Multilevel Flexible Specification of the Production Function in Health Economics

Authors:
Luca Grassetti - Dept. of Statistical Sciences - University of Padova
Via C. Battisti, 241 - 35121 Padova, Italy
E-mail: grassetti@stat.unipd.it
Enrico Gori - Dept. of Statistics - University of Udine
E-mail: e.gori@dss.uniud.it
Simona C. Minotti - Dept. of Econ. and Social Sciences - Catholic University of Piacenza
E-Mail: simona.minotti@unicatt.it

Abstract
Previous studies on efficiency of hospitals often refer to quite restrictive functional forms for the technology. In this paper, referring to a study about some hospitals in Lombardy, we propose a statistical model based on the translogarithmic function - the most widely used flexible functional form (Christensen, Jorgenson and Lau, 1973). More specifically, in order to take into consideration the hierarchical structure of the data (as in Gori, Grassetti and Rossi, 2004), we propose a multilevel model, ignoring for the moment the one-side error specification, typical of stochastic frontier analysis. Given this simplification, however, we are easily able to take into account some typical econometric problems as, for example, 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) in our context.

The interpretation of the results is surely an interesting administrative instrument for decision makers in order to analyze 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. 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, which is the most widely used flexible functional form for economic functions (Christensen, Jorgenson and Lau, 1973; Kim, 1992; Grant, 1993; Ryan and Wales, 2000). This model allows to obtain also closed form measures of scale elasticity and scale efficiency, readily computable from the fitted model. 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 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:

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

The success of this functional form in many econometric applications is due to its flexibility (for example, 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 (an input variable) 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.

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

The data involved in the study are characterized by the presence of two levels of data collection. While the first one regards the single wards in each hospital, the second is identified by the hospitals.

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 just said the data involved in the study are characterized by a two level hierarchical structure. More, in depth, the considered dataset refers to 1478 first level units (wards) observed in 178 second level units (hospitals) of Italian region Lombardy in 1997. This dataset, described in Gori, Grassetti and Rossi (2004), 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, includes 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 cases treated in the observed year), which is the output variable (indicated by \(NHosp\)). The beds number (indicated by \(Beds\)) is available for each ward in a hospital. 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.
Then, some other informations are taken into consideration. First of all, in order to distinguish the hospitalizations by degree of complexity, there is the case-mix index, which represents the relative level of case complexity in each ward, given the regional mean. In particular, we use a standardized version of the logarithm of this index (indicated by $std(ln(\text{CMix}))$). 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 Disease Related Groups (DRG) weight variable, calculated at hospital and ward level. The DRG code derive 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 $\mu_{\text{Weig,ij}}$ and $\mu_{\text{Weig,i}}$) and represents the mean complexity of observed cases at ward and hospital level. The second one is the standard deviation of DRG weights ($\sigma_{\text{Weig,ij}}$ and $\sigma_{\text{Weig,i}}$) 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.

In Table 1 are summarized the descriptive statistics of the variables involved in the model.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Min</th>
<th>Median</th>
<th>Max</th>
<th>Mean</th>
<th>Stand. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$ln(\text{N Hosp})$</td>
<td>0.000</td>
<td>6.903</td>
<td>9.806</td>
<td>6.589</td>
<td>1.947</td>
</tr>
<tr>
<td>$ln(\text{Beds})$</td>
<td>0.000</td>
<td>3.296</td>
<td>6.234</td>
<td>3.166</td>
<td>0.926</td>
</tr>
<tr>
<td>$ln(\text{Staff})$</td>
<td>2.833</td>
<td>6.497</td>
<td>8.322</td>
<td>6.442</td>
<td>1.147</td>
</tr>
<tr>
<td>mean DRG Weight (Ward lev.)</td>
<td>0.160</td>
<td>0.850</td>
<td>8.320</td>
<td>1.076</td>
<td>0.776</td>
</tr>
<tr>
<td>s.d. DRG Weight (Ward lev.)</td>
<td>0.000</td>
<td>0.420</td>
<td>6.870</td>
<td>0.682</td>
<td>0.937</td>
</tr>
<tr>
<td>mean DRG Weight (Hospital lev.)</td>
<td>0.630</td>
<td>0.850</td>
<td>2.130</td>
<td>0.886</td>
<td>0.030</td>
</tr>
<tr>
<td>s.d. DRG Weight (Hospital lev.)</td>
<td>0.120</td>
<td>0.670</td>
<td>1.950</td>
<td>0.695</td>
<td>0.093</td>
</tr>
<tr>
<td>$std(ln(\text{CMix}))$</td>
<td>-10.350</td>
<td>-0.023</td>
<td>2.935</td>
<td>0.000</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 1: Descriptive Statistics

The pair-wise scatter plot in Figure 1 shows the relation between the output variable and the input variables and in particular highlights the strong dependence between hos-
pitalizations number and beds number. From in-depth studies this relationship can not be considered linear.

