American Journal of Agricultural Economics (1999), 81: 894-901

USING CROSS-SECTION DATA TO ADJUST TECHNICAL EFFICIENCY INDEXES ESTIMATED WITH PANEL DATA

Antonio M. Álvarez and Eduardo González

Abstract

This article proposes a procedure to incorporate cross sectional information in the estimation of technical efficiency indexes obtained from panel data. A conventional index of technical efficiency is estimated in a first stage using panel data on inputs and outputs. Then, the individual effects from the first stage are adjusted using cross sectional information, obtaining a corrected technical efficiency index. The model is applied to a panel of 82 Spanish dairy farms, where only cross sectional information about input quality is available. An analysis of variance is performed between some variables and both the corrected and the uncorrected indexes, finding that the conclusions derived from both analyses are different.

Key words: dairy farms, input quality, technical efficiency, panel data.

Antonio M. Álvarez (alvarez@uniovi.es) is associate professor in the Department of Economics at the University of Oviedo, Spain. Eduardo González (efidalgo@uniovi.es) is teaching assistant in the Department of Business at the University of Oviedo, Spain.

The authors gratefully acknowledge valuable comments received from Rafael Álvarez, Manuel Arellano, Carlos Arias, Boris Bravo, José M. Campa, Art Havenner, Knox Lovell, and two anonymous reviewers. The authors also thank editor Spiro Stefanou for his helpful suggestions.

The measurement of technical efficiency (TE) has been a popular field of research since the seminal paper by Farrell. His approach has given rise to a growing body of literature based on the notion of frontiers, which has focused mainly on the estimation of efficiency indexes. The interest in this type of study lies in the relationship between inefficiency and production costs. An inefficient firm is wasting inputs because it does not produce the maximum output attainable, given the amount of inputs used, and therefore there is room to reduce average cost. The main drawback of many TE studies is that they provide few practical prescriptions for firms. One of the problems with the Farrell index is that it is a scalar measure and does not give information about improper utilization of specific inputs.

Some studies investigate the causes of technical inefficiency. The usual procedure has been to run a regression of the technical efficiency index on several variables (e.g., firm size, education, age) in what has been termed in the literature as the second step[1].

Most of these studies interpret the estimated TE index as a measure of management with the assumption that the role played by differences in omitted variables is negligible. However, a TE index cannot, in general terms, be interpreted as a measure of management, because it may be confounded by unobservables, such as unmeasured inputs, differences in input (or output) quality, and differences in technology. This argument can be traced back to Timmer who argued that inefficiency seems to be due to definitional and measurement problems in the variables. Page stresses the importance of input quality, suggesting that, if inputs were more fully specified and variables more completely defined, much of the apparent variation in efficiency levels would presumably disappear.

In this article we assess the importance of not accounting for input quality in the estimation of TE indexes[2]. It is a common problem to have panel data on inputs and outputs but only cross sectional information on input quality. We develop a method to combine panel and cross sectional data in the estimation of technical efficiency indexes to assist the estimation of TE indexes in these situations. The method proceeds in two stages. First, a production function is estimated using panel data on inputs and outputs. Then, cross sectional data on input quality are used to adjust the individual effects from the first stage. The model is applied to a sample of Spanish dairy farms.

The next section reviews the panel data techniques to estimate the TE indexes, followed by a discussion of the procedure used to correct the indexes. We then present the data and summarize the empirical results along with a comparative analysis of the corrected and uncorrected indexes.

The Measurement of Technical Efficiency

A frontier production function gives the maximum output attainable for any combination of inputs, in such a way that all firms lay on or below the frontier. The deviation of a firm from its frontier is interpreted as technical inefficiency and, therefore, a measure of TE is given by the ratio of current to maximum production, the latter being given by the production on the frontier.

