Empirical Economics (2004) 29:373–381
DOI 10.1007/s00181-003-0173-5

EMPIRICAL ECONOMICS

Quantile estimation of frontier production function

Cristina Bernini, Marzia Freo, Attilio Gardini

Department of Statistics, University of Bologna, Via Belle Arti 41, Bologna, Italy
(e-mail: freo@stat.unibo.it)

First version received: June 2001/Final version received: December 2002

Abstract. The purpose of the paper is to provide new information on the performance of frontier estimation methods, using data from Italian hotel industry. Quantile regression is also suggested as solution to frontier production function estimation. It is shown that, while the choice of estimation methods among conventional techniques significantly affects the economic analysis, quantile regression provides valuable new information by estimating the whole spectrum of production functions corresponding to different efficiency levels. In addition, the method makes available a coherent framework to analyze the performance of the conventional techiniques.

Key words: Production function, stochastic frontier model, semiparametric frontier model, quantile regression

JEL classification: C14, C16, D24

1. Introduction

Many aspects of economic analysis of productive systems are based on production function, which in economic theory describes the technical relationship between inputs and outputs of technically efficient productive units.

In this paper we provide a critical review of production function estimation in econometric literature; to this aim different and gradually more real-

We would like to thank the Co-Editor, the Associate Editor and an anonymous referee for comments and suggestions. The research was supported by the University Research Council and the National Research Council. The usual disclaimer applies.

The estimates were computed using the Roger Koenker and StatLib S-Plus routine of quantile regression and the Tim Coelli and CEPA Web site FRONTIER 4.1 Program. The data set is provided by the Ho.Re.Ca. survey conducted by ISTAT in 1992.

istic hypotheses on the efficiency distribution over the population of firms are discussed. The original purpose is to reinstate statistical and economic analysis of productive frontier in regression quantile framework. The quantile regression allows critical discussions on the main features of the frontier analysis, pointing out non neutral effects of technical inefficiency and wide heterogeneity of productive plans. The application is on Italian hotel industry data.

In Sect. 2 we present the economic framework. Section 3 provides some literature solutions to frontier estimation (the classical regression, the stochastic frontier and the semiparametric models) and a critical review of the empirical results on an economic perspective. In Sect. 4, as an alternative approach to frontier production analysis, the quantile regression estimator is proposed and empirically implemented on the hotel data set. Concluding remarks are given in Sect. 5.

2. Frontier production function

Economic analysis of productive systems is fundamentally based on production function specification and estimation. The production function is a frontier function since it describes the maximum output attainable by the firms which implement efficient production processes. In particular, if y_n^∂ is the maximum output given the input vector $\mathbf{x}_n$, for the n-th firm, then the theoretical frontier production function is represented by the equation

$$y_n^\partial = f(\mathbf{x}_n) \tag{1}$$

which corresponds to the technology upper limiting the production possibility set. At the same time, the observable production process (y_n) could be at most equal to its maximum, that is

$$y_n \leq y_n^\partial \quad . \tag{2}$$

The inequality (2) implies a truncated statistical distribution for the variable y_n over the productive processes population, whose implementation involves modifications in the conventional estimation techniques based on Gaussian error term, such as standard linear regression.

For frontier estimation it is customary to distinguish between parametric and nonparametric methods. Following the parametric approach, the frontier is represented by a probabilistic relation, $f(\mathbf{x};\, \theta)$, derived by means of econometric techniques. This approach [1], [9] is better suited for frontier estimation than standard techniques since it explicitly attempts to estimate the parameters of production function on the frontier. To this purpose, the function error term is supposed to consist of a Gaussian random component, v_n, and of a non-negative random component, u_n, which explains inefficiency and has to be subtracted from the production function in the following way:

$$y_n^\partial = f(\mathbf{x}_n) + v_n - u_n \tag{3}$$

$$v_n \sim N(0, \sigma_v^2) \tag{4}$$

$$u_n \sim D^+(\mu, \sigma_u^2) \tag{5}$$

$$u_n \perp v_n \tag{6}$$

Empirical implementation of (3)–(6) on the firms data provides estimates of the economic parameters and efficiency scores for each firm by means of distance measures obtained separating the (in)efficiency component from the overall error term.

