left(x_2\right) \mathrm{d}x_2 = 0$. Therefore, $f_1\left(x_1\right)$ 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 $f\left(x_1, x_2\right)$. Moreover, [30] expressed that if f was a twice-differentiable smooth function, then by using the backfitting algorithm we would generate reliable semiparametric estimators. A brief explanation of the backfitting algorithm is as follows [19]. Consider equation (8), the backfitting algorithm initially estimates $\hat{f}_1\left(x_1\right)$ and then, while fixing the fitted function $\hat{f}_1\left(x_1\right)$, 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 $\hat{f}_2\left(x_2\right)$. 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 [14,30].

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), [19, p.192] showed that the CV method minimizes $\sum_{-(i,t)} \left[Y_{it} - \hat{\alpha} - \hat{f}_{-(it)}(x_{it})\right]^2$ by estimating $\hat{f}_{-(it)}$ using a two-step procedure, known as “the leaving out the observation $\left(y_{it}, x_{it}\right)$.” In the first step, for a fixed firm *i*, $i = 1,2,...,n_t$ and in every sequence of time period $t, t = 1,2,...,T$ 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 was mentioned in ‘Introduction’ Section, we measured technical efficiency scores of the Wisconsin dairy producers using the COLS method whose estimators are consistent and unbiased similar to the least squares estimators [31]. In addition, 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 [5]. Notwithstanding its shortcomings, the COLS method is a popular approach in the analysis of panel data in estimating technical efficiency scores [32,33,16,10,19,34].

3. RECENT STUDIES of EFFICIENCY IN THE U.S. DAIRY INDUSTRY

There are relatively a number of articles in the literature that attempted to measure the efficiency of dairy farms in the United States using both the parametric and nonparametric frontier functions. In this section, we highlighted the results of some of these studies in chronological order. For example, [35] studied technical, allocative, and scale efficiency of owner-operators of dairy farms in Utah using a stochastic production frontier function. The authors used a cross-sectional data contacting 116 families from a population of 510 in the State of Utah. The sample data were separated by size (small, medium, and large) based on dollar sales during 1985. Kumbhakar et al. [35] found that large farms (having more than 100 milking cows) were technically more efficient than small farms (having less than 50 milking cows). In another study, [36] extended a stochastic efficiency decomposition model to analyze technical, allocative, and economic efficiency. The researchers used a cross-section data collected from 511 New England dairy farms to estimate a Cobb-Douglas stochastic production frontier. Results showed that the mean overall efficiency was 70.2 per cent, and that, on average, there was little difference between technical (80.3 per cent) and allocative (84.6 per cent) efficiency. Ahmad and Bravo-Ureta [37] used an unbalanced panel data to decompose dairy output growth into technological progress, technical efficiency, and input-growth for a sample of 1,072 observations collected from 96 dairy farms in Vermont between 1971 and 1984. Results showed that the average technical efficiency was approximately 77 per cent and the size effect (56 per cent) played a greater role than productivity growth (44 per cent) in increasing milk production. Haghiri and Simchi [38] observed that estimated efficiency scores were sensitive to the choice of functional forms and argued one reason of such discrepancies in results might come from the fact that parametric frontier functions are unable to detect the true relationships between variables used in the model. To demonstrate such relationship, [38] used the alternating conditional expectation (ACE) algorithm to estimate the technical efficiency of a sample data collected from New York dairy farms from 1997 to 1998. The ACE algorithm showed that a power functional form was an appropriate model for the sample data. The estimated mean technical efficiency was found to be approximately 67.0 per cent. In another study, [16] compared the estimated technical efficiency scores of Ontario and New York dairy producers between 1992 and 1998 using a nonparametric stochastic frontier model. The result showed that during the period of study, New York dairy farmers produced milk more efficiently than Ontario dairy producers, but the magnitude of the difference was small. The estimated mean technical efficiency for the former group was 60.2 per cent as compared to 53.2 per cent for the latter. [16] also found no correlation between farm size and the estimated technical efficiency scores. Finally, [10] used a nonparametric extension of estimating generalized quadratic Box-Cox models and estimated the parameters of the model by utilizing the additivity and variance stabilization algorithm. The new method accounts for random noise in the data and relaxes the sensitivity of technical efficiency scores to the choice of functional form. It also provides more flexible choices for estimating the parameter of the dependent variable. Haghiri and Simchi [10] specified the model to measure technical efficiency scores of New York dairy producers from 1990 to 2000. Results showed the sample producers did not use resources efficiently, as the estimated mean technical efficiency scores were found to be 0.663.

