IMA Journal of Management Mathematics (2005) **16**, 383–398
doi:10.1093/imaman/dpi017
Advance Access publication on April 29, 2005

Multilevel flexible specification of the production function in health economics

LUCA GRASSETTI†
Department of Statistical Sciences, University of Padova, 35121 Padova, Italy

ENRICO GORI‡
Department of Statistics, University of Udine, 33100 Udine, Italy

AND

SIMONA C. MINOTTI§
Department of Economics and Social Sciences, Catholic University of Piacenza, 29100 Piacenza, Italy

Previous studies on hospitals' efficiency often refer to quite restrictive functional forms for the technology (Aigner *et al.*, 1977, *J. Econom.*, **6**, 21–37) . In this paper, referring to a study about some hospitals in Lombardy, we formulate convenient correctives to a statistical model based on the translogarithmic function—the most widely used flexible functional form (Christensen *et al.*, 1973, *Rev. Econ. Stat.*, **55**, 28–45). More specifically, in order to take into consideration the hierarchical structure of the data (as in Gori *et al.*, 2002, *Stat. Appl.*, **14**, 247–275), we propose a multilevel model, ignoring for the moment the one-side error specification, typical of stochastic frontier analysis (Aigner *et al.*, 1977, *J. Econom.*, **6**, 21–37). Given this simplification, however, we are easily able to take into account some typical econometric problems as, e.g. heteroscedasticity. The estimated production function can be used to identify the technical inefficiency of hospitals (as already seen in previous works), but also to draw some economic considerations about scale elasticity, scale efficiency and optimal resource allocation of the productive units. We will show, in fact, that for the translogarithmic specification it is possible to obtain the elasticity of the output (regarding an input) at hospital level as a weighted sum of elasticities at ward level. Analogous results can be achieved for scale elasticity, which measures how output changes in response to simultaneous inputs variation. In addition, referring to scale efficiency and to optimal resource allocation, we will consider the results of Ray (1998, *J. Prod. Anal.*, **11**, 183–194) to our context. The interpretation of the results is surely an interesting administrative instrument for decision makers in order to analyse the productive conditions of each hospital and its single wards and also to decide the preferable interventions.

Keywords: elasticity; multilevel models; production function analysis; scale efficiency; scale elasticity; translog function.

1. Introduction

Previous studies on hospitals' efficiency often refer to quite restrictive functional forms for the technology (Aigner *et al.*, 1997). In this paper, referring to a study about some hospitals in Lombardy, we

†Email: grassetti@stat.unipd.it
‡Email: e.gori@dss.uniud.it
§Email: simona.minotti@unicatt.it

formulate convenient correctives to a statistical model based on the translogarithmic function, which is the most widely used flexible functional form for economic functions (Christensen *et al.*, 1973; Kim, 1992; Grant, 1993; Ryan & Wales, 2000) and obtains also closed form measures of scale elasticity and scale efficiency, readily computable from the fitted model (Ray, 1998). The aim of this work is, indeed, to provide an administrative instrument, based on stochastic production function analysis, which is able both to identify non-standard productive conditions and to propose convenient correctives, in the sense of inputs re-allocation.

The translogarithmic specification is the second-order Taylor approximation of a generic production function. In the simple case of one output (y) and two input variables (x_1, x_2), it is equal to

$$\ln(y) = \alpha_0 + \alpha_1 \ln(x_1) + \alpha_2 \ln(x_2) + \frac{\beta_{11}}{2}[\ln(x_1)]^2 + \beta_{12} \ln(x_1) \ln(x_2) + \frac{\beta_{22}}{2}[\ln(x_2)]^2. \qquad (1)$$

The success of this functional form in many econometric applications is due to its flexibility (e.g. the elasticity of the output with respect to an input is not constant, as for the Cobb–Douglas, but it is a function of the inputs). This flexibility allows a large adaptability of the model, but, at the same time, increases the multicollinearity.

In the present case of study, the choice of the translogarithmic specification is mainly connected to the non-linearity of the relation between beds number (one of the input variables) and the hospitalizations number (the output variable). In addition, it allows to make some economic considerations about scale elasticity, scale efficiency and optimal resource's allocation of the productive units (Ray, 1998).

In order to take into consideration the hierarchical structure of the data, common in health context (Leyland & Goldstein, 2001), we propose a multilevel model, ignoring for the moment the one-side error specification, common in stochastic frontier analysis (Aigner *et al.*, 1997). Given this simplification, however, we are able to take into account some typical econometric problems as e.g. heteroscedasticity.

The data involved in the study are, in fact, characterized by the presence of two levels of data collection. While the first one regards the single wards in each hospital, the second one is identified by the hospitals. These two levels characterize the natural hierarchical structure of the Health System in a microeconomic framework.

The paper is organized as follows. Section 2 describes the data involved in the study. Section 3 illustrates the model proposed. Section 4 presents the results of the analysis. Section 5 proposes some economic considerations on the results. Section 6 gives some conclusions.

2. The data

As illustrated in the previous section, the data involved in the study are characterized by a two-level hierarchical structure. More specifically, the considered data set refers to 1478 first-level units (wards) observed in 178 second-level units (hospitals) of the Italian region Lombardy in 1997. This data set, described in Gori *et al.* (2005), has been obtained from a deterministic linkage of three distinct archives: the discharge database, the beds allocation database and the staff database. These three archives are collected at different levels of detail. The discharge files are available for each hospitalization and, among the other variables, include the ward and the hospital identifiers. This information allows to calculate the total number of hospitalizations for each ward in a hospital (i.e. the total number of in-patients dismissed in the observed time period), which will be our output variable (indicated by NHosp). The beds number is available for each ward in a hospital (indicated by Beds). The staff variable (Staff) is available only at hospital level, without the possibility of distinguishing among substructures, and corresponds to the total number of doctors, nurses and administrative staff.

TABLE 1 *Descriptive statistics*

Variable	Min	Median	Max	Mean	Standard deviation
ln(NHosp)	0.000	6.903	9.806	6.589	1.947
ln(Beds)	0.000	3.296	6.234	3.166	0.926
ln(Staff)	2.833	6.497	8.322	6.442	1.147
Mean DRG weight (ward level)	0.160	0.850	8.320	1.076	0.776
s.d. DRG weight (ward level)	0.000	0.420	6.870	0.682	0.937
Mean DRG weight (hospital level)	0.630	0.850	2.130	0.886	0.030
s.d. DRG weight (hospital level)	0.120	0.670	1.950	0.695	0.093
std[ln(CMix)]	−10.350	−0.023	2.935	0.000	1

Then, some other information is taken into consideration. First of all, in order to distinguish the hospitalizations by degree of complexity, we consider the case-mix index, which represents the relative level of case complexity in each ward, calculated with respect to the regional mean value. In particular, we use a standardized version of the logarithm of this index (indicated by std[ln(CMix)]). This variable allows to distinguish between wards with complexity measure under and over the average, the latter probably characterized by a structural lower number of hospitalizations. This variable can be considered as a modifier of the effect of other input variables as we will see in the following paragraphs. Then, in order to take into account the type of hospitalizations disease, an addictive linear component is introduced in the production function model, consisting of four variables. These four variables are two transformations of the diagnosis related groups (DRG) weight variable, calculated at hospital and ward level. The DRG code derives from an international clusterization of the hospitalization cases; the DRG weight provides a 'proxy' measure of the complexity of every treated case. The first transformation is the mean of DRG weights (indicated by $\mu_{\mathrm{Weig},ij}$ and $\mu_{\mathrm{Weig},i}$) and represents the mean complexity of observed cases at ward and hospital level. The second one is the standard deviation of DRG weights ($\sigma_{\mathrm{Weig},ij}$ and $\sigma_{\mathrm{Weig},i}$) and allows to control the variability of the complexity in treated cases at the two levels considered. These factors are introduced to affect only the level of production and not the production process itself.