Unfortunately, such a specification presents the negative feature for productive processes analysis of a uniform, i.e. neutral, reduction in factors' returns for the technically inefficient firms [5]. On the contrary, technical inefficiency may have different or opposite effects on factors' productivities; in addition, inefficiency could be related to the factor mix of the specific productive process.

In order to improve the implicit neutrality of efficiency distribution it is possible [7] to re-formulate the stochastic frontier model (3)–(6) substituting the unconditional mean in (5) with a conditional mean dependent on a set of explanatory factors, $\mathbf{z}_n$,

$$u_n \sim D^+(\mu_n, \sigma_u^2) \tag{7}$$

$$\mu_n = \delta' \mathbf{z}_n \ . \tag{8}$$

The rationale is to control for $\mathbf{z}_n$ factors in order to obtain a parametric model admitting consistent estimates from both economic and statistical points of view. It should be noticed that the specification in (3), (4), (7) and (8) corresponds to the more known model proposed by Battese and Coelli [2] to detect variables which may influence the efficiency.

In contrast to the parametric approach, more flexibility is offered by the nonparametric method, which does not impose any explicit functional form on the frontier. It is obtained by means of linear programming techniques through the envelopment of the data representing productive plans. Since this method does not accommodate for stochastic shocks, the deviations from the frontier are entirely attributed to inefficiency; the two fundamental methodologies are Free Disposal Hull -FDH- [4] and Data Envelopment Analysis – DEA [3]. The nonparametric methodologies well approximate the best practice frontier. Nevertheless, since deterministic, they are influenced by measurement errors. Moreover they focus on the efficiency measurement not providing economic parameters estimates of the production function.

An interesting solution for efficiency analysis is constituted by the semiparametric method [11], [12] that combines non parametric and parametric approaches in a two stage manner, as follows: FDH is used to identify efficient units; these results are carried over in the form of dummy variables which are incorporated in the statistical regression used to estimate the frontier. The semiparametric method provides results for the economic analysis of productive frontier system and partially controls the determinism of the nonparametric filter.

In order to compare the ability of previous production function models to satisfy frontier analysis, an empirical application is carried out.

3. Parametric and semiparametric approaches

The hotel industry has been one of the most interesting Italian service sector in the last decades and it is characterized by largely heterogenous production plans constrained either by the technology or by the localization [6].

A Cobb-Douglas frontier production function, for which factors' returns are equal to parameters, is specified. The capital and labor inputs are respectively measured by bed-places and by a seasonally adjusted weighted average of annual employees. Differently from the usual, the output is expressed in value instead of in quantity units; the value added has been preferred to the physical output - nights - since nights do not represent a uniform share of hotel outputs, excluding all the others leisure services (catering, café, sports systems and other supplied services etc.).

In order to measure the technical (in)efficiency effects on the production function parameters, three specifications, corresponding to different economic and statistical hypotheses on productive system efficiency, have been evaluated. The general production equation is:

$$\begin{aligned} \ln y_n &= \beta_0 + \beta_k \ln k_n + \beta_l \ln l_n + v_n - u_n \\ v_n &\sim N(0, \sigma_v^2) \end{aligned} \tag{9}$$

where y_n, k_n, l_n are respectively output, capital and labor variables for the n-th firm. The hypotheses on output technical efficiency, expressed in terms of the inverse of the Debreu-Farrel ratio, e_n, are summarized in Table 1.

In model I all the firms are assumed to operate efficiently. Deviations from the frontier are attributed to the stochastic shock which is explained through a Gaussian component. In consequence, the classical regression model, implicit in a wide part of literature, is obtained.

In the following two specifications (model II, IIb), (in)efficiency among firms is admitted by subtracting a truncated normal stochastic component from the previous classical regression model.

The last specification (model III) is obtained through the semiparametric approach; in the first step the inefficient observations are eliminated by an FDH filter. A productive unit is considered efficient if it has an efficiency score greater than 0.95, that is

$$e_n = \frac{y_n}{y_n^\partial} \geq 0.95.$$

In the second step, the parametric model (9) is estimated on the filtered efficient firms.