As the frontier literature has evolved, the stochastic frontiers have superseded the early deterministic models. The conceptual implications of the stochastic frontier are important for the interpretation of inefficiency (Aigner, Lovell, and Schmidt). Also, Schmidt and Sickles demonstrate that efficiency measures are inconsistent when estimated from cross section data. Panel data based econometric models can generate consistent estimates that account for unobservable heterogeneity among firms. A stochastic frontier production function with panel data can be written as:

(1) $y_{it} = \alpha + \mathbf{x_{it}}'\beta + \lambda_t + \varepsilon_{it} - v_i$

where y_{it} is the output of firm i in period t, $\mathbf{x_{it}}$ the vector of inputs, the parameters λ_t are the time effects and α is the intercept. The error term is composed of two terms: a one sided error term (v_i) that represents technical inefficiency and a symmetrical error term (ε_{it}) that is independent and identically distributed with zero mean and constant variance (σ_ε). This model transforms into (2) by setting $\alpha_i = \alpha - v_i$:

(2) $y_{it} = \alpha_i + \mathbf{x_{it}}'\beta + \lambda_t + \varepsilon_{it}$

where α_i are the individual effects that can be considered fixed or random. The appropriate choice of the model depends on the correlation between the individual effect and the explanatory variables.

For the fixed effects model, which we apply in this article, relative indexes of technical efficiency can be computed from the comparison of the individual effects. In the case of a logarithmic specification, the expression is[3]:

(3) $TE_i = \exp(\alpha_i - \max \alpha_j)$

where TE_i is the TE level of firm i. This index takes the value 1 for the firm with the largest individual effect. The remaining firms obtain indexes lower than 1, reflecting the existence of unobservables making them less efficient.

The Adjusting Procedure

It is common to interpret the efficiency indexes as measures of management. However, differences in TE can be attributed to unmeasured inputs, differences in input quality and different technologies. Management can be one of these unmeasured inputs.

Farrell and succeeding studies focus on TE measures as including differences in input quality. Inputs are usually assumed to be homogeneous but this can be an oversimplification. For example, feedstuffs for dairy cows are an aggregate measure of many different types of feed that can vary in quality. Also, land, as measured by hectares, does not take into account differences in slope or soil fertility[4].

The above discussion suggests that complementary information about the heterogeneity captured by the fixed effect, such as input quality, can be used to adjust the TE indexes. Based on the assumption of a common technology, the remaining differences among firms will be largely due to management.

The relevant information to calculate the TE index is contained in the individual effects. A large individual effect implies that there are unobservable factors that make this firm more productive than others. Our intent is to determine which part of the individual effect is due to the quality of the inputs used and which part is not. One way to proceed is by estimating a frontier of individual effects depending on the level of quality. Given a level of quality, the frontier determines the potential individual effect, which can be compared with the actual individual effect to adjust the level of TE.

We perform this estimation by Corrected Ordinary Least Squares. This method involves regressing the estimated fixed effects ($\hat{\alpha}_i$) on a set of variables that measure the quality of the inputs ($\mathbf{Z_i}$) via Ordinary Least Squares

(4) $\hat{\alpha}_i = a + \mathbf{Z_i'b} + u_i$,

and then correct the residuals by the largest positive one. The fitted value $\hat{\hat{\alpha}}_i$ is corrected by the largest positive residual to yield α_i^*, the potential individual effect for the level of quality $\mathbf{Z_i}$ given by the frontier[5].

(5) $\alpha_i^* = \hat{\hat{\alpha}}_i + \max \hat{u}_j$

The adjusted index of TE uses α_i^* to yield

(6) $TE_i^* = \exp(\hat{\alpha}_i - \alpha_i^*)$

Figure 1 illustrates the adjustment procedure. The line FF' is the individual-effects frontier representing the maximum α attainable for each level of quality. The conventional index of TE for firm i is measured as exp($\hat{\alpha}_i$ -maxα_j). This index does not allow that firm i is using less quality than firm h (the one with the largest $\hat{\alpha}$) and, thus, part of the difference between their individual effects is due to the difference in input quality. The index is adjusted by making comparisons to the potential individual effect given the level of quality, α^*_i. The adjusted index increases because the new measure takes into account the fact that this firm uses less quality than firm h. For firms with more quality than firm h, the adjusted index will be smaller[6].

<<<<<<<<<<<<<<<<<<<<INSERT FIGURE 1 HERE>>>>>>>>>>>>>>>>>>>>>>

Data

This study uses technical and accounting data from a balanced panel of 82 dairy farms observed in the period 1987-1991. The farms are located in Asturias, a region in Northern Spain, where dairy farming is the main agricultural activity. These farms participate in a voluntary record keeping program and, despite the wide range of sizes, they are family farms that are assumed to use a common technology.