Table 1 summarizes the descriptive statistics of the variables involved in the model.

The pairwise scatter plot in Fig. 1 shows the relation between the output variable and the input variables and in particular highlights the strong dependence between hospitalizations number and beds number. From in-depth studies we conclude that this relationship cannot be considered linear.

3. The model

In order to model the number of hospitalizations we use a modified translog specification. In particular, the complete translog model for the hospitalizations number, function of the beds number and the staff of observed units is modified by the addition of two other components. The first one is a multiplicative term that linearly depends on case-mix index. The second one is the additive linear component illustrated in Section 2 (indicated in the following by $f(\mathrm{Weights})$).

Given the hierarchical structure of the data, we propose to adopt a multilevel model, with two levels and random intercept, which can be formulated as follows:

$$\ln(\mathrm{NHosp}_{ij}) = \alpha(\mathrm{CMix}_{ij}) + \beta(\mathrm{CMix}_{ij}, \mathrm{Beds}_{ij}) + \gamma(\mathrm{CMix}_{ij}, \mathrm{Staff}_i) + \delta(\mathrm{CMix}_{ij}, \mathrm{Beds}_{ij}, \mathrm{Staff}_i) + f(\mathrm{Weights}) + u_i + \epsilon_{ij}, \quad (2)$$

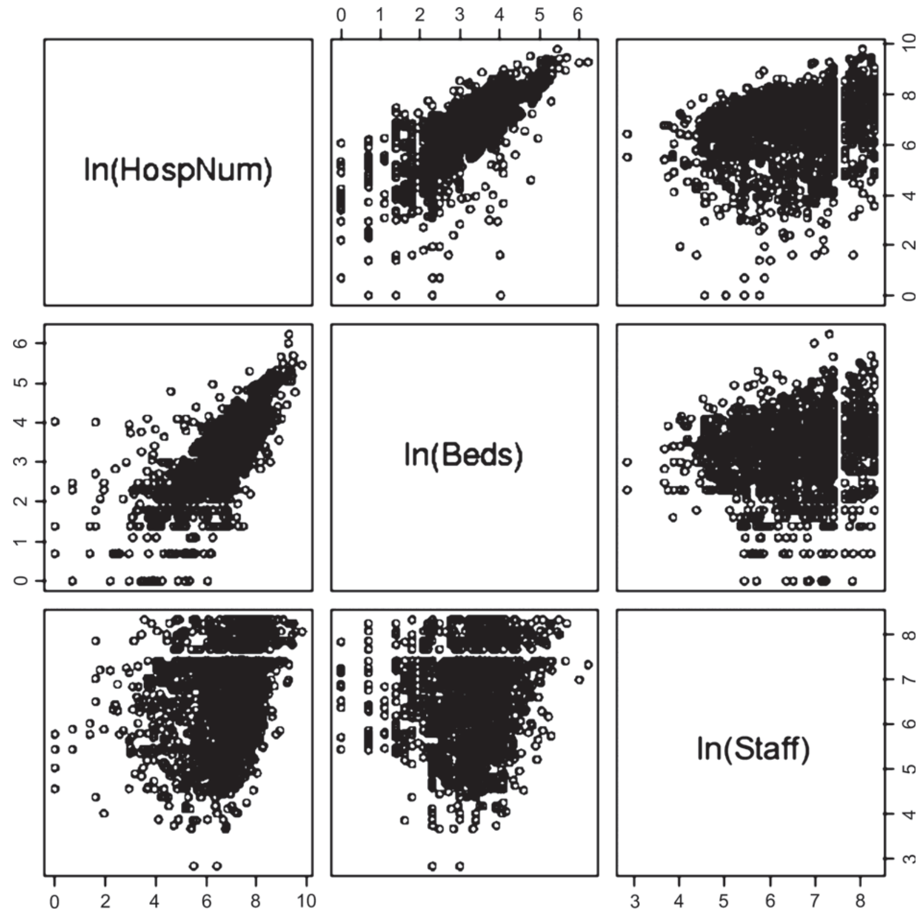

FIG. 1. Pairwise scatter plot among output variable and input variables.

where every single function is defined as

$$\begin{aligned}
\alpha(\mathrm{CMix}_{ij}) &= \alpha_0 + \alpha_1 \mathrm{std}[\ln(\mathrm{CMix}_{ij})],\\
\beta(\mathrm{CMix}_{ij}, \mathrm{Beds}_{ij}) &= [\beta_0 + \beta_1 \mathrm{std}[\ln(\mathrm{CMix}_{ij})]]\left[\ln(\mathrm{Beds}_{ij}) + \frac{\beta_2}{2}\ln(\mathrm{Beds}_{ij})^2\right],\\
\gamma(\mathrm{CMix}_{ij}, \mathrm{Staff}_i) &= [\gamma_0 + \gamma_1 \mathrm{std}[\ln(\mathrm{CMix}_{ij})]]\left[\ln(\mathrm{Staff}_i) + \frac{\gamma_2}{2}\ln(\mathrm{Staff}_i)^2\right],\\
\delta(\mathrm{CMix}_{ij}, \mathrm{Beds}_{ij}, \mathrm{Staff}_i) &= [\delta_0 + \delta_1 \mathrm{std}[\ln(\mathrm{CMix}_{ij})]][\ln(\mathrm{Beds}_{ij}) \ln(\mathrm{Staff}_i)],\\
f(\mathrm{Weights}) &= \lambda_1 \mu_{\mathrm{Weig},ij} + \lambda_2 \sigma_{\mathrm{Weig},ij} + \lambda_3 \mu_{\mathrm{Weig},i} + \lambda_4 \sigma_{\mathrm{Weig},i},
\end{aligned}$$

$u_i \sim N(0, \sigma_u^2)$ are the residuals at hospital level, $i = 1, \dots, N$, and $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$ are the residuals at ward level, $j = 1, \dots, M$, with $u_i \perp \epsilon_{ij}$.