Figure 1: Pair-wise Scatter Plot among output variable and input variables

3 The Model

In order to model the hospitalization phenomenon 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 from case-mix index (indicated by $std(ln(CMix_{ij}))$). The second one is the additive linear component illustrated in Section 2 (indicated in the following by $f(Weights)$).
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:

\[
\ln(N_{\text{Hosp}_{ij}}) = \alpha(CMix_{ij}) + \beta(CMix_{ij}, Beds_{ij}) + \\
\quad + \gamma(CMix_{ij}, Staff_i) + \delta(CMix_{ij}, Beds_{ij}, Staff_i) \\
\quad + f(Weights) + u_i + \epsilon_{ij},
\]  \hspace{1cm} (2)

where

\[
\alpha(CMix_{ij}) = \alpha_0 + \alpha_1 \text{std}(\ln(CMix_{ij})) \\
\beta(CMix_{ij}, Beds_{ij}) = [\beta_0 + \beta_1 \text{std}(\ln(CMix_{ij}))] \cdot [\ln(Beds_{ij}) + \frac{\delta_0}{2} \ln(Beds_{ij})^2] \\
\gamma(CMix_{ij}, Staff_i) = [\gamma_0 + \gamma_1 \text{std}(\ln(CMix_{ij}))] \cdot [\ln(Staff_i) + \frac{\delta_0}{2} \ln(Staff_i)^2] \\
\delta(CMix_{ij}, Beds_{ij}, Staff_i) = [\delta_0 + \delta_1 \text{std}(\ln(CMix_{ij}))] \cdot [\ln(Beds_{ij}) \cdot \ln(Staff_i)] \\
f(Weights) = \lambda_1 \mu_{\text{Weights}_{ij}} + \lambda_2 \sigma_{\text{Weights}_{ij}} + \lambda_3 \mu_{\text{Weights}_{i}} + \lambda_4 \sigma_{\text{Weights}_{i}}.
\]

\(u_i \sim N(0, \sigma^2_u)\) are the residuals at hospital level, \(i=1, \ldots, N\),

\(\epsilon_{ij} \sim N(0, \sigma^2_e)\) are the residuals at ward level, \(j=1, \ldots, M\), with \(u_i \perp \epsilon_{ij}\).

Using a classical notation for the coefficients of mixed models, equation (2) can be re-written as:
\[ \ln(N\text{Hosp}_{ij}) = \alpha_0 + \alpha_1 \ln(Beds_{ij}) + \frac{\alpha_2}{2}(\ln(Beds_{ij})^2) + \alpha_3 \ln(Staff_{i}) + \]
\[ + \frac{\alpha_4}{2}(\ln(Staff_{i})^2) + \alpha_5 (\ln(Beds_{ij}) \cdot \ln(Staff_{i})) + \alpha_6 (\ln(Beds_{ij}) \cdot \text{std}(\ln(CMix_{ij}))) + \]
\[ + \alpha_7 (\ln(Staff_{i}) \cdot \text{std}(\ln(CMix_{ij}))) + \alpha_8 (\ln(Staff_{i}) \cdot \text{std}(\ln(CMix_{ij}))) + \]
\[ + \alpha_9 (\ln(Beds_{ij})^2 \cdot \text{std}(\ln(CMix_{ij}))) + \alpha_{10} (\ln(Staff_{i})^2 \cdot \text{std}(\ln(CMix_{ij}))) + \]
\[ \alpha_{11} \text{std}(\ln(CMix_{ij})) + \lambda_1 \mu_{eig,ij} + \lambda_2 \sigma_{eig,ij} + \lambda_3 \mu_{eig,i} + \lambda_4 \sigma_{eig,i} + \epsilon_{ij}, \]

where \( \alpha_0 = \alpha_0 + u_i, \alpha_1 = \beta_0, \alpha_2 = \beta_0/\beta_2, \) etc.