The estimates of the different specifications described above for the production frontier, presented in Table 2, show interesting structural characteristics of the Italian accommodation industry. In particular, according to model I, the hotel system turns out to be characterized by increasing returns to scale. The labor elasticity is greater than the capital one so that the Italian hotel industry results in a high labor-intensity productive structure. In model II both labor and scale returns are lower than in model I; confidence intervals for all coefficients, except the intercept, are greatly overlapping, meaning that

Table 1. Hypotheses on models I, II, IIb and III

Model	Economic Hp.	Statistical Hp.
I	$e_n = 1$	$u_n = 0$
II	$e_n \leq 1$	$u_n \sim N^+(\mu_1, \sigma_v^2)$
IIb	$e_n \leq 1$	$u_n \sim N^+(\mu_n, \sigma_v^2)$ $\mu_n = \delta' \mathbf{z}_n$
III	$e_n \leq 1$	$y_n = [f(\mathbf{x}_n) + v_n] I_{[y_n \geq 0.95 y_n^\partial]}$

Table 2. Estimates of models I, II, IIb and III

Model	I		II		II b*		III	
	coeff.	c.i. 95%	coeff.	c.i. 95%	Coeff.	c.i. 95%	coeff.	c.i. 95%
β_0	8.506	[8.350,8.661]	9.293	[9.162,9.423]	9.499	[9.364;9.634]	9.705	[9.554;9.856]
β_k	0.607	[0.560;0.655]	0.571	[0.534;0.608]	0.509	[0.472;0.547]	0.597	[0.550;0.644]
β_l	0.738	[0.702;0.774]	0.708	[0.680;0.737]	0.701	[0.673;0.728]	0.529	[0.494;0.564]
Scale	1.345	[1.314;1.376]	1.280	[1.253;1.307]	1.200	[1.183;1.237]	1.126	[1.098;1.154]
μ			−3.399	[−4.070;−2.728]	−9.796	[−14.301;−5.290]		
δ_1					−2.400	[−3.144;−1.656]		
δ_2					−1.314	[−2.025;−0.604]		
δ_3					0.614	[0.317;0.911]		
δ_4					−11.895	[−15.971;−7.820]		
σ^2_V	0.611		0.210		0.219		0.154	
σ^2_u			2.888		9.623			

* Explanatory factors for inefficiency are respectively dummies for presence of café and catering (z_1), convention room (z_2), sporting (z_3) and reservation systems (z_4).

the two models provide descriptions of productive structures not significantly different from each other.

A quite different structure is stylized by semiparametric estimates: labor return is measured significantly smaller than in the other models, while capital elasticity does not substantially change ($\simeq 0.6$). The results indicate that the technically efficient firms implement productive processes different from the other firms.

Therefore, there is reason to believe that, in stochastic frontier framework, it would not be possible to detect non-uniform effects of technical inefficiency on productive mix.

To take into account for non-neutral (in)efficiency, some different explanatory variables of the efficiency (presence of café and catering, convention room, sports and reservation systems) have been introduced in the specification IIb. Although the (in)efficiency specification (7) is significant, the estimates of production function parameters are remarkably similar to those of model II confirming that the stochastic frontier model is not adequate to approximate the productive frontier. This could happen because the way in which frontier shifts is a function of the specific introduced factors and, even more generally, because stochastic frontier is a conditional expectation estimator instead of the maximal value estimator which the frontier analysis requests.

Therefore, the problem can be restated in a more coherent manner by means of the quantile regression estimator, because describing the maximum output y attainable from a given vector of input $\mathbf{x}$ is the same as estimating the technological equation on the observations in the highest percentiles of the conditional distribution of $y|\mathbf{x}$, that is quantile regression estimation.

4. Alternative approach: Quantile estimation

Quantile regression provides a description of a response variable as a conditional function of a set of covariates, broader than the methods based on conditional means (ordinary least square or maximum likelihood) do. The interesting feature consists of extending the analysis from mean or median values to the full range of other conditional quantile functions, providing an analytical description of an ordered set of technological relationships corresponding to different levels of efficiency.