4. EMPIRICAL ANALYSIS

4.1 Data Description

This study used 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 used 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 were used. The dependent variable was defined as the total fluid milk production of 3.7 per cent fat content per year measured in pounds. We measured land in hectare as annual total (owned and rented) tillable area. Lack of data prevented us from separating pasture and planted areas. We defined labor as annual total equivalent worker unit (hereafter, ewu), which contained both hired and family labor as well as the working hours of the main operator where it was applicable. We added all the payments made to rent machinery and farm equipment, purchase oil and parts, and depreciation together and considered them as the total farm-specific machinery accrual expenses. Livestock expenses were computed by adding the amount of money spent on cattle lease, breeding, livestock replacement, veterinary, and milk marketing. We calculated 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 hey. We also included accrual expenses related to non-farm production processes as miscellaneous expenditures. These expenses included electricity bills, insurance invoices, interest payments, and building depreciation. We deflated the last four groups of variables using the appropriate Producer Price Index (*PPI*), and converted all the variables (except socioeconomic variables) to logarithmic forms to mitigate possible heteroskedasticity in the model. Finally, we divided 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 was 18,784 pounds with a standard deviation of 3,244 during the period of study. The average tillable area in the sample data was 4.39 hectares and the mean labor used was 0.035 ewu per each milking cow. The amount of money spent on purchasing grain and roughage ($US 7.79 per milking cows) was the highest accrual expenses in the sample data. Table 1 also shows that each dairy farm in the sample data, on average, had 93 milking cows during the period of the study. Finally, the average age of the primal farm operator was 43 years and the average years of schooling was approximately 14 years.

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.

4.2 Estimation Results

We, first, estimated the mean response function f defined in equation (2) by using *R* (version 2.10.1), and then measured the individual technical efficiency scores using the COLS approach. In particular, we considered 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 assumed a parametric relationship between the dependent variable and the socioeconomic variables on the ground that dairy farm operators accumulated more experience each year as they got older and participated in formal and informal school training. Table 2 presents the estimated technical efficiency scores of the sample data, which was classified by the group mean performance (hereafter, efficiency class interval).

Table 2. Technical efficiency of Wisconsin dairy producers

Efficiency class interval	No. of firms	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.

For each of the efficiency class intervals and the overall average estimated technical efficiency scores, we obtained lower and upper bounds of confidence limit by using a bootstrap technique, so-called the bias-corrected and accelerated percentile method with 1000 replications [39]. The result showed that the average technical efficiency scores of the sample data were 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 was higher than that of the ones which were reported in [16] and [10] for the similar studies conducted for the New York dairy producers, lower than that of what [36] stated for the New England dairy farms, and about the same magnitude for the Vermont dairy industry [37]. The estimated mean technical efficiency scores in the aforementioned studies was 60.2, 66.3, 80.3, and 77.2 per cent, respectively. These findings are important to policymakers because the same U.S. policy implemented in all the regions during the period of study. Table 2 also shows that the majority (84.8 per cent) of the sample dairy farms fell in the category of 66-85 per cent technical efficiency. Fig. 1 shows the estimated technical efficiency class interval of the sample data.