The variables used in the estimation of the production frontier are:

Milk (M) : Milk production (thousands of liters).

Labor (L) : Total cost of labor (hundreds of thousands of pesetas)

Land (H) : Total farm area (hectares).

Cows (C) : Number of milking cows.

Feedstuffs (F) : Total amount of feedstuffs fed to the dairy cows (tons). Because the farms have different replacement rates, feedstuffs have been adjusted to include only concentrates given to milking cows.

Roughage (R) : Total expenditures necessary to produce forage crops. It includes expenses such as seeds, sprays, fertilizer and depreciation of machinery (hundreds of thousands of pesetas).

Since labor and roughage are measured in monetary units, they were deflated by an index of prices paid by farmers. Table 1 presents descriptive statistics of the variables. The coefficients of variation indicate the existence of an important degree of heterogeneity among the production decisions of the farms in the sample.

<<<<<<<<<<<<<<<INSERT TABLE 1 HERE>>>>>>>>>>>>>>>>>>>>>>>>>

Empirical Results

The functional form employed in the empirical analysis is the translog production function with dummies to control for individual effects[7]:

$$(7) \qquad \ln M_{it} = \alpha_i + \Sigma_j \beta_j \ln x_{jit} + 1/2\Sigma_j\Sigma_k \beta_{jk} \ln x_{jit} \ln x_{kit} + v_{it}$$

$$\beta_{jk} = \beta_{kj} \qquad for \quad k \neq j$$

where the subscript i makes reference to the farm, t is the time period and j and k denote inputs[8].

Expression (7) was computed by OLS using the WITHIN estimator for a fixed effects model (Schmidt and Sickles). We applied White's estimator of the variance-covariance matrix which corrects for heteroscedasticity to obtain efficient estimates. The results are presented in Table 2, yielding an R^2=0.98 and mean technical efficiency equal to 72%.

<<<<<<<<<<<<<<<INSERT TABLE 2 HERE>>>>>>>>>>>>>>>>>>>>>>>>>

Table 3 shows the input elasticities evaluated at the geometric mean of the sample.

<<<<<<<<<<<<<<<INSERT TABLE 3 HERE>>>>>>>>>>>>>>>>>>>>>>>>>

All elasticities are positive and significantly different from zero, as expected. The largest values correspond to the inputs cows and feedstuffs implying, at the margin, that the main contribution to milk increases are due to these two inputs.

The cross sectional information about input quality comes from an auxiliary survey carried in 1991. The variables are used to adjust the TE indices and are[9]:

Age (A) : Age of first manager. There is no expected sign for this variable, since it incorporates two opposing effects of experience and aging.

Artificial pasture (P) : Proportion of artificial pasture and forage crops in total land. The remaining land is natural pasture, which is less productive. Therefore a positive sign is expected.

Genetics (G) : This variable is measured by the cost (pesetas) of the most expensive semen dose. It is a proxy for the genetic level of the dairy herd. Therefore, the expected sign is positive.

Silage (S) : Total amount of silage per cow (cubic meters). This feed is supposed to boost milk yields and therefore a positive sign is expected.

PTA (T) : This variable reflects the level of part-time agriculture. It is a dummy variable that takes the value 1 if the family members earn money from other sources.

Zone (Z) : This is a dummy variable that takes the value 0 for the farms on the coast and 1 for the farms inland. There is no expected sign for this variable.

Table 4 provides some descriptive information about these variables.

<<<<<<<<<<<<<<<<<INSERT TABLE 4 HERE>>>>>>>>>>>>>>>>>>>>>>>>>>>>

The quality adjustment was performed using the following model:

(8) $$\hat{\alpha}_i = a + b_A \ln A_i + b_P \ln P_i + b_G \ln G_i + b_S \ln S_i + b_T \ln T_i + b_Z \ln Z_i + u_i$$

where $\hat{\alpha}_i$ is already expressed in logarithms. The results from the estimation are reported in Table 5. All the coefficients present the expected sign and are significantly different from zero, except the artificial pasture variable. The part time agriculture and geographic zone dummies are not statistically significant. The age variable has a positive and significant coefficient suggesting that the effect of experience is more important than the effect of aging. The mean efficiency stays at a level of 72%[10].