Using a classical notation for the coefficients of mixed models, (2) can be re-written as

$$\begin{aligned}\ln(\mathrm{NHosp}_{ij}) = {} & \alpha_{0i} + \alpha_1 \ln(\mathrm{Beds}_{ij}) + \frac{\alpha_2}{2}[\ln(\mathrm{Beds}_{ij})^2] + \alpha_3 \ln(\mathrm{Staff}_i) + \frac{\alpha_4}{2}[\ln(\mathrm{Staff}_i)^2] \\ & + \alpha_5[\ln(\mathrm{Beds}_{ij}) \ln(\mathrm{Staff}_i)] + \alpha_6[\ln(\mathrm{Beds}_{ij}) \,\mathrm{std}\,[\ln(\mathrm{CMix}_{ij})]] \\ & + \alpha_7[\ln(\mathrm{Staff}_i) \,\mathrm{std}\,[\ln(\mathrm{CMix}_{ij})]] + \alpha_8[\ln(\mathrm{Beds}_{ij}) \ln(\mathrm{Staff}_i) \mathrm{std}\,[\ln(\mathrm{CMix}_{ij})]] \\ & + \alpha_9[\ln(\mathrm{Beds}_{ij})^2 \mathrm{std}\,[\ln(\mathrm{CMix}_{ij})]] + \alpha_{10}[\ln(\mathrm{Staff}_i)^2 \mathrm{std}\,[\ln(\mathrm{CMix}_{ij})]] \\ & + \alpha_{11} \mathrm{std}\,[\ln(\mathrm{CMix}_{ij})] + \lambda_1 \mu_{\mathrm{Weig},ij} + \lambda_2 \sigma_{\mathrm{Weig},ij} + \lambda_3 \mu_{\mathrm{Weig},i} + \lambda_4 \sigma_{\mathrm{Weig},i} + \epsilon_{ij}, \quad (3)\end{aligned}$$

where $\alpha_{0i} = \alpha_0 + u_i$, $\alpha_1 = \beta_0$, $\alpha_2 = \beta_0\beta_2$, etc.

Note that the model defined in (3) can be seen as an unconstrained version of model (2). Equation (3) is obtained by the substitution of $\alpha(\mathrm{CMix}_{ij})$, $\beta(\mathrm{CMix}_{ij})$, $\gamma(\mathrm{CMix}_{ij})$, $\delta(\mathrm{CMix}_{ij})$ and $f(\mathrm{Weights})$ in (2). This operation leads to a linear form with some restriction on parameter values, e.g. coefficients connected to ln(Beds) correspond to β_0, β_1, $\frac{\beta_0\beta_2}{2}$ and $\frac{\beta_1\beta_2}{2}$, that are only three different coefficients in model (2) and, respectively, α_1, α_6, α_9 and $\frac{\alpha_2}{2}$ in model (3). In order to justify the re-parameterization in (3), we have performed a hypothesis test about non-linear restrictions (Godfrey, 1988). The Wald test statistic presents a p-value of 0.0699, which allows to accept the null hypothesis. Given this result, the following analysis is based on model (3), without any constraints on the interaction parameters. The choice of formulation (3) is due not only to the ease in estimation by means of conventional statistical software, but also because it allows an immediate interpretability of the coefficients.

Then, regarding the errors at first level, ϵ_{ij}, usually called residuals, the estimation results, performed with the statistical software R (see Cribari-Neto & Zarkos, 1999; Ihaka & Gentleman, 1996), show a high level of heteroscedasticity (see Fig. 2(A)). In order to include this information in the model, we have defined a dependence of ϵ from one of the most significant inputs (the beds number), by means of an exponential multiplicative variance function. Then, the corresponding model assumes that the within-group errors are heteroscedastic, with variance function equal to

$$\mathrm{Var}(\epsilon_{ij}) = \sigma_\epsilon^2 |\mathrm{Beds}_{ij}|^{2\theta}. \quad (4)$$

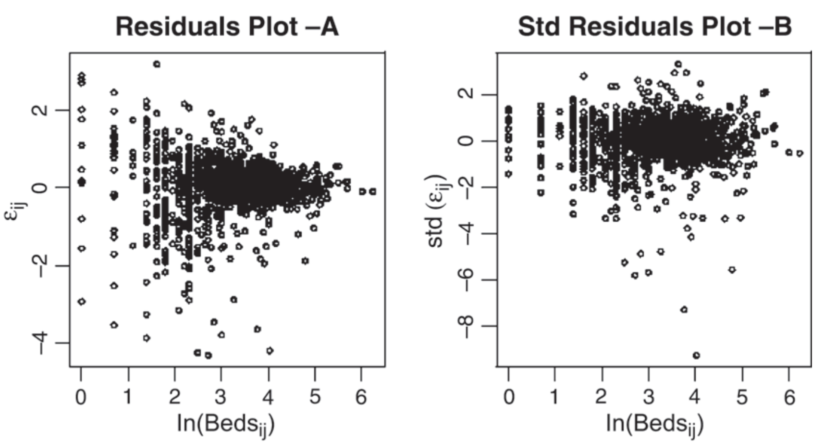

FIG. 2. Effect on the residuals of inclusion of heteroscedasticity in the model.

The plot of the standardized residuals (estimated by considering the inclusion of heteroscedasticity in the model) in Fig. 2(B) shows that the problem of heteroscedasticity has been reduced.

Finally, the second level errors, u_i, can be interpreted as efficiency indicators, as illustrated, e.g. in Gori *et al.* (2005), and can be estimated by means of the 'Empirical Bayes' approach as described in Pinheiro & Bates (2002) and Verbeke & Molemberghs (2000).

4. The results

In Table 2 we summarize the estimated coefficients for fixed and random effects of the model introduced in (3), which takes also into consideration the heteroscedasticity. For variance components, as Wald statistics based on the asymptotic standard error are not reliable, we provide the 95% confidence intervals (CIs). For the parameters in the linear mixed-effects model, approximate CIs are obtained by using the normal approximation of the distribution of the maximum likelihood (ML) estimators.

The prior aim of this paper is to focus on the economic interpretation of the estimated model. The obtained coefficients show that quadratic forms of original variables have negative effects on the number of hospitalizations as well as the complexity indexes (std[ln(CMix$_{ij}$)], $\mu_{\text{Weig},ij} \cdot \sigma_{\text{Weig},ij}$ and $\mu_{\text{Weig},i}$). Only $\sigma_{\text{Weig},i}$ presents a positive effect. A positive effect is, also, connected to the input variables considered in their original scale, also taking into consideration their interaction with the standardized case-mix index. Some coefficients are not significantly different from 0. In Table 2 we report also these least values in order to give a complete representation of the model specified in (3).