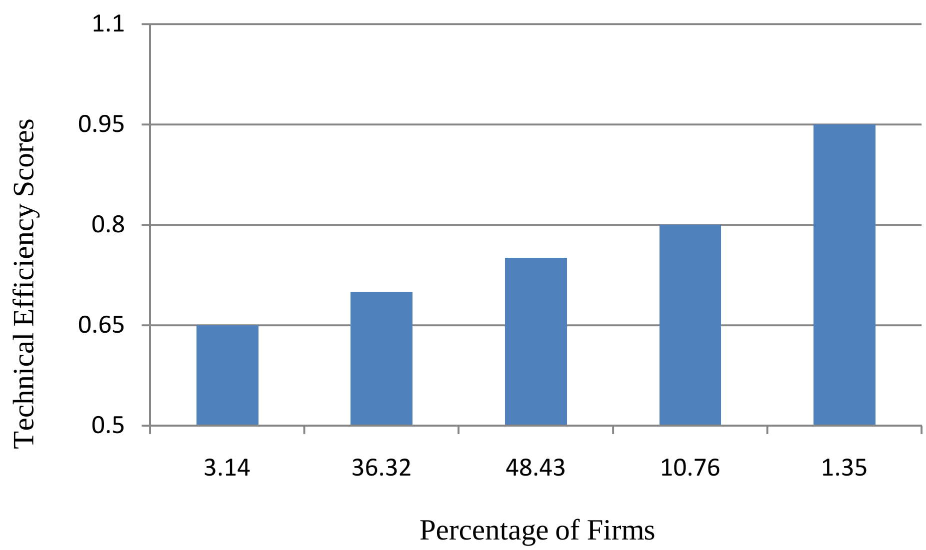

Fig. 1. Technical efficiency class interval

To investigate variations among the estimated technical efficiency scores in the sample data, we conducted a two-factor ANOVA test with no replication. The computed *F*-test value (12.02) rejected the null hypothesis of no variation amongst 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 varied significantly from year-to-year. Furthermore, we used 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 allowed us to find whether the estimated technical efficiency scores and time trend are independent of each other [40]. Since the calculated value of the chi-square (110.93) with 20 degrees of freedom exceeded the statistic critical value (37.56) at the 0.05 level of significance it was reasonable to conclude that technological changes were occurred in the sample data between 1993 and 1998.

To examine the distribution type of the estimated technical efficiency scores we conducted a univariate normality test for the sample data [41]. The calculated *p*-value of the statistic (0.441) indicated that the null hypothesis of normal distribution of the estimated technical efficiency scores at the 0.05 level of significance was not rejected. In addition, we specified a regression model to investigate factors that caused inefficiency in the sample data. To do this, we used 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 showed that none of the four predictors was 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) were all positive while it was negative for milk system (-0.0041). Precedent studies also reported similar findings to what we found in this paper. For instance, [36] concluded that the analysis of the relationships between the estimated technical and allocative efficiencies and four socioeconomic variables, such as farm size, education, extension, and experience revealed that the socioeconomic variables did not affect the magnitude of the efficiency scores of the New England dairy producers. Finally, we found no correlation between farm size and the estimated technical efficiency scores**.** The estimated sample correlation coefficient (0.065) was positive, but it was not significant at the 0.05 level.

5. CONCLUSION AND OUTLOOK

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 would affect productive efficiency of the U.S. dairy producers. Several methods have been introduced in the literature to estimate frontier functions; none of them provides accurate results because the estimation of efficiency scores is time-and-data specific. As a result, the implication of any policies made from the results of such studies would lead to misallocation of scarce resources. 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 within the framework of generalized additive models to estimate technical efficiency scores of a sample data collected from the Wisconsin dairy industry database between 1993 and 1998. The parameters of the model were estimated using generalized alpine smoothing technique. The result showed that the sample of Wisconsin dairy producers was, on average, 77.8 per cent technically efficient and the distribution of the estimated technical efficiency scores was normal. From managerial perspective, this means that the same amount of output could have been produced with 22.2 per cent fewer input used if all dairy farms had operated at 100 per cent efficiency. In other words, producers could decrease their production cost per unit and earn higher profit margins. This is an important issue since the U.S. dairy industry was not under the supply management policy, which guarantees a higher output price than of the prevalent market price. We suggest the following areas for further research. First, lack of the information related to the input and output prices for each of the firms in the sample prevented us from measuring allocative and overall efficiency scores. Second, we used labor in terms of physical quantity and not in terms of dollar values. This might raise the question that how sensitive the findings of the research would be if labor was measured in terms of dollar values. Finally, since the estimation of marginal products is not part of the objectives of this study, we suggest another study that uses the gradient estimation method for additive nonparametric and semiparametric regression models to measure the marginal product indices [42].