Given a real valued random variable y and a set of covariates $\mathbf{x}$, the α-quantile function ($\alpha \in [0, 1]$) is the linear conditional quantile function

$$Q_\alpha(y|\mathbf{x} = \tilde{\mathbf{x}}) = \tilde{\mathbf{x}}'\beta_{(\alpha)}$$

which can be estimated by solving ([8], [10])

$$\beta_{(\alpha)} = \arg\min_{\beta} \sum_{n=1}^{N} |\alpha - I_{(y_n - x_n'\beta < 0)}|(y_n - x_n'\beta). \tag{10}$$

The α-quantile estimator measures the relationship among an input variables vector, $\mathbf{x}$, and the output, y, on the α quantile of conditional distribution $y_{(\alpha)}|\mathbf{x}$. Moreover quantile regression estimator is robust to deviations from distributional hypotheses which is an appealing characteristic in the

Table 3. Estimates of models IV, V, VI

Model	IV		V		VI	
	coeff.	c.i. 95%	coeff.	c.i. 95%	coeff.	c.i. 95%
β_0	8.888	[8.749;9.047]	9.798	[9.657;9.981]	10.100	[9.905;10.334]
β_k	0.509	[0.458;0.551]	0.499	[0.450;0.547]	0.560	[0.489;0.619]
β_l	0.794	[0.759;0.832]	0.695	[0.661,0.721]	0.609	[0.547;0.643]
Scale	1.303	[1.283;1.322]	1.194	[1.161;1.226]	1.169	[1.104;1.232]

production function context because of the asymmetric distribution of stochastic error.

In the frontier production function framework, quantile regression allows valuable new information; in fact, it provides a whole spectrum of production functions corresponding to different quantiles of conditional distribution of output given inputs, as closer to the frontier as α increases.

Table 3 shows the estimates of the production function for $\alpha = 0.500$, $\alpha = 0.900$, $\alpha = 0.975$ quantiles (respectively models IV, V and VI). The obtained quantile regression results provide new information for economic analysis of Italian hotel industry.

In the conditional median relation (Model IV) as in expected values models (Model I-II), labor elasticity considerably exceeds the capital one and the scale parameter is still about equal to 1.3. As the production function approaches the frontier, the positive spread between labor and capital elasticities diminishes and it becomes negative very close to the frontier. The returns to scale are just greater than constant.

In a greater detail, Fig. 1 points out the non-neutral effect of the (in) efficiency on the inputs mix that does not change uniformly as the frontier is approached. In fact, as α increases, that is as the technical efficiency increases, the labor productivity remains substantially constant up to $\alpha = 0.700$, then it begins to decrease, gradually up to $\alpha = 0.900$ quantile, suddenly for $\alpha > 0.900$ quantiles. The capital productivity reveals a quasi-opposite behavior: it remains substantially constant up to $\alpha = 0.900$ quantile and it increases over quantiles higher than 0.900. Thus, labor return overcomes capital return for the hotels operating on lower quantiles, which are the most part. On the other hand, capital return is equal or even greater than labor return for the most technically efficient hotels (on the highest quantiles).

Figure 2 depicts what happens to scale returns: as efficiency increases, scale returns diminish. The classic as well as the stochastic production function models overestimate touristic scale returns, and there is evidence that the estimated increasing returns to scale could be due to a mispecification of the model for the technical inefficiency.

In conclusion, while in the conventional frontier framework the choice of the estimation method significantly affects economic analysis, the quantile estimation allows us to distinguish among technological relations at different efficiency levels. Furthermore quantile regression provides valuable new information to verify performance and suitability of the conventional techniques in the frontier production function estimation literature.

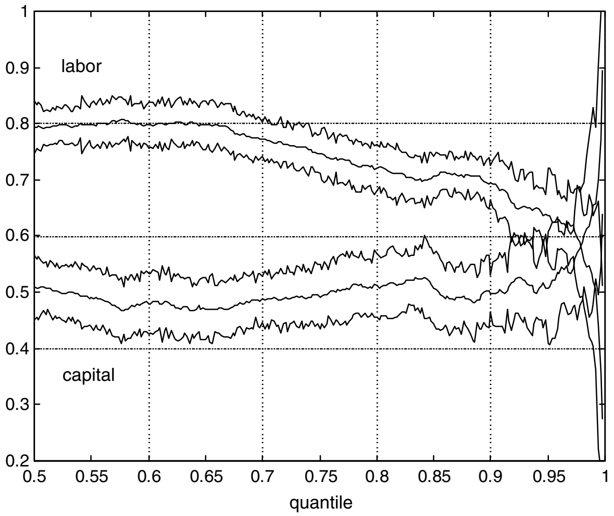

Fig. 1. Factors returns (with confidence intervals)

