Three Approaches to the Analysis of Cost Function in Health Care

*Martin DLOUHÝ – Lenka FLUSSEROVÁ**

Abstract

In this paper, we describe three methods of cost-function analysis in health care: classical econometric analysis, frontier analysis, and survivor analysis. The strength of classical econometric analysis is a highly developed methodology of hypothesis testing; the weakness is that it is necessary to deal with many problems related to estimation technique, such as multicollinearity, autocorrelation, heteroscedasticity, etc. The strength of stochastic frontier analysis is that it incorporates random shocks in efficiency evaluation; on the other hand, strong assumptions about the distribution of efficiency have to be made. The advantages of survivor analysis are its simplicity and the possibility to include the factors that are otherwise hard to measure; the disadvantages are the application only in the long-term studies, and the provision of no specific information on the character of cost function in the studied industry. Different methods have strengths and weaknesses and the choice of the appropriate method depends on the objective of the study.

Keywords: *cost function, econometric analysis, frontier analysis, survivor analysis, health care*

JEL Classification: C10, D24, I10

1. Introduction

A cost function describes the relationship between output (total product) and cost; it shows the minimal cost of producing the given level of output. In health care, output can be measured in the number of outpatient consultations, the number of admissions, the number of inpatient days and the like. The total cost function

* Martin DLOUHÝ – Lenka FLUSSEROVÁ, Vysoká škola ekonomická v Praze, Katedra ekonometrie, nám. W. Churchilla 4, 130 67 Praha 3, Česká republika; e-mail:dlouhy@vse.cz

***Acknowledgement**:* The paper was supported by the Grant Agency of the Czech Republic, project No. 402/06/0150 *Quantitative Models for the Analysis of Economic Efficiency on Imperfect Markets.*

can be written as $TC = c(y)$, where y is the level of output. The cost of production can be also related to inputs (the number of beds, physicians or nurses) and prices of these inputs. In this case, the total cost function is expressed as

$$TC = p_1 x_1 + p_2 x_2 + ... + p_k x_k \tag{1}$$

where x_j (j = 1, 2,…, k) are the quantities of inputs (factors of production) and p_j (j = 1, 2,…, k) are the prices of these inputs. In empirical studies, researchers also relate cost with external factors, such as the location of hospital, the type of ownership, or the chain affiliation. Estimation of the cost function offers: (1) to determine marginal cost, (2) to explore the existence of scale economies, (3) to explore the existence of economies of scope, (4) to evaluate the cost efficiency of production unit.

Marginal cost is the cost related to production of the additional unit of output. The marginal cost is the first derivative of the total cost function. *Economies of scale* describe the situation in which the long-term average cost (*AC*) of production declines with increasing output *y*. The opposite of economies of scale are *diseconomies of scale*, which describe situation in which $AC(y) < AC(y + d)$; $d > 0$. This rise of average cost is explained by inefficiencies in the management of large organizations. *Economies of scope* are possible only for the multi-product case. Hospitals that produce both inpatient and outpatient care have a choice whether to provide a particular service on the outpatient or inpatient basis. According to economies of scope, it is more efficient to produce two products together than separately. It looks obvious in theory, but it is less obvious in practice. Concentration of services in health centers and hospitals may lead to excessive utilization of accessible services. If this is the case, the average costs of services are lower, but the average cost per patient could actually increase. Thus, under some circumstances, it may be a really difficult question whether the concentration of services means a higher quality or just more expensive medicine. *Evaluation of cost efficiency*: technical efficiency means that a firm does the best possible job, without any waste. If inputs are substitutes, different combinations of inputs can produce the same output. In this case, a manager should choose the least costly alternative of production. There is a direct link between a production function and a cost function: technical efficiency is a necessary condition for cost efficiency.

2. Classical Econometric Analysis

The cost function is correctly specified if all important variables are included and if the right functional form is chosen. The strength of econometric analysis is a highly developed methodology of hypothesis testing – both the selection of variables and the selection of functional form can be tested (Gujarati, 2003). The cost function is usually specified as a multiple regression equation, with the dependent variable being either the total cost or the average cost. In many studies, the recurrent cost is only used because of problems with the capital cost, which is measured in historical prices. Depending on the objective of the study and the data availability, the empirical econometric studies include a great variety of explanatory variables. It is possible to divide these explanatory variables into three main categories:

a) Input-related explanatory variables

For example, the number of departments, the number of beds, the number of physicians, the number of nurses, the number of outpatient departments.

b) Output-related explanatory variables

For example, the number of outpatient consultations, the number of admissions (discharges), the number of inpatient days, the case mix (diagnosis, age, race), the indicators of quality of services (hospital mortality, readmissions).

c) Other explanatory variables

For example, the type of ownership (private or public), the for-profit or non-for-profit status, the market competitiveness, the affiliation with a multi-hospital system, the affiliation with a medical school, location (urban or rural), and so forth.