<<<<<<<<<<<<<<<<<INSERT TABLE 5 HERE>>>>>>>>>>>>>>>>>>>>>>>>>>>>

Figure 2 illustrates the relationship between the corrected and the uncorrected indexes, whose correlation is 0.88. The departures from the 45º line reflect the importance of input quality as a component of the uncorrected measure. The efficiency of the farms above the line was underestimated with the primitive measure, and overestimated for the farms below the line. On average both indexes are the same (72%). However, an analysis of the sources of efficiency needs to determine correctly the particular index of every farm.

<<<<<<<<<<<<<<<<<INSERT FIGURE 2 HERE>>>>>>>>>>>>>>>>>>>>>>>>>>>

Analysis of the TE Index

In this section, we analyze the relationship between the index of TE and several variables of size and productivity. An analysis of variance is performed on these variables for both the uncorrected and the corrected index of TE to assess the effect of the correction in the conclusions. Farms are ranked based on their indexes and split into 3 groups of equal size labeled low, average, and high efficiency. For each group we calculate the mean of each variable and test whether the differences among the groups are statistically significant or not. In particular, we study two broad groups of variables: size variables (milk, cows and land) and productivity variables (average products).

Table 6 presents the results of the analysis of variance before the correction. The exact definition of the three efficiency intervals is given in brackets. With respect to the size variables, there is a clear positive relationship of TE with milk, while cows and land are not significant. From the five average products, only those corresponding to cows, land, and labor have a significant positive relationship with TE.

<<<<<<<<<<<<<<<INSERT TABLE 6 HERE>>>>>>>>>>>>>>>>>>>>>>>>>

However, the previous results were obtained from an index of TE that incorporates the effect of input quality differences. For example, the fact that the most efficient farms have the highest milk yield per cow is not informative if these farms are also the farms that have the best cows. For instance, the genetics variable controls for differences in the quality of the cows. Thus, it is appropriate to compare the former results with those obtained using the adjusted TE index.

The results of the analysis of variance using the corrected index of TE are shown in Table 7. The previous relationship between TE and milk disappears, suggesting that most of the correlation was due to differences in input quality. We now find a strong negative relationship between TE and land and cows, which did not emerge in the earlier analysis. The largest farms now appear to be the least efficient once the effect of quality has been addressed. This may suggest, holding management capacity constant, that the farms in the largest size category are experiencing limits to a manager’s span of control over the farm operation. This suggests that these farms are operating at a level consistent with the upward-sloping portion of the average cost curve[11].

The average products follow similar patterns to those before the correction, but now the differences in liter of milk per kilogram of feedstuffs are statistically significant.

The smaller differences found in the average products of cows, land and labor suggest that some of those differences were due to the quality of the inputs.

<<<<<<<<<<<<<<<INSERT TABLE 7 HERE>>>>>>>>>>>>>>>>>>>>>>>>>

Summary and Conclusions

In this article we present a model where the TE index obtained from panel data is adjusted using additional cross sectional information. First we estimated a translog production function using a fixed effects model and then we regressed the individual effects on a set of quality variables obtained through an auxiliary survey made in one year of the panel. The results from this regression are used to compute the corrected TE index, which is analyzed using Analysis of Variance. This procedure is useful in the case that relevant information, such as input quality, is not available for all the periods of the panel. The main finding in this article is that the results obtained in TE studies are more sensitive to measurement error than it had been previously thought.

The results show that some of the conclusions obtained in studies of TE may depend heavily on the information about input quality. For instance we see that, before the adjustment, TE appears to be positively related with the farm’s size. Once the effect of input quality has been considered this relationship turns to be the opposite. Thus, a wrong recommendation could be given to the farms if the information about input quality is not available.

References

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

Battese, G.E., and T.J. Coelli. “A Model for Technical Inefficiency Effects in a Stochastic Frontier Production Function for Panel Data.” *Empiric. Econ.* 20(1995):325-332.

Bravo-Ureta, B.E., and A.E. Pinheiro. “Efficiency Analysis of Developing Country Agriculture: A Review of the Frontier Function Literature.” *Agr. and Resour. Econ. Rev.* 22(April 1993):88-101.