TABLE 2 *Estimated coefficients of the mixed model with heteroscedasticity*

	Value	Standard error	p-Value
Fixed effects coefficients			
α_0	0.7941	1.0367	0.4434
α_1	1.2328	0.1541	0.0000
α_2	−0.0165	0.0163	0.3101
α_3	0.9683	0.3033	0.0017
α_4	−0.0630	0.0251	0.0129
α_5	−0.0114	0.0236	0.6275
α_6	0.3265	0.1756	0.0632
α_7	0.9390	0.2315	0.0001
α_8	0.0185	0.0232	0.4252
α_9	−0.0443	0.0195	0.0232
α_{10}	−0.0774	0.0183	0.0000
α_{11}	−3.2004	0.8323	0.0001
λ_1	−0.4798	0.0493	0.0000
λ_2	−0.0591	0.0493	0.2305
λ_3	−1.2266	0.3833	0.0016
λ_4	0.6379	0.2689	0.0188
Random effects coefficients		Lower bound	Upper bound
θ	−0.3773	−0.4170	−0.3375
σ_u	0.3502	0.3026	0.4053
σ_ϵ	2.0625	1.8123	2.3474

Number of observations: 1478, number of groups: 178.

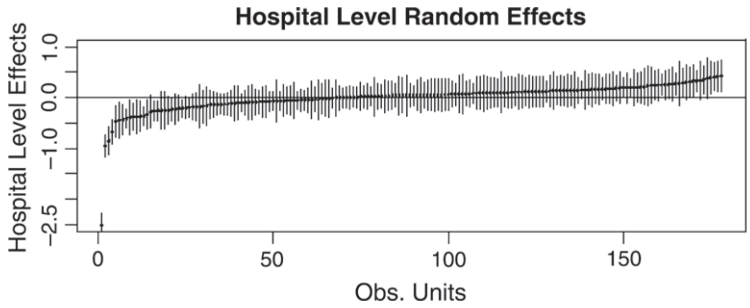

FIG. 3. CIs for estimated BLUP random effects (u_i) at hospital level.

In Fig. 3 we have summarized the estimated best linear unbiased predictor (BLUP) error components. CIs have been calculated on the basis of pairwise comparison theory given in Goldstein & Healy (1995). The estimated BLUP error components and their standard errors have been computed following Chapter 7 of Pinheiro & Bates (2002). The point estimates and their CIs are easily obtained from fixed effects and variance ML estimates.

5. Some economic considerations

As just pointed out, the aim of this article is to investigate the elasticity of the hospitalizations number with respect to the inputs of the model, in particular Staff and Beds, which are, indeed, the only variables directly under the control of hospitals' administration (Regional Agency for Health Care). As mentioned previously, case-mix is treated as an exogenous variable, in order to distinguish among different types of hospitalizations. Looking at our analysis results, in fact, it results that elasticity is strongly affected by this variable; we will show in the following how elasticity varies for its different levels.

The *elasticity* of the output y for an input x_r, defined as the marginal productivity of x_r, divided by the average productivity of x_r, i.e.

$$e(x_r) = \frac{\partial y}{\partial x_r} \Big/ \frac{y}{x_r} = \frac{\partial \ln(y)}{\partial \ln(x_r)}, \tag{5}$$

is the percentage change in output associated with a unitary percent change in the rth input, holding all other inputs constant, and represents a unit-free measure of the marginal productivity (Chambers, 1988).

From this formulation we obtain directly the elasticity at ward level, denoted by e^{W} (coefficient notations refer to (2)). For Beds we have:

$$\begin{aligned} e^{\mathrm{W}}(\mathrm{Beds})_{ij} &= \frac{\partial \ln(\mathrm{NHosp}_{ij})}{\partial \ln(\mathrm{Beds}_{ij})} \\ &= [\beta_0 + \beta_1 \mathrm{std}[\ln(\mathrm{CMix}_{ij})]][1 + \beta_2 \ln(\mathrm{Beds}_{ij})] \\ &\quad + [\delta_0 + \delta_1 \mathrm{std}[\ln(\mathrm{CMix}_{ij})]] \ln(\mathrm{Staff}_i), \end{aligned} \tag{6}$$

and analogously, for Staff:

$$\begin{aligned} e^{\mathrm{W}}(\mathrm{Staff})_{ij} &= \frac{\partial \ln(\mathrm{NHosp}_{ij})}{\partial \ln(\mathrm{Staff}_i)} \\ &= [\gamma_0 + \gamma_1 \mathrm{std}[\ln(\mathrm{CMix}_{ij})]][1 + \gamma_2 \ln(\mathrm{Staff}_i)] \\ &\quad + [\delta_0 + \delta_1 \mathrm{std}[\ln(\mathrm{CMix}_{ij})]] \ln(\mathrm{Beds}_{ij}). \end{aligned} \tag{7}$$

In both cases, the elasticity depends on the three explicative variables of the model. In particular, as one can see in Appendix—Table A2, elasticities are mainly affected by the case-mix. Given a positive case-mix value (considered in its standardized form), the estimated elasticity of the hospitalizations number with respect to Beds (consider the right-hand side of the table) decreases with Beds, when Staff is fixed. Specularly it increases with Staff, when Beds are fixed. On the contrary, given a negative case-mix value, specular patterns can be observed. Finally, given a null value of case-mix, elasticities are almost constant.

Table A1, in Appendix, summarizes the estimated elasticity for Beds and Staff calculated for every observed macro units (i.e. at hospital level). Equation (7) can be, in fact, reformulated at hospital level by writing output y as the sum of outputs of single micro unit, reducing itself to the weighted sum of elasticities at ward level:

$$
\begin{aligned}
e^{\mathrm{H}}(\mathrm{Staff})_i &= \frac{\partial \sum_j y_{ij}}{\partial \mathrm{Staff}_i}\,\frac{\mathrm{Staff}_i}{\sum_k y_{ik}} = \sum_j \left(\frac{\partial y_{ij}}{\partial \mathrm{Staff}_i}\,\frac{\mathrm{Staff}_i}{\sum_k y_{ik}}\,\frac{y_{ij}}{y_{ij}}\right),\\
&= \sum_j \left(\frac{\partial y_{ij}}{\partial \mathrm{Staff}_i}\,\frac{\mathrm{Staff}_i}{y_{ij}}\,\frac{y_{ij}}{\sum_k y_{ik}}\right),\\
&= \sum_j \left(\frac{\partial y_{ij}}{\partial \mathrm{Staff}_i}\,\frac{\mathrm{Staff}_i}{y_{ij}}\,w_{ij}\right) = \sum_j \left(\frac{\partial \ln y_{ij}}{\partial \ln \mathrm{Staff}_i}\,w_{ij}\right),\\
&= \sum_j e^{\mathrm{W}}(\mathrm{Staff})_{ij}\,w_{ij},
\end{aligned} \tag{8}
$$

where e^{H} indicates the elasticity at hospital level and w_{ij} are to be intended as weights of the elasticities at ward level.