In hospital economics, the relation between the number of beds and cost is especially studied. This is probably so for two reasons, first, the relation between the hospital size and the cost is a key relation of hospital economics indicating scale of activity, and second, the data on hospital beds are usually the best available data on hospitals at all. The initial, single-equation model of hospital cost is written as

$$\textit{Average Cost} = \beta_0 + \beta_1(\textit{Number of Beds}) + \beta_2(\textit{Number of Beds})^2 + u \qquad (2)$$

A nonlinear relationship in the second cost function (2), which was introduced by the square of the number of beds, enables to study economies of scale. If larger hospitals are able to achieve scale economies, the average cost curve in relation to scale of activity is *L*-shaped. An alternative assumption is that, at some point, the average cost begins to grow due to the inefficient management control over a large organization. Such average cost curve is *U*-shaped and implicitly assumes that there exists an optimum size of a hospital. However, it seems that the number of beds is not as good indicator of hospital size and cost as it was in the past. With the development of modern (high-cost) medical technology, the outpatient services are now a far more important part of a modern hospital than decades ago.

In order to assess the quality of methods and results of the studies of hospital cost functions, Posnett (Posnett, 2002) suggests that three important considerations are: (a) an appropriate unit of measurement, (b) adjustment for case mix, and (c) adjustment for input prices.

• *Unit of measurement*. As a measure of efficiency, cost per case is superior to cost per day. A hospital that improved its efficiency by reducing the average length of stay may have higher average cost per day because costs are typically higher in the first few days after admission.

• *Adjustment for case mix*. A difference in case mix is one of the most obvious determinants in cost per case. The studies that do not adequately adjust for differences in resource intensity between hospitals are difficult to interpret, especially in relation to economies of scale. If larger hospitals attract a more resource-intensive case mix, unit costs may be higher even in the presence of economies of scale.

• *Adjustment for input prices*. Costs are a function of the input mix and the price of individual inputs. If differences in the costs of inputs faced by a hospital are not adjusted for, they may confound any true underlying relationships between size and efficiency or other variables.

3. Frontier Analysis

If there is no recognition of statistical error, the level of efficiency is calculated as the distance between the observation and the regression line. If the residual is positive, the unit is relatively efficient, and vice versa, if the residual is negative, the production unit is relatively inefficient. The shortcoming of such approach is that it concentrates on the estimation of average behavior, not on the best performance. The classical regression analysis was, therefore, extended to the *deterministic frontier analysis* and *stochastic frontier analysis* (Kumbhakar and Lovell, 2000). Both methods can be applied to production and cost functions. Estimation of the input-oriented cost efficiency is even a more interesting case than estimation of the output-oriented technical efficiency because cost efficiency can be decomposed into input-oriented technical inefficiency and input allocative inefficiency. Because technical efficiency is a necessary condition for cost efficiency, a cost efficient unit has to be technically efficient.

Let us define that E_i is the total expenditure of production unit i, y_i is the vector of outputs produced by unit i, x_i is the vector of inputs, w_i is the vector of input prices faced by unit i, and β is the vector of unknown technology parameters to be estimated. Let us suppose that outputs and inputs are nonnegative and prices are positive. The total expenditure of ith unit is $E_i = w_i^T x_i$. The cost frontier $c(y_i, w_i;\ \beta)$ is common for all production units. The cost efficiency of ith unit, CE_i, may be expressed as the ratio of minimum feasible cost to expenditure:

$$CE_i = \frac{c(y_i,\ w_i;\ \beta)}{E_i} \tag{3}$$

For a cost efficient unit, the observed expenditure E_i equals to the minimum feasible cost $c(y_i,w_i;\beta)$, and therefore the cost efficiency $CE_i = 1$. If the observed expenditure E_i is higher than the minimum feasible cost, the production unit is not cost efficient and $CE_i < 1$. This formulation (3) is *a deterministic cost frontier*, which ignores random shocks and attributes the higher expenditure of the unit to cost inefficiency. Notice that cost efficiency of the unit can be estimated without observing the input vector x_i.