COMPETING INTERESTS

The author has declared that no competing interests exist.

REFERENCES

1. Chennareddy V. Production efficiency in south Indian agriculture. Journal of Farm Economics. 1967;49(4):816-820.
2. Lau LJ, Yotopoulos PA. A test for relative efficiency and applications to Indian agriculture. American Economic Review. 1971;61(1):94-109.
3. Farrell MJ. The measurement of productive efficiency. Journal of the Royal Statistics Society Series. 1957;120(3):253-290.
4. Aigner DJ, Lovell CAK, Schmidt P. Formulation and estimation of stochastic frontier production functions. Journal of Econometrics. 1977;6(1):21–37.
5. Haghiri M. 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. 2003.
6. Kumbhakar SC, Lovell CAK. Stochastic frontier analysis. 1st edition. Cambridge: Cambridge University Press; 2003.
7. Haghiri M. Applied nonparametric regression analysis: the choice of generalized additive models. Review of Economics and Finance. 2013;3(1):25-34.
8. Kumbhakar SC, Tsionas EG. Nonparametric stochastic frontier models. 2nd North American Productivity and Efficiency Conference, Union College, Schenectady, Albany, New York, June 20-22; 2002.
9. Anderson DP, Chaisantikulawat T, Tan Khee GA, Kebbeh M, Lin N, Shumway CR. Choice of functional form for agricultural production analysis. Review of Agricultural Economics. 1996;18(2):223-231.
10. Haghiri M, Simchi A. A nonparametric extension of generalized quadratic Box-Cox models in measuring technical efficiency. Applied Economics. 2010;42(1):1-8.
11. Li Q, Huang CJ, Li D, Fu T. Semiparametric smooth coefficient models. Journal of Business and Economic Statistics. 2002;20(3):412-422.
12. Muller M. Estimation and testing in generalized partial linear models: a comparative study. Statistics and Computing. 2001;11(4):299-309.
13. Sun K, Kumbhakar SC. Semiparametric smooth-coefficient stochastic frontier model. Economics Letters. 2013;120(2):305-309.
14. Wahba G. Spline models for observational data, 1st edition. Philadelphia: CBMS-NSF Regional Conference Series SIAM (Society for Industrial and Applied Mathematics); 1990.
15. Richmond J. Estimating the efficiency of production. International Economic Review. 1974;15(2):515-521.
16. Haghiri M, Nolan JF, Tran KC. Assessing the impact of economic liberalization across countries: a comparison of dairy industry efficiency in Canada and the USA. Applied Economics. 2004;36(11):1233-1243.
17. Hastie TJ, Tibshirani RJ. Generalized additive models. 1st edition. London: Chapman and Hall; 1990.
18. McCullagh P, Nelder J. Generalized linear models, 2nd edition. London: Chapman and Hall; 1989.
19. Kneip A, Simar L. A general framework for frontier estimation with panel data. Journal of Productivity Analysis. 1996;7(2/3):187- 212.
20. Schimek MG, Turlach BA. Additive and generalized additive models. In: Schimek MG, editor. Smoothing and regression: approaches, computation, and application. 1st edition. New Jersey: John Wiley and Sons; 2000:277-327.