Note that the model defined in (3) can be seen as an unconstrained version of model (2). Equation (3) is obtained by substitution of \( \alpha(CMix_{ij}), \beta(CMix_{ij}), \gamma(CMix_{ij}), \delta(CMix_{ij}) \) and \( f(Weights) \) in (2). This operation leads to a linear form with some restriction on parameter values. For example, coefficients connected to \( \ln(Beds) \) correspond to \( \beta_0, \beta_1, \frac{\beta_0 \beta_3}{2} \) and \( \frac{\beta_0 \beta_3}{2} \), that are only three different coefficients in model (2) and, respectively, \( \alpha_1, \alpha_6, \alpha_9 \) and \( \frac{\alpha_2}{2} \) in model (3). In order to justify the re-parametrization 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 to the immediate interpretability of the estimated coefficients.

Then, regarding the errors at first level, \( \epsilon_{ij} \), usually called residuals, the estimation results, performed with the statistical software R (see Cribari-Neto and Zarkos, 1999; Ihaka and Gentleman, 1996), show a high level of heteroscedasticity (see Figure 2-A). In order
to include this information in the model, we have defined a dependence of $\epsilon$ 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:

$$Var(\epsilon_{ij}) = \sigma^2_{ij} |Bed_{ij}|^{2\theta}.$$  

(4)

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

Figure 2: Effect on the residuals of inclusion of heteroscedasticity in the model

Finally, the $2^{nd}$ level errors, $u_i$, can be interpreted as efficiency indicators, as illustrated for example in Gori, Grassetti and Rossi (2004), and can be estimated by “Empirical Bayes” (EB) approach as described in Pinheiro and Bates (2000) and Verbeke and Molenberghs (2000).
4 The Results

In Table 2 we summarize the estimated coefficients for fixed and random effects of the model introduced in equation (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. For the parameters in the linear mixed-effects model approximate confidence intervals are obtained by using the normal approximation of the distribution of the maximum likelihood 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 number of hospitalization as well as the various complexity indexes (\(std(ln(CMix_{ij}))\), \(\mu_{W_{eig,ij}}\), \(\sigma_{W_{eig,ij}}\), \(\mu_{W_{eig,i}}\) and \(\sigma_{W_{eig,i}}\)). A positive effect is, indeed, connected to the original input variables, also taking into consideration their interaction with the standardized case-mix index.

![Hospital Level Random Effects](image)

Figure 3: Confidence intervals for estimated BLUP random effects \((u_i)\) at hospital level

In Figure 3 we have summarized the estimated BLUP error components. Confidence intervals have been calculated on the basis of pairwise comparison theory given in Goldstein and Healy (1995). The estimated BLUP error components and their standard errors have been computed following chapter 7 of Pinheiro and Bates (2000). The point es-
<table>
<thead>
<tr>
<th>Fixed Effects Coefficients</th>
<th>Value</th>
<th>Std.Error</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_0$</td>
<td>0.7941</td>
<td>1.0367</td>
<td>0.4434</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>1.2328</td>
<td>0.1541</td>
<td>0.0000</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>-0.0165</td>
<td>0.0163</td>
<td>0.3101</td>
</tr>
<tr>
<td>$\alpha_3$</td>
<td>0.9683</td>
<td>0.3033</td>
<td>0.0017</td>
</tr>
<tr>
<td>$\alpha_4$</td>
<td>-0.0630</td>
<td>0.0251</td>
<td>0.0129</td>
</tr>
<tr>
<td>$\alpha_5$</td>
<td>-0.0114</td>
<td>0.0236</td>
<td>0.6275</td>
</tr>
<tr>
<td>$\alpha_6$</td>
<td>0.3265</td>
<td>0.1756</td>
<td>0.0632</td>
</tr>
<tr>
<td>$\alpha_7$</td>
<td>0.9390</td>
<td>0.2315</td>
<td>0.0001</td>
</tr>
<tr>
<td>$\alpha_8$</td>
<td>0.0185</td>
<td>0.0232</td>
<td>0.4252</td>
</tr>
<tr>
<td>$\alpha_9$</td>
<td>-0.0443</td>
<td>0.0195</td>
<td>0.0232</td>
</tr>
<tr>
<td>$\alpha_{10}$</td>
<td>-0.0774</td>
<td>0.0183</td>
<td>0.0000</td>
</tr>
<tr>
<td>$\alpha_{11}$</td>
<td>-3.2004</td>
<td>0.8323</td>
<td>0.0001</td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>-0.4798</td>
<td>0.0493</td>
<td>0.0000</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>-0.0591</td>
<td>0.0493</td>
<td>0.2305</td>
</tr>
<tr>
<td>$\lambda_3$</td>
<td>-1.2266</td>
<td>0.3833</td>
<td>0.0016</td>
</tr>
<tr>
<td>$\lambda_4$</td>
<td>0.6379</td>
<td>0.2689</td>
<td>0.0188</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random Effects Coefficients</th>
<th>Value</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta$</td>
<td>-0.3773</td>
<td>-0.4170</td>
<td>-0.3375</td>
</tr>
<tr>
<td>$\sigma_u$</td>
<td>0.3502</td>
<td>0.3026</td>
<td>0.4053</td>
</tr>
<tr>
<td>$\sigma_e$</td>
<td>2.0625</td>
<td>1.8123</td>
<td>2.3474</td>
</tr>
</tbody>
</table>