A *stochastic cost frontier* is formulated as $[c(y_i, w_i; \beta) \exp\{v_i\}]$, where $c(y_i, w_i; \beta)$ is the deterministic part and $\exp\{v_i\}$ is the unit-specific stochastic part of the frontier. The input-oriented cost efficiency is then given by the ratio

$$CE_i = \frac{c(y_i, w_i; \beta) \exp\{v_i\}}{E_i} \tag{4}$$

The composed error term ε_i in the stochastic cost frontier model is defined as $v_i + u_i$, where v_i is the two-sided random-noise component, and u_i is the nonnegative cost inefficiency component. The composed error ε_i is asymmetric and positively skewed because $u_i \geq 0$. The cost frontier must be linearly homogeneous in input prices: $c(y_i, \lambda w_i; \beta) = \lambda c(y_i, w_i; \beta)$ for $\lambda > 0$. One solution is the restriction that the sum of the technology parameters β_j equals one, or another solution is that the cost frontier model is reformulated. Let us assume that the stochastic cost frontier takes the Cobb-Douglas functional form. The reformulated stochastic cost frontier is then written as

$$\ln\left(\frac{E_i}{w_{ki}}\right) = \ln \beta_0 + \beta_y \ln y_i + \sum_{j=1}^{k-1} \beta_j \left(\frac{w_{ji}}{w_{ki}}\right) + v_i + u_i \tag{5}$$

where w_{ji} is the price of *j*th input faced by *i*th unit, and *k* is the number of inputs. The first part of equation (5) measures the relation between the expenditure E_i and output y_i, and the second part of equation (5) measures the relation between the expenditures and input prices faced by unit *i*. The measure of cost efficiency for the *i*th production unit is calculated as

$$CE_i = \exp\{-u_i\} \tag{6}$$

The estimates of cost efficiency can be obtained by the mean or the mode point estimators $E(u_i|\varepsilon_i)$ and $M(u_i|\varepsilon_i)$. They are given by

$$E(u_i|\varepsilon_i) = -\sigma_* \left[\frac{\varphi(\varepsilon_i\lambda/\sigma)}{1-\Phi(\varepsilon_i\lambda/\sigma)} - \left(\frac{\varepsilon_i\lambda}{\sigma}\right)\right] \tag{7}$$

and by

$$M(u_i|\varepsilon_i) = \begin{cases} -\varepsilon_i \left(\dfrac{\sigma_u^2}{\sigma^2} \right) & if\ \varepsilon_i \le 0, \\ 0 & otherwise \end{cases} \tag{8}$$

where ε_i is the composed error term, σ_v^2 and σ_u^2 are distribution parameters of v_i and u_i, $\Phi(\cdot)$ is the cumulative density function of standard normal distribution, $\varphi(\cdot)$ is the density function of standard normal distribution, $\sigma = (\sigma_v^2 + \sigma_u^2)^{1/2}$, $\lambda = \sigma_u/\sigma_v$, and $\sigma_* = (\sigma_v^2 \sigma_u^2)/\sigma^2$. When the point estimates of u_i are obtained, the cost efficiency of each producer (6) is estimated by $exp\{-u_i\}$.

The main advantage of stochastic frontier analysis is that it is able to incorporate random shocks in efficiency evaluation. On the other hand, strong assumptions about the distribution of efficiency have to be made (e.g., exponential distribution of cost efficiency is assumed).

4. Survivor Analysis

Several complicated issues must be considered by a researcher in estimating cost functions—measuring of case mix, handling difference in quality, and distinguishing between the short-run and long-run costs. A *survivor analysis*, developed by Stigler (1958), is recommended by health economics textbooks as a simple alternative to estimating cost functions (Fedstein, 1998; Folland, Goodman and Stano, 2001). The idea of the method is straightforward: those categories that grow relative to the rest of the industry are assumed to have some advantage over the other ones. In the long-run, the distribution of providers should tend toward an optimum, which is, by the analysis, identified as the category (-ies) with the fastest growth. Categories may be defined by the size of hospital or of group practice (when estimating economies of scale), by the specialty, by the type of ownership, by location, and so forth. For example, if scale economies exist then, in the long-term, hospitals that are too small or too large will exit the market and there will be only those of the optimum size.