It can be simply demonstrated, by means of the limit of the difference quotient, that analogous results can be achieved for Beds, obtaining

$$
e^{\mathrm{H}}(\mathrm{Beds})_i = \sum_j \left(\frac{\partial \ln y_{ij}}{\partial \ln \mathrm{Beds}_{ij}}\,w_{ij}\right) = \sum_j e^{\mathrm{W}}(\mathrm{Beds})_{ij}\,w_{ij}. \tag{9}
$$

The estimates of elasticity at hospital level are summarized in Fig. 4.

From this figure one can notice that, for Beds, the elasticity is concentrated around the unity. For Staff, however, the plot shows that almost all units work in over-dimensioned conditions, and a few have even reached a congested estate (the negative values). From a re-allocative point of view, we can then identify situations where elasticity values can justify an input increase (elasticities greater than one) and cases for which an additional input brings up no proportional output improvement (elasticities lower than one).

Considering the sum of the elasticities for Beds and Staff, we obtain a slight interpretation of elasticity results as the local returns to scale (so called *elasticity of scale*)

$$
e(x) = e(\mathrm{Beds})_i + e(\mathrm{Staff})_i, \tag{10}
$$

where x indicates the input vector (Beds, Staff). The elasticity of scale is a scalar-valued measure of how output changes in response to simultaneous input variation (Chambers, 1988). Here simultaneous input variation is restricted to variations that do not change relative input utilization; i.e. the ratios (x_r/x_s) are constant for all r and s.

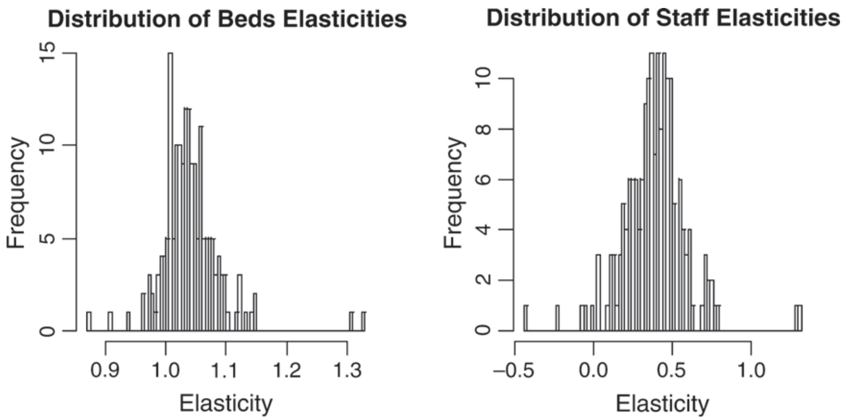

FIG. 4. Histograms of hospital elasticities for Beds and Staff.

> By geometrical point of view, the elasticity of scale is interpretable as measuring how accurately the distance between isoquants in input space reflects the distance in output space. In particular, there are three possible characterizations of production functions. If $e(x) = 1$, the production function exhibits constant returns to scale, and the isoquants are evenly spaced. If $e(x) < 1$, the production function exhibits decreasing returns to scale, and the distance between isoquants in input space overestimates the distance in output space. Finally, if $e(x) > 1$, the distance in input space under-estimates the distance in output space, and the production function exhibits increasing returns to scale; isoquants, therefore, tend to be more crowded together as one moves along a ray from the origin (Chambers, 1988).

The preceding concepts have some economical interpretations.

> Assume that, by a given endowment of inputs, the goal is producing as much output as possible and that we can decide whether or not it would be better to split up the resource endowment equally into m separate operations or to produce everything in one large operation. For convenience, suppose also that both alternatives are equally costly. If the available technology is characterized by decreasing returns to scale, there is no incentive to centralize the operation, and it is better to split up the operation; exactly analogous arguments show that when $e(x) = 1$, centralization and decentralization are indifferent, and when $e(x) > 1$, centralization is preferable (Chambers, 1988).

These situations can be interpreted as the estimated productive conditions of each hospital and each ward. Focusing our analysis on hospital-level results, an estimated scale elasticity under the unity identifies hospitals that present a congested condition, where investments are not useful and it is better to reduce the dimension; on the contrary, hospitals that have a higher return to scale are the ones with unused productive capacity, which could increase their dimension. This interpretation could be very useful from an administrative point of view. Table A1, in Appendix, shows also the estimated scale elasticity for each hospital. It can be noted, however, that the differences among hospitals are almost entirely due to Staff elasticity. Staff and Beds values are also given in these tables in order to allow a straightforward interpretation of the estimated values.

Another interesting measure, relating to the returns to scale characteristics of a technology, is the *scale efficiency*, which measures the ray average productivity at the observed input scale on the production frontier relative to the maximum ray average productivity attainable at an input bundle x characterized by $e(x) = 1$, defined by Banker (1984) as the *most productive scale size*.

> It needs to be emphasized that scale efficiency is lower than one whenever the observed input mix is not scale-optimal, i.e. where locally constant returns to scale does not hold. Scale elasticity, on the other hand, can be either greater than or less than unity. Only at the *most productive scale size* both measures equal unity and are, therefore, equal to one another. Elsewhere, they differ and the magnitude of scale elasticity does not directly reveal anything about the level of scale efficiency (Ray, 1998).

Ray (1998) developed an input-oriented measure of scale efficiency, directly obtainable from an empirically estimated single output–multiple input translog production function, e.g. regarding the translog model

$$\ln(\mathrm{NHosp}) = \alpha_0 + \alpha_1 \ln(\mathrm{Beds}) + \alpha_2 \ln(\mathrm{Staff}) + \frac{\beta_{11}}{2}[\ln(\mathrm{Beds})]^2 + \beta_{12} \ln(\mathrm{Beds}) \ln(\mathrm{Staff}) + \frac{\beta_{22}}{2}[\ln(\mathrm{Staff})]^2, \tag{11}$$

and in the absence of technical inefficiency, the scale efficiency is equal to

$$\mathrm{SE}(x) = \exp\left\{\frac{(1 - e(x))^2}{2\beta}\right\}, \tag{12}$$

where $\beta = \sum_{r=1}^{2} \sum_{s=1}^{2} (\beta_{rs})$.

In the more general case involving technical inefficiency, it is equal to

$$\mathrm{SE}(x) = \exp\left\{\frac{\left[1 - \sqrt{e(x)^2 - 2\beta\theta}\right]^2}{2\beta}\right\}, \tag{13}$$

where θ is the technical inefficiency (in our case we can obtain it from a modified ordinary least squares (MOLS) or Corrected ordinary least squares (COLS) transformation of the second-level error component). This measure can be computed for each hospital from the fitted translog production function.

As previously seen, a practical implication of returns to scale analysis is that any input characterized by increasing returns to scale should be expanded, while one with diminishing returns should be scaled down in order to attain full scale efficiency. Supposing it is possible to change relative input utilization, an interesting question would be: what level of Staff combined with the given quantity of Beds would result in a scale-optimal input mix?