Table 2: Estimated coefficients of the mixed model with heteroscedasticity
estimates and their confidence intervals (C.I.) are easily obtained from fixed effects and variance maximum likelihood (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 hospital control. As mentioned previously, the case-mix index 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.

$$
 e(x_r) = \frac{\partial y}{\partial x_r} \div \frac{y}{x_r} = \frac{\partial \ln(y)}{\partial \ln(x_r)},
$$

is the percentage change in output associated with a unitary percent change in the $r$-th 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 $e^W$.

For Beds we have:

$$
 e^W(Beds)_{ij} = \frac{\partial \ln(NHosp_{ij})}{\partial \ln(Beds_{ij})} =
$$

$$
 = [\beta_0 + \beta_1 \text{std}(\ln(CMix_{ij}))] \cdot [1 + 2\beta_2 \ln(Beds_{ij})] +
$$

$$
 + [\delta_0 + \delta_1 \text{std}(\ln(CMix_{ij}))] \cdot \ln(Staff_i). \tag{6}
$$
Analogously, for \( Staff \) we have:

\[
ed^W_{Staff ij} = \frac{\partial \ln(N\text{Hosp}_{ij})}{\partial \ln(Staff_{ij})} = [\gamma_0 + \gamma_1 \text{std}(\ln(CMix_{ij}))] \cdot [1 + 2\gamma_2 \ln(Staff_i)] + \\
+ [\delta_0 + \delta_1 \text{std}(\ln(CMix_{ij}))] \cdot \ln(Beds_{ij}).
\] (7)

In both cases, the elasticity depends on the three explicative variables of the model. As one can see in Appendix A (see Table 6), it is mainly affected by the case-mix value. In fact, given a positive index, the estimated elasticity of the hospitalizations number with respect to \( Beds \) decreases with \( Beds \), when \( Staff \) is fixed, and 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 the index, both elasticities are almost constant.

Tables 4 and 5, in Appendix A, summarize the estimated elasticity for \( Beds \) and \( Staff \) calculated for every observed macro units (i.e. at hospital level). Equation (5) 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:

\[
ed^H(Staff)_i = \frac{\partial \sum_j \frac{y_{ij}}{\text{Staff}_i} \frac{\text{Staff}_i}{\sum_k y_{ik}}}{\partial \ln \frac{\text{Staff}_i}{\sum_k y_{ik}}} = \sum_j \left( \frac{\partial y_{ij}}{\partial \text{Staff}_i} \frac{\text{Staff}_i}{\sum_k y_{ij}} \right) = \sum_j \left( \frac{\partial y_{ij}}{\partial \ln \frac{\text{Staff}_i}{\sum_k y_{ik}}} \frac{\text{Staff}_i}{\sum_k y_{ik}} \right) \sum_j e^W_{Staff ij} w_{ij},
\] (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:

\[
ed^H(Beds)_i = \sum_j \left( \frac{\partial \ln y_{ij}}{\partial \ln Beds_{ij}} \right) w_{ij} = \sum_j e^W(Beds)_{ij} w_{ij},
\] (9)

The estimates of elasticity at hospital level are summarized in Figure 4.
From this figure one can notice that, for Beds, the elasticity is concentrated around the unity. For Staff, however, the plot shows that quite all units work in over-dimensional 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 slightly interpretation of elasticity results as the local returns to scale (so called elasticity of scale)

$$e(x) = e(\text{Beds},i) + e(\text{Staff},i),$$

where $x$ indicates the input vector $(\text{Beds}, \text{Staff})$. 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; that is, the ratios $(x_r/x_s)$ are constant for all $r$ and $s$.