An advantage of classical, univariate survivor analysis is that the method includes both the factor to be investigated and all other factors. The analysis thus includes factors that are hard to measure in econometric studies of cost function. On the other hand, a limitation of the survivor analysis is that it is not able to isolate the effects of those factors. This limitation can be moderated by taking an explicit account of such factors in an expanded, multivariate survivor analysis. The linear version of the multivariate survivor analysis takes the form

$$s_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki} + u_i, \tag{9}$$

where s_i is the change in market share of group i, and x_{1i}, x_{2i}, ..., x_{ki} are the explanatory variables (factors). An alternative is a binary growth model, but in this type of model, the information is lost in converting a continuous variable into a binary one.

The univariate survivor analysis has one advantage over other methods: it is its simplicity. But both the univariate and multivariate survivor analyses, like the other methods of cost analysis, are not able to overcome the fact that the governments, not the market forces play a significant role in the health sector. Changes in the structure of the health care market may rather demonstrate planned governmental interventions than a real economic struggle of providers for their survival. However, the survivor analysis of governmental policy seems to be also an interesting application.

According to Koutsoyiannis (1979), the survivor analysis suffers from serious limitations. The survivor analysis assumes that: (a) the firms pursue the same objectives (hospitals are both private and public); (b) the firms operate in similar environments so that they do not have locational or other ad-vantages (e.g., public hospitals receive subsidies); (c) prices of factors and technology are not changing (e.g., prices of health technology grow rapidly); (d) the firms operate in a very competitive market structure, that is, there are no barriers to entry or collusive agreements, since under such conditions inefficient (high-cost) firms would probably survive for long periods of time (hospitals are local monopolies, the entry is regulated). Another shortcoming of the survival analysis is that it is not able to explain cases where the size distribution of firms remains constant over time. If the share of the various plant sizes does not change over time, this does not imply that all scales of plant are equally efficient. Koutsoyiannis (1979) also argues that the survivor analysis indicates only the broad shape of the long-run cost curve, but it does not show the actual magnitude of economies of scale.

5. Applications

Although the methodology of cost function can be applied to a variety of production units in health care (physician, department, hospital, community health center, health system), the health care applications are dominated by the studies where the production unit is a hospital. The analysis of cost function with methods of quantitative economic analysis are relatively rare in the Czech health system, therefore, we also present illustrative examples from other countries.

In their study on hospital efficiency, Dlouhý and Strnad (1999) estimated hospital total cost function on a sample of 40 Czech acute-care hospitals, the 1997 data. The independent variables were: the number of beds, the number of beds in intensive care units, and the number of outpatient departments. All parameters were statistically significant. As the model assumed the constant average cost, the existence of economies of scale could not be tested, which was a weakness of the study.

Frech and Mobley (1995) tried to resolve the problem of hospital economies of scale by the multivariate survival analysis. They analyzed hospital data from California for years 1983 and 1989. In the continuous version of their survival model, the dependent variable was defined as a change in market share, measured in inpatient days, for a given group of hospitals. The explanatory variables included the level of output (average daily census), the chain affiliation, the Herfindahl index of market concentration, the change in local market-level proportion of hospital revenues under discount contracts, and some other variables. The study found the existence of scale economies possibly up to a size as large as 220 bed-days (370 beds).

Over the 1981 – 1982 through 1991 – 1992 periods, Bilodeau, Crémieux et al. (2004) studied all short-term hospitals in the Province of Québec, Canada. They used two-step analysis: first, they estimated technical efficiency; second, the costs of technical inefficiency were calculated. Bilodeau et al. found that significant inefficiencies up to 17% could have been saved through improved performance. Post-estimation analysis that included quantitative measures of care suggested that differences in performance were attributable to differences in management and unobservable quality of care rather than patient case mix. The interesting result was that the 20 worst performing hospitals out of 119 over the period studied had accounted for nearly 50% of total inefficiencies.

Mark (1996) empirically examined whether an ownership affects the quality of psychiatric care. He estimated a quality deviation function and a frontier cost function using data on psychiatric hospitals in California for the years 1984 – 1989. The estimation of parameters was obtained by two-stage least squares. The analysis found the evidence that nonprofit psychiatric hospitals experienced fewer complaints and violation than for-profit psychiatric hospitals. There was no evidence that nonprofit hospitals were more inefficient than for-profit hospitals. Those findings supported the view that describes nonprofit hospitals as offering advantages in markets characterized by the asymmetric information.