Ray (1998) developed an index, in the two-input case, which measures the extent to which the observed quantity of Staff differs from what would be the optimal level in light of the size of its actual Beds.

Let B_0 be the exogenously fixed quantity of Beds, S_0 the observed quantity of Staff and (S_0^*) the optimal quantity of Staff (in the sense of *most productive scale size*). Ray (1998) defined

$$\sigma = \frac{S_0^*}{B_0} \bigg/ \frac{S_0}{B_0} = \frac{S_0^*}{B_0}\,\frac{B_0}{S_0} = \frac{S_0^*}{S_0}. \tag{14}$$

Clearly, $\sigma = 1$ if and only if the observed pair (S_0, B_0) is itself scale-optimal. Otherwise, $\sigma > 1$ implies that the actual Beds–Staff ratio is higher than the optimal. Similarly, $\sigma < 1$ implies excessive number of Staff relative to the optimal level.

For the translog case the index is equal to

$$\sigma = \exp\left\{\frac{1 - e(S_0, B_0)}{\beta_{22} + \beta_{12}}\right\}, \tag{15}$$

i.e. it depends on both the observed scale elasticity, $e(S_0, B_0)$, and the estimated values of the parameters in the denominator. Thus, the mere fact that increasing returns to scale hold at any given input bundle does not by itself imply that the observed quantity of Staff is too low.

It should be noted, however, that if $\beta_{22} + \beta_{12} = 0$, there will not exist a finite S_0^* for the given B_0 (Ray, 1998).

6. Conclusions

The decision to measure the efficiency of health services by the given model is due to both interpretability and flexibility of the functional form. The model proposed is certainly a simplified version of the complete econometric model specification (some other variables, in fact, can affect the analysed phenomenon) but, also at this preliminary stage, some of the obtained results are really closed to the desirable hypotheses. Then one can conclude that the application of this methodology provides useful and reliable results. Some attention should be focused, however, on typical econometric problems, like outlier detection; this is, in fact, still an unsolved problem in the multilevel framework, because it is not easy to identify at which level outliers should be searched (see Langford & Toby, 1998; Barnett & Toby, 1994).

By means of deterministic COLS and MOLS approaches (Greene, 1993), decision makers can interpret the random effect BLUP estimates $(\hat{u}_i)$ as indicators of structure efficiency. The larger the effect, the better the productive process. Then, the interpretation of estimated elasticities provides some information about the productive conditions of observed units.

Our results show that, as beds elasticities are mainly concentrated around unity, the interest of decision makers should be focused on estimated staff elasticities. Both staff elasticity and scale elasticity highlight the presence of over- and under-dimensioned units, i.e. situations where a re-allocation of staff is necessary.

The sample used for the analysis included different kinds of hospital structures. As reported in Table 3, while structures classified as ‘hospitals’ and ‘classified hospitals’ are almost homogenous in terms of staff elasticity, ‘private and public clinics’ present a large variability. ‘Research structures’ are instead characterized by a low average staff elasticity and, consequently, worse productive conditions, maybe due to the different goal of these structures. Here staff is quite completely devoted to research and the health service is considered only as a secondary aim.

In our analysis, the Case-Mix is treated as an exogenous variable, which does not interfere with hospital politics. This is correct in the case that we consider as fixed the service demand for each structure. As one can see in Table A2 of Appendix, substantial changes in case-mix index cause the raise of different elasticity patterns.

Future developments will consider the generalization of Ray’s (1998) results to our model specification and the consequent estimation of scale efficiency and σ-index for scale optimality of input mix.

TABLE 3 *Elasticity summary statistics of hospitalization structures clusters*

	Percentiles						Standard	Observed
	0.01	0.25	0.5	0.75	0.99	Mean	deviation	number
Beds elasticities structure type								
Hospitals	0.971	1.015	1.032	1.051	1.088	1.033	0.025	118
Research institutes	0.913	0.992	1.013	1.036	1.086	1.010	0.042	18
Classified hospitals	1.010	1.022	1.025	1.042	1.072	1.035	0.022	5
Private and public clinics	0.919	1.028	1.068	1.099	1.317	1.067	0.074	51
Staff elasticities structure type								
Hospitals	0.003	0.302	0.375	0.449	0.600	0.369	0.123	118
Research institutes	−0.045	0.187	0.265	0.354	0.703	0.282	0.183	18
Classified hospitals	0.269	0.321	0.353	0.426	0.563	0.387	0.104	5
Private and public clinics	−0.326	0.253	0.478	0.604	1.297	0.453	0.303	51

In conclusion, we think that the presented methodology and relative results can be considered really interesting for the decision making processes. In fact, both random and fixed effect estimates are easily interpretable. From an administrative point of view, they can be used to identify the different productive observed conditions (both technical and scale inefficiency) and, in case, to decide the preferable interventions.

Acknowledgements

Section 1 as well as the supervision of the paper is due to EG, the remaining sections were developed jointly by LG and SCM. If need be, Sections 2–4 could be attributed to LG and Sections 5–6 to SCM.

REFERENCES

AIGNER, D. J., LOVELL, C. A. K. & SCHMIDT, P. J. (1977) Formulation and estimation of stochastic frontier production models. *J. Econometrics*, **6**, 21–37.

BANKER, R. D. (1984) Estimating the most productive scale size using data envelopment analysis. *Eur. J. Oper. Res.*, **17**, 35–44.

BARNETT, V. & TOBY, L. (1994) *Outliers in Statistical Data*. Chichester: Wiley.

CHAMBERS, R. G. (1988) *Applied Production Analysis. A Dual Approach*. Cambridge: Cambridge University Press.

CHRISTENSEN, L. R., JORGENSON, D. W. & LAU, L. J. (1973) Transcendental logarithmic production frontiers. *Rev. Econ. Stat.*, **55**, 28–45.

CRIBARI-NETO, F. & ZARKOS, S. G. (1999) R: yet another econometric programming environment. *J. Appl. Econ.*, **14**, 319–329.

GODFREY, L. G. (1988) *Misspecification Tests in Econometrics: The Lagrange Multiplier Principle and Other Approaches*. New York: Cambridge University Press.

GOLDSTEIN, H. & HEALY, M. J. R. (1995) The graphical presentation of a collection of means. *J. R. Stat. Soc. Ser. A–G*, **159**, 505–513.

GORI, E., GRASSETTI, L. & ROSSI, C. (2002) Linear mixed models in efficiency analysis: evidence from validation procedures. *Stat. Appl.*, **14**, 247–275.

GRANT, J. H. (1993) The translog approximate function—substitution among inputs in manufacturing evaluated at sample means. *Econ. Lett.*, **41**, 235–240.

GREENE, W. H. (1993) The econometric approach to efficiency analysis. *The Measurement of Productive Efficiency* (H. O. Fried, C. A. K. Lovell & S. S. Schmidt eds). New York: Oxford University Press.