Bravo-Ureta, B., and L. Rieger. “Alternative Production Frontier Methodologies and Dairy Farm Efficiency.” *J. Agr. Econ.* 41(1990):215-226.

Farrell, M.J. “The Measurement of Productive Efficiency.” *J. Royal Statist. Soc., Series A,* Part 3 (1957):253-290.

Greene, W. “Maximum Likelihood Estimation of Econometric Frontier Functions.” *J. Econometrics* 13(May 1980):27-56.

Hallam, D., and F. Machado. “Efficiency Analysis with Panel Data: A Study of Portuguese Dairy Farms.” *Eur. Rev. Agr. Econ.* 23(1996):79-93.

Lass, D.A., and C.M. Gempesaw II. “Estimation of Firm-Varying, Input-Specific Efficiencies in Dairy Production.” *Northeastern. J. Agr. and Resour. Econ.* 21(October 1992):142-150.

Page, J.M. Jr. “Firm Size and Technical Efficiency: Applications of Production Frontiers to Indian Survey Data.” *J. Develop. Econ.*, 16(September/October 1984):129-152.

Pitt, M.N., and L.F. Lee. “The Measurement and Sources of Technical Inefficiency in the Indonesian Weaving Industry.” *J. Develop. Econ.* 9(August 1981):43-64.

Schmidt, P., and R. C. Sickles. “Production Frontiers and Panel Data.” *J. Bus. and Econ. Statist.* 2(October 1984):367-374.

Stigler, G. “The X-istence of X-Efficiency.” *Amer. Econ. Rev.* 66(March 1976):213-216.

Tauer, L.W., and K.P. Belbase. “Technical Efficiency of New York Dairy Farms.” *Northeastern J. Agr. and Resour. Econ.* 16(April 1987):10-16.

Timmer, C. P. “Using a Probabilistic Frontier Production Function to Measure Technical Efficiency.” *J. Polit. Econ.* 79(July/August 1971):776-794.

White, H. “A Heteroscedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity.” *Econometrica* 48(May 1980):817-838.

American Journal of Agricultural Economics (1999), 81: 894-901

Footnotes

[1] An early example is the paper by Pitt and Lee while a more recent contribution is Hallam and Machado. Bravo-Ureta and Pinheiro summarize the pros and cons of this approach, while a recent paper by Battese and Coelli criticizes this second step because the efficiency indexes are not identically distributed.

[2] Throughout the rest of the article, we will refer only to differences in input quality. However, all the arguments and the adjustment procedure can be applied directly to the case of differences in unmeasured inputs.

[3] The maximum fixed effect corresponds to the observation with v_i=0 and therefore coincides with the intercept of expression (1).

[4] Stigler noted that differences in observed inefficiency can be attributed to different technologies. Lass and Gempesaw have tried to account for this possibility using a random coefficients model. However, TE studies usually assume that firms operate with a common technology. In fact, this argument can be considered as a special case of differences in input quality, since the adoption of a new technology implies using capital stock from a different vintage.

[5] Given that this procedure uses the largest positive residual to perform the adjustment, the results can be affected by the existence of outliers.

[6] For simplicity, we assume in Figure 1 that the firm with the largest individual effect is also the firm that determines the frontier.

[7] The Cobb-Douglas specification was tested versus the Translog and rejected at the 1% significance level.

[8] Time effects are not included due to multicollinearity problems. In our case, year dummies reflect climate variations, including rain and temperature. When climate conditions are adverse farmers use larger quantities of roughage and feedstuffs and, thus, these variables account for the climate factors, making the use of year dummies redundant.

American Journal of Agricultural Economics (1999), 81: 894-901

[9] Note that there is not a one to one correspondence between these variables and the five inputs considered. For instance, the input *land* is qualified by the *proportion of artificial pasture* and also by the *zone* dummy variable, as it reflects differences in slope and climate conditions. On the other hand, not only *age* but also *PTA* qualifies *labor*. The adjustment proposed in this paper accounts for these types of relationships by estimating the impact of all the quality variables on the fixed effect.

[10] We also estimated equation (8) using several interaction terms. However, this procedure introduced high multicollinearity, causing most of the coefficients to be non significant.