Posnett (2002) reviewed the literature on economies of scale and found that larger hospitals did not have to have necessarily lower average costs and better outcomes. The research literature suggests that if economies on scale exist, they appear to be fully exploited at a relatively low scale, somewhere in the range of 100 – 200 beds. This size is likely a minimum size of acute general hospital that is able to provide all complementary medical, diagnostic and support services. For hospitals with 200 and more beds, we may observe a roughly constant average cost. Somewhere in the range of 300 – 600 beds the average cost may be expected to rise. It is important, however, to bear in mind that optimum hospital size is a direct function of the health care needs of the population that it is designed to serve.

Conclusion

In this paper, we described three methods of cost-function analysis in health care: classical econometric analysis, frontier analysis, and survivor analysis. The strength of classical econometric analysis is a highly developed methodology of hypothesis testing – both the selection of variables and the selection of functional form can be tested. The weakness of econometric analysis is that it is necessary to deal with many problems related to estimation technique, such as multicollinearity, autocorrelation, heteroscedasticity, etc. The strength of stochastic frontier analysis is that it incorporates random shocks in efficiency evaluation. On the other hand, strong assumptions about the distribution of efficiency have to be made. The advantages of survivor analysis are the simplicity and the possibility to include the factors that are otherwise hard to measure. The disadvantages of survivor analysis are: the method can be applied only in the long-term studies (not applicable to cross-sectional data); it does not provide specific information on the character of cost function in the studied industry. Hence results of survivor analysis are not controlled for possible confounding factors, such as case mix and so on. No method of cost analysis is able to overcome the fact that the governments, not the market forces play a significant role in the health sector. Changes in the structure of the health care market may rather demonstrate governmental interventions than a struggle of health providers for their economic survival.

Health care is an application area with very specific context and characteristics. But this complexity is the reason why health care as an application area is so challenging for researchers. The main topic of this study was showing how quantitative economic analysis of cost function could be useful in evaluating efficiency of resources in the health system. Different approaches have strengths and weaknesses and the choice of the appropriate method depends, of course, on the objective of the study.

References

[1] BILODEAU, D. – CRÉMIEUX, P.-Y. – JAUMARD, B. – OUELLETTE, P. – VOVOR, T. (2004): Measuring Hospital Performance in the Presence of Quasi-fixed Inputs: An Analysis of Québec Hospitals. Journal of Productivity Analysis, *21*, pp. 183 – 199.

[2] DLOUHÝ, M. – STRNAD, L. (1999): Nemocnice – kvalita, efektivita, finance [A Hospital: Quality, Efficiency, Financing]. Praha: Grantová agentura Ministerstva zdravotnictví ČR. <http://www.izpe.cz//files/vysledky/jinevysledky/02.pdf>.

[3] FELDSTEIN, P. J. (1998): Health Care Economics. Fifth Edition. Albany, NY: Delmar Publishers.

[4] FOLLAND, S. – GOODMAN, A. C. – STANO, M. (2001): The Economics of Health and Health Care. Third Edition. Upper Saddle River: Prentice Hall.

[5] FRECH, H. E. III – MOBLEY, L. R. (1995): Resolving the Impasse on Hospital Scale Economies: A New Approach. Applied Economics, *27*, pp. 286 – 296.

[6] GUJARATI, D. N. (2003): Basic Econometrics. Fourth Edition. New York: McGraw-Hill.

[7] KOUTSOYIANNIS, A. (1979): Modern Microeconomics. Second Edition. London and Basingstoke: The Macmillan Press.

[8] KUMBHAKAR, S. C. – LOVELL, C. A. K. (2000): Stochastic Frontier Analysis. Cambridge: Cambridge University Press.

[9] MARK, T. L. (1996): Psychiatric Hospital Ownership and Performance: Do Nonprofit Organizations offer Advantages in Market Characterized by Asymmetric Information? The Journal of Human Resources, *31*, pp. 631 – 649.

[10] POSNETT, J. (2002): Are Bigger Hospitals Better? In: M. McKee and J. Healy (eds.): Hospitals in Changing Europe. Buckingham: Open University Press.

[11] STIGLER, G. J. (1958): The Economics of Scale. The Journal of Law and Economics, *1*, 54 – 71.