IHAKA, R. & GENTLEMAN, R. (1996) R: a language for data analysis and graphics. *J. Comput. Graph. Stat.*, **5**, 299–314.

KIM, H. Y. (1992) The translog production function and variable returns to scale. *Rev. Econ. Stat.*, **74**, 546–552.

LANGFORD, I. H. & TOBY, L. (1998) Outliers in multilevel data. *J. R. Stat. Soc. Ser. A–G*, **161**, 121–160.

LEYLAND, A. H. & GOLDSTEIN, H. (2001) *Multilevel Modelling of Health Statistics*. Chichester: Wiley.

PINHEIRO, J. & BATES, D. (2002) *Mixed-Effects Models in S and S-PLUS Series: Statistics and Computing*. New York: Springer.

RAY, S. C. (1998) Measuring scale efficiency from a translog production function. *J. Prod. Anal.*, **11**, 183–194.

RYAN, D. L. & WALES, T. J. (2000) Imposing local concavity in the translog and generalized Leontief cost functions. *Econ. Lett.*, **67**, 253–260.

VERBEKE, G. & MOLEMBERGHS, G. (2000) *Linear Mixed Models for Longitudinal Data*. New York: Springer.

Appendix: Elasticity results

TABLE A1 *Estimated elasticities computed with respect to each input variable and scale elasticity*

HospID	Beds Elasticity	Staff Elasticity	Scale Elasticity	Beds	Staff	HospID	Beds Elasticity	Staff Elasticity	Scale Elasticity	Beds	Staff
1	1.010	0.330	1.339	170	351	46	1.006	0.238	1.244	128	246
2	1.031	0.367	1.397	853	1327	47	1.049	0.445	1.494	162	322
3	1.015	0.311	1.326	622	1187	48	1.016	0.240	1.256	75	202
4	0.997	0.238	1.235	630	1042	49	1.031	0.374	1.405	165	303
5	1.001	0.185	1.186	151	217	50	1.006	0.133	1.138	50	102
6	1.039	0.401	1.440	293	579	51	1.068	0.518	1.586	271	485
7	1.007	0.206	1.213	268	285	52	1.062	0.472	1.534	329	676
8	1.060	0.475	1.535	123	228	53	1.021	0.113	1.135	58	55
9	1.054	0.411	1.465	400	95	54	1.025	0.353	1.378	351	559
10	1.141	0.765	1.907	126	143	55	1.122	0.697	1.819	168	219
11	1.061	0.509	1.570	214	559	56	1.125	0.710	1.835	110	168
12	1.044	0.363	1.406	77	154	57	1.068	0.478	1.546	80	48
13	1.050	0.444	1.494	368	702	58	1.048	0.150	1.198	282	309
14	1.046	0.443	1.489	140	280	59	1.133	0.719	1.852	130	134
15	1.042	0.422	1.464	102	234	60	1.146	0.713	1.860	300	365
16	1.042	0.426	1.469	258	408	61	0.872	-0.430	0.442	160	170
17	1.073	0.569	1.642	270	932	62	1.023	0.245	1.268	71	65
18	1.022	0.321	1.343	201	321	63	1.084	0.564	1.648	60	55
19	1.074	0.520	1.594	180	182	64	1.328	1.303	2.631	136	232
20	0.989	0.193	1.182	106	126	65	0.995	0.261	1.256	260	350
21	1.068	0.471	1.538	100	100	66	1.039	0.397	1.436	182	240
22	1.043	0.244	1.287	170	183	67	1.039	0.378	1.417	91	98
23	0.965	-0.223	0.742	60	42	68	1.061	0.464	1.525	190	185
24	1.028	0.296	1.324	32	91	69	1.092	0.601	1.692	198	358
25	1.036	0.432	1.468	525	926	70	1.116	0.758	1.874	380	681
26	1.049	0.439	1.488	130	267	71	1.086	0.743	1.830	60	70
27	1.059	0.494	1.553	140	331	72	1.016	0.226	1.243	144	96
28	1.027	0.288	1.315	92	180	73	1.025	0.307	1.332	105	242
29	1.026	0.315	1.340	342	538	74	1.045	0.445	1.490	495	943
30	1.056	0.482	1.538	172	461	75	1.084	0.557	1.640	129	264
31	1.006	0.233	1.239	260	474	76	1.030	0.366	1.396	212	377
32	1.089	0.606	1.695	349	663	77	1.008	0.263	1.271	157	240
33	1.088	0.581	1.669	72	240	78	0.963	0.179	1.142	163	226
34	0.998	0.240	1.238	587	977	79	1.033	0.352	1.384	140	244
35	1.044	0.446	1.491	549	1270	80	1.028	0.342	1.371	186	232
36	1.053	0.482	1.534	391	857	81	1.063	0.443	1.506	170	321
37	1.089	0.592	1.681	188	434	82	1.064	0.487	1.551	68	165
38	1.033	0.374	1.408	258	469	83	1.074	0.563	1.637	279	585
39	1.033	0.374	1.407	493	1016	84	0.986	0.122	1.108	78	182
40	1.036	0.413	1.449	511	1347	85	1.306	1.290	2.596	130	229
41	1.015	0.354	1.368	575	1109	86	1.100	0.570	1.670	148	107
42	1.010	0.198	1.209	223	333	87	0.979	0.172	1.151	142	152
43	1.047	0.414	1.461	141	262	88	0.982	0.175	1.156	135	174
44	1.010	0.321	1.331	504	1085	89	1.032	0.387	1.418	319	415
45	1.022	0.334	1.357	238	493	90	1.047	0.449	1.496	420	551