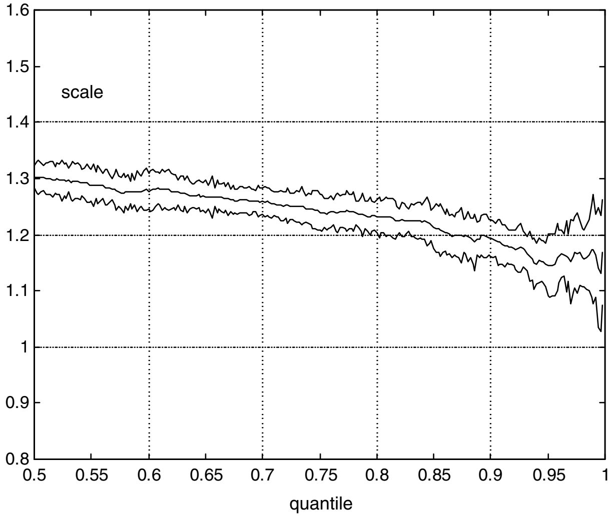

Fig. 2. Scale returns (with confidence intervals)

5. Conclusions

This paper is concerned with the methodological question of frontier production functions estimation. Empirical implications of conventional techniques are evaluated using data from the Italian hotel industry.

The discrepancies between the stochastic frontier and semiparametric models' estimates cast doubt on the ability of the expected value estimators in describing the frontier behavior. For the purpose of frontier production function estimation, we therefore consider the problem in terms of quantile estimation, with very interesting results.

Quantile estimation shows significant biases in the interpretations of the relevant economic parameters of the frontier production analysis (factors and scale returns) by conventional techniques. Moreover it allows the analysis for different levels of efficiency providing the whole spectrum of production functions. These findings are useful for broader interpretation of the empirical context of heterogenous technologies.

References

1. Aigner DJ, Lovell CAK, Schmidt P (1977) Formulation and estimation of stochastic frontier production function model. Journal of Econometrics 6:21–37
2. Battese GE, Coelli T (1995) A model for technical inefficiency effects in a stochastic frontier production function for panel data. Empirical Economics 20:325–332
3. Charnes A, Cooper WW, Rhodes E (1978) Measuring the efficiency of decision making units. European Journal of Operational Research 2:429–444
4. Deprins D, Simar L, Tulkens H (1984) Measuring labor inefficiency in post offices. In: Marchand M, Pestieau P, Tulkens H (eds) The performance of public enterprises: Concepts and measurements. North Holland, Amsterdam, pp 243–267
5. Freo M (1999) Analisi della struttura produttiva del settore manifatturiero attraverso la metodologia delle frontiere di produzione. PhD Thesis, University of Bologna , Bologna
6. Bernini C, Gardini A, Guizzardi A (1998) Il sistema congressuale italiano: caratteri tipologici e distribuzione territoriale. In: Presidenza del Consiglio dei Ministri (eds) Ottavo rapporto sul turismo italiano, Mercury, Firenze, pp 357–380
7. Huang CJ, Liu JT (1994) Estimation of a non-neutral stochastic frontier production function. Journal of Productivity Analysis 5:171-180
8. Koenker R, Bassett G (1978) Regression quantiles. Econometrica 46:33—50
9. Meeusen W, van de Broeck J (1977) Efficiency estimation from Cobb-Douglas production functions with composed errors. International Economic Review 18:435–444
10. Pagan A, Ullah A (1999) Nonparametric econometrics. Cambridge University Press, Cambridge
11. Simar L (1992) Estimating efficiencies from frontier models with panel data: a comparison of parametric, nonparametric and semiparametric methods with bootstrapping. Journal of Productivity Analysis 3:167–203
12. Thiry B, Tulkens H (1992) Allowing for inefficiency in parametric estimation of production functions for urban transit firms. Journal of Productivity Analysis 3:45–65