[11] This result is at odds with similar studies in the dairy sector. Bravo-Ureta and Rieger and Tauer and Belbase found a positive relationship between TE and cows although they did not include cows as an input in the production function. Hallam and Machado found positive correlation between TE and total farm value added.

Table 1.- Descriptive statistics of the data

Variable	Mean	Coeff. Variation	Min	Max
Milk	100.1	0.52	27.6	386.3
Labor	12.1	0.25	6.3	24.7
Land	12.1	0.32	5.5	25.0
Cows	21.0	0.39	8.1	56.0
Feedstuffs	31.2	0.67	1.1	149.3
Roughage	8.9	0.59	0.95	36.5

Table 2.- Estimates of the Translog Frontier Production Function

Parameter	Coefficient	t-ratio	Parameter	Coefficient	t-ratio
β_L	0.06	0.18	β_{LH}	0.01	0.10
β_H	-0.77	-1.35	β_{LC}	-0.19	-1.20
β_C	0.43	0.83	β_{LF}	0.12	2.00*
β_F	0.21	0.92	β_{LR}	0.13	2.30*
β_R	0.004	0.02	β_{HC}	0.65	2.51*
β_{LL}	-0.03	-0.19	β_{HF}	-0.17	-1.22
β_{HH}	-0.07	-0.22	β_{HR}	-0.16	-1.77*
β_{CC}	-0.26	-0.80	β_{CF}	-0.06	-0.66
β_{FF}	0.16	4.23*	β_{CR}	0.05	0.56
β_{RR}	0.08	1.56	β_{FR}	-0.06	-1.75*

* Significant at the 0.1 level.

Table 3.- Input elasticities at the geometric mean

Input	Elasticity	t-ratio
Labor	0.10	2.88*
Land	0.14	2.56*
Cows	0.68	12.43*
Feedstuffs	0.26	8.87*
Roughage	0.04	1.99*

* Significant at the 0.1 level.

Table 4.- Descriptive statistics of the input quality variables

Variable	Mean	Coeff. Variation	Min	Max
Age	46.3	0.19	25.0	63.0
Artificial pasture	0.41	0.54	0.03	1.0
Genetics	6079	0.72	1000	15000
Silage	2.49	0.99	0.02	11.4
PTA	0.47	1.05	0	1
Zone	0.58	0.85	0	1

Table 5.- Estimation of the regression to correct the fixed effects

Variable	Parameter	Coefficient	t-ratio
Intercept	a	1.48	4.04*
Age	b_A	0.12	1.73*
Artificial pasture	b_P	-0.02	-0.71
Genetics	b_G	0.07	2.68*
Silage	b_S	0.02	1.93*
Part time agriculture (PTA)	b_T	0.04	1.40
Geographic zone	b_Z	0.02	0.44

* Significant at the 0.1 level.

Table 6.- Analysis of variance of the uncorrected index of TE

	TECHNICAL EFFICIENCY			
	Low (51-67)	Average (68-74)	High (75-100)	F-test
SIZE VARIABLES				
Milk	85.0	100.5	114.4	2.5*
Cows	20.9	21.0	21.0	0.0
Land	13.2	12.0	11.2	1.9
PRODUCTIVITY VARIABLES				
Milk/cow	3.9	4.6	5.4	42.1*
Milk/land	6.5	8.1	10.3	12.1*
Milk/labor	6.5	7.9	10.2	9.4*
Milk/feedstuffs	3.5	3.6	3.8	0.5
Milk/roughage	11.2	12.1	12.5	0.8

* Significant at the 0.1 level.

Table 7.- Analysis of variance of the quality-adjusted index of TE

	TECHNICAL EFFICIENCY			
	Low (53-68)	Average (69-75)	High (76-100)	F-test
SIZE VARIABLES				
Milk	96.4	102.6	101.3	0.1
Cows	22.6	21.4	19.0	3.0*
Land	13.1	13.3	10.1	7.4*
PRODUCTIVITY VARIABLES				
Milk/cow	4.1	4.5	5.2	14.2*
Milk/land	7.6	7.5	10.0	6.0*
Milk/labor	7.4	8.1	9.3	2.2
Milk/feedstuffs	3.1	3.8	3.9	4.7*
Milk/roughage	11.7	11.5	11.7	0.7

* Significant at the 0.1 level.