TABLE A1 *Continued*

HospID	Beds Elasticity	Staff Elasticity	Scale Elasticity	Beds	Staff	HospID	Beds Elasticity	Staff Elasticity	Scale Elasticity	Beds	Staff
91	0.976	0.022	0.998	92	128	135	1.057	0.480	1.537	135	327
92	0.974	-0.056	0.918	108	184	136	1.066	0.476	1.541	174	211
93	1.035	0.357	1.392	35	47	137	1.057	0.461	1.518	95	181
94	1.060	0.493	1.553	236	616	138	1.042	0.423	1.465	201	534
95	1.052	0.482	1.534	366	667	139	1.146	0.733	1.879	94	85
96	1.037	0.378	1.414	167	368	140	1.054	0.451	1.505	269	541
97	1.016	0.319	1.335	197	292	141	1.037	0.404	1.441	405	789
98	1.036	0.415	1.451	155	272	142	1.079	0.503	1.582	180	79
99	1.020	0.345	1.366	203	378	143	1.123	0.794	1.917	744	210
100	1.031	0.383	1.414	168	300	144	1.017	0.358	1.375	1021	1644
101	1.008	0.284	1.292	90	211	145	1.004	0.263	1.267	481	732
102	1.083	0.601	1.684	118	391	146	1.057	0.489	1.546	230	487
103	1.029	0.247	1.276	100	92	147	1.042	0.460	1.502	917	2135
104	1.009	0.267	1.276	378	624	148	1.030	0.411	1.440	895	1653
105	1.095	0.554	1.649	275	430	149	1.035	0.400	1.436	724	1407
106	1.066	0.484	1.550	105	107	150	1.015	0.327	1.342	706	1461
107	1.073	0.492	1.564	145	133	151	1.018	0.390	1.408	1574	3198
108	1.078	0.552	1.630	319	421	152	0.997	0.301	1.298	2155	3802
109	0.996	0.024	1.019	86	65	153	1.029	0.415	1.445	747	1558
110	1.052	0.445	1.497	320	325	154	1.012	0.337	1.348	941	1630
111	1.047	0.402	1.449	120	113	155	1.010	0.360	1.370	1298	2570
112	1.018	0.288	1.306	122	146	156	1.032	0.426	1.458	644	1462
113	1.023	0.251	1.274	86	123	157	0.971	0.087	1.058	494	1081
114	1.028	0.368	1.396	102	192	158	1.015	0.399	1.414	746	2369
115	1.035	0.335	1.370	140	171	159	1.007	0.370	1.377	1608	4113
116	1.057	0.419	1.476	125	205	160	1.009	0.348	1.358	564	1529
117	1.071	0.548	1.619	345	661	161	1.008	0.280	1.289	766	1642
118	1.047	0.465	1.512	452	745	162	1.024	0.384	1.408	613	1401
119	1.104	0.621	1.724	100	89	163	1.025	0.215	1.240	80	103
120	0.974	0.027	1.001	76	61	164	0.991	0.306	1.297	750	2520
121	1.098	0.608	1.706	160	135	165	0.936	0.108	1.044	510	1511
122	1.100	0.559	1.659	30	17	166	1.036	0.407	1.443	207	525
123	1.003	0.182	1.185	57	39	167	1.032	0.428	1.460	1394	2965
124	1.076	0.508	1.584	147	142	168	0.990	0.103	1.093	53	192
125	1.059	0.435	1.494	162	229	169	1.004	0.206	1.210	173	253
126	1.079	0.590	1.668	147	210	170	1.016	0.273	1.289	130	194
127	1.040	0.450	1.490	472	902	171	1.037	0.259	1.296	64	57
128	1.020	0.246	1.266	95	124	172	1.091	0.731	1.823	189	410
129	0.991	-0.001	0.990	97	115	173	1.056	0.542	1.597	993	3150
130	1.023	0.345	1.368	120	127	174	0.993	0.137	1.130	154	321
131	1.041	0.373	1.413	137	143	175	0.908	-0.071	0.837	110	747
132	1.077	0.485	1.562	145	113	176	1.010	0.337	1.347	242	487
133	1.056	0.467	1.524	164	320	177	1.038	0.325	1.363	159	247
134	1.019	0.199	1.218	90	142	178	1.001	0.203	1.205	287	390

TABLE A2 *Estimated elasticities computed with respect to each input by given values of case-mix index. Negative values of case-mix index correspond to easier treatments which need more Beds than Staff. In a specular way positive case-mix identifies the treatments that need more Staff than Beds*

Elasticity of NHosp calculated with respect to Staff given different input mix and different levels of case-mix (std[ln(CMix)])							
	Beds						
Staff	10	20	30	40	50	100	500
Case-mix = 2							
10	1.902	1.920	1.930	1.938	1.943	1.961	2.002
50	1.201	1.219	1.229	1.236	1.242	1.260	1.301
200	0.597	0.615	0.625	0.633	0.638	0.656	0.697
500	0.198	0.216	0.226	0.233	0.239	0.257	0.298
Case-mix = 1							
10	1.277	1.282	1.285	1.287	1.288	1.293	1.305
50	0.825	0.830	0.833	0.835	0.837	0.841	0.853
200	0.436	0.441	0.444	0.446	0.447	0.452	0.464
500	0.179	0.183	0.186	0.188	0.190	0.195	0.206
Case-mix = 0							
10	0.652	0.644	0.639	0.636	0.634	0.626	0.607
50	0.449	0.441	0.437	0.433	0.431	0.423	0.405
200	0.275	0.267	0.262	0.259	0.256	0.248	0.230
500	0.159	0.151	0.147	0.143	0.141	0.133	0.114
Case-mix = −1							
10	0.027	0.006	−0.006	−0.015	−0.021	−0.042	−0.090
50	0.073	0.053	0.041	0.032	0.025	0.004	−0.044
200	0.113	0.093	0.081	0.072	0.065	0.044	−0.004
500	0.140	0.119	0.107	0.098	0.092	0.071	0.023
Case-mix = −2							
10	−0.598	−0.632	−0.652	−0.665	−0.676	−0.710	−0.788
50	−0.303	−0.336	−0.356	−0.370	−0.381	−0.414	−0.492
200	−0.048	−0.081	−0.101	−0.115	−0.126	−0.159	−0.237
500	0.121	0.087	0.067	0.053	0.043	0.009	−0.069

Elasticity of NHosp calculated with respect to beds given different input mix and different levels of case-mix (std[ln(CMix)])							
	Beds						
Staff	10	20	30	40	50	100	500
Case-mix = 2							
10	1.460	1.315	1.229	1.169	1.122	0.976	0.637
50	1.502	1.356	1.271	1.210	1.163	1.017	0.679
200	1.537	1.391	1.306	1.246	1.199	1.053	0.714
500	1.561	1.415	1.329	1.269	1.222	1.076	0.738
Case-mix = 1							
10	1.295	1.211	1.162	1.127	1.100	1.015	0.819
50	1.307	1.223	1.173	1.138	1.111	1.027	0.831
200	1.317	1.232	1.183	1.148	1.121	1.036	0.841
500	1.323	1.239	1.190	1.154	1.127	1.043	0.847
Case-mix = 0							
10	1.131	1.108	1.094	1.085	1.077	1.055	1.001
50	1.112	1.089	1.076	1.066	1.059	1.036	0.983
200	1.096	1.073	1.060	1.050	1.043	1.020	0.967
500	1.086	1.063	1.050	1.040	1.033	1.010	0.957
Case-mix = −1							
10	0.966	1.004	1.027	1.043	1.055	1.094	1.183
50	0.917	0.956	0.978	0.995	1.007	1.046	1.135
200	0.876	0.914	0.937	0.953	0.965	1.004	1.094
500	0.848	0.887	0.910	0.926	0.938	0.977	1.066
Case-mix = −2							
10	0.801	0.901	0.959	1.001	1.033	1.133	1.365
50	0.723	0.823	0.881	0.923	0.955	1.055	1.287
200	0.655	0.755	0.814	0.855	0.888	0.988	1.220
500	0.611	0.711	0.770	0.811	0.843	0.943	1.176