21. Hardle W, Mammen E, Muller M. Testing parametric versus semiparametric modelling in generalized linear models. Journal of the American Statistical Association. 1998;93(444):1461–1474.
22. Severini TA, Staniswalis JG. Quasi-likelihood estimation in semiparametric models. Journal of the American Statistical Association. 1994;89(426):501–511.
23. Friedman JH. Multivariate adaptive regression splines. The Annals of Statistics. 1991;19(1):1-67.
24. Ozmen A, Batmaz I, Weber GW. Robust conic multivariate adaptive regression splines (RCMARS) and an application to precipitation modeling. Technical Report No.219. Institute of Applied Mathematics. Middle East Technical University. Ankara. Turkey; 2012.
25. Taylan P, Weber GW, Liu L, Yerlikaya-Ozkurt F. On the foundations of parameter estimation for generalized partial linear models with B-splines and continuous optimization. Computers & Mathematics with Applications. 2010;60(1):134-143.
26. Yerlikaya-Ozkurt F, Batmaz I, Weber GW. A review of conic multivariate adaptive regression splines (CMARS): a powerful tool for predictive data mining. Technical Report No.15. Institute of Applied Mathematics. Middle East Technical University. Ankara. Turkey; 2012.
27. Eubank RL. Spline regression. In: Schimek MG, editor. Smoothing and regression: approaches, computation, and application. 1st edition. New Jersey: John Wiley and Sons; 2000:1-18.
28. Stone M. Cross-validatory choice and assessment of statistical predictions. Journal of the Royal Statistical Society Series B. 1974;36(2):111-147.
29. Chen R, Hardle W, Linton OB, Sevarance-Lossin E. Nonparametric estimation of additive separable regression models. Statistical Theory and Computational Aspects of Smoothing. Physica Verlag. 1996;247-265.
30. Friedman JH, Stuetzle W. Projection pursuit regression. Journal of the American Statistical Association. 1981;76(376):817-823.
31. Greene WH. The econometric approach to efficiency analysis. In: Fried HO, Lovell CAK, Schmidt SS, editors. 1st edition. The measurement of productive efficiency and productivity change. London: Oxford University Press; 2008:1-158.
32. Cornwell C, Schmidt P, Sickles R. Production frontiers with cross sectional and time series variation in efficiency levels. Journal of Econometrics.1990;46(1/2):185-200.
33. Evans DA, Tandon CM, Lauer J. 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; 2000.
34. Simar L. Estimating efficiencies from frontier models with panel data: a comparison of parametric, nonparametric, and semiparametric methods with bootstrapping. Journal of Productivity Analysis. 1992;3(1/2):171-203.
35. Kumbhakar SC, Biswas B, Bailey D. A study of economic efficiency of Utah dairy farmers: a system approach. The Review of Economics and Statistics. 1989;71(4):595-604.
36. Bravo-Ureta BE, Rieger L. Dairy farm efficiency measurement using stochastic frontiers and Neoclassical duality. American Journal of Agricultural Economics. 1991;73(2):421-428.
37. Ahmad M, Bravo-Ureta BE. An econometric decomposition of dairy output growth. American Journal of Agricultural Economics. 1995;77(4):914-921.
38. Haghiri M, Simchi A. A general solution to Popper or production. Empirical Economics Letters. 2003;2(1):51-73.

39. Briggs AH, Mooney CZ, Wonderling DE. Constructing confidence intervals for cost-effectiveness ratios: an evaluation of parametric and nonparametric techniques using Monte Carlo simulation. Statistics in Medicine. 1999;18(23):3245-3262.
40. Levine DM, Berenson ML, Krehbiel TC, Stephen DF. Statistics for manager using MS Excel. 6th edition. New Jersey: Prentice Hall; 2011.
41. Shapiro SS, Wilk MB. An analysis of variance test for normality (complete samples). Biometrika. 1965;52(3/4):591-611.
42. Ozabaci D, Henderson DJ. Gradients via oracle estimation for additive nonparametric regression models with application to returns to schooling. Working papers. Department of Economics, Finance and Legal Studies. University of Alabama. Tuscaloosa. Alabama. USA; 2012.

Peer-review history:
The peer review history for this paper can be accessed here:
http://www.sciencedomain.org/review-history.php?iid=247&id=20&aid=1773