“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 underestimates 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 hospitals level results, an estimated scale elasticity under the unity identifies hospitals presenting a congested condition, where investments are not useful and it is better to reduce the dimension; on the contrary, hospitals that have an higher return to scale are the ones with unused productive capacity, which could increase their dimension. This interpretation could be very useful by the administrative point of view. Tables 4 and 5, in Appendix A, shows also the estimated scale elasticity for each hospital. It can be noted, however, that the differences among hospitals are due quite all to \(\text{Staff}\) elasticity. \(\text{Staff}\) and \(\text{Beds}\) values are also given in these tables in order to allow a straightforward interpretation of the estimated values.

Another interesting measure, related to the returns to scale 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 the most productive scale size (Banker, 1984), which is an input bundle $x$ characterized by $e(x) = 1$. “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, 1988). Ray developed an input-oriented measure of scale efficiency, directly obtainable from an empirically estimated single output multiple input translog production function. For example, regarding the translog model

$$
\ln(N Hosp) = \alpha_0 + \alpha_1 \ln(Beds) + \alpha_2 \ln(Staff) + \frac{\beta_{11}}{2} [\ln(Beds)]^2 + \beta_{12} \ln(Beds) \ln(Staff) + \frac{\beta_{22}}{2} [\ln(Staff)]^2
$$

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

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

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

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

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

where $\theta$ is the technical inefficiency (in our case we can obtain it from a MOLS or COLS transformation - Greene, 1993 - 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. Ray (1998) defines

\[
\sigma = \frac{S_0^*}{B_0} \cdot \frac{B_0}{S_0} = \frac{S_0^*}{S_0}. \quad (14)
\]

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:

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

that is 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 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 analyzed 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 more attention should be focused, however, on typical econometric problems, like for example 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 and Toby, 1998 and Barnett and 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 is the effect, the better is 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 sta elasticities. Both sta 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 sta elasticity, “Private and Public Clinics” present a large variability. “Research structures” are instead characterized by a low average sta elasticity and, consequently, worse productive conditions, maybe due to the different goal of these structures. Here sta is quite completely devoted to research and the health service is considered only as a secondary aim.
<table>
<thead>
<tr>
<th>Beds Elasticities</th>
<th>Structure Type</th>
<th>y0.01</th>
<th>y0.25</th>
<th>y0.5</th>
<th>y0.75</th>
<th>y0.99</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Obs. Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hospitals</td>
<td>0.971</td>
<td>1.015</td>
<td>1.032</td>
<td>1.051</td>
<td>1.088</td>
<td></td>
<td>1.033</td>
<td>0.025</td>
<td>118</td>
</tr>
<tr>
<td>Research Institutes</td>
<td>0.913</td>
<td>0.992</td>
<td>1.013</td>
<td>1.036</td>
<td>1.086</td>
<td></td>
<td>1.010</td>
<td>0.042</td>
<td>18</td>
</tr>
<tr>
<td>Classified Hospitals</td>
<td>1.010</td>
<td>1.022</td>
<td>1.025</td>
<td>1.042</td>
<td>1.072</td>
<td></td>
<td>1.035</td>
<td>0.022</td>
<td>5</td>
</tr>
<tr>
<td>Private and Public Clinics</td>
<td>0.919</td>
<td>1.028</td>
<td>1.068</td>
<td>1.099</td>
<td>1.317</td>
<td></td>
<td>1.067</td>
<td>0.074</td>
<td>51</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Staff Elasticities</th>
<th>Structure Type</th>
<th>y0.01</th>
<th>y0.25</th>
<th>y0.5</th>
<th>y0.75</th>
<th>y0.99</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Obs. Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hospitals</td>
<td>0.003</td>
<td>0.302</td>
<td>0.375</td>
<td>0.449</td>
<td>0.600</td>
<td></td>
<td>0.369</td>
<td>0.123</td>
<td>118</td>
</tr>
<tr>
<td>Research Institutes</td>
<td>-0.045</td>
<td>0.187</td>
<td>0.265</td>
<td>0.354</td>
<td>0.703</td>
<td></td>
<td>0.282</td>
<td>0.183</td>
<td>18</td>
</tr>
<tr>
<td>Classified Hospitals</td>
<td>0.269</td>
<td>0.321</td>
<td>0.353</td>
<td>0.426</td>
<td>0.563</td>
<td></td>
<td>0.387</td>
<td>0.104</td>
<td>5</td>
</tr>
<tr>
<td>Private and Public Clinics</td>
<td>-0.326</td>
<td>0.253</td>
<td>0.478</td>
<td>0.604</td>
<td>1.297</td>
<td></td>
<td>0.453</td>
<td>0.303</td>
<td>51</td>
</tr>
</tbody>
</table>

Table 3: Elasticity summary statistics of hospitalization structures clusters

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. For a few diseases with a lower case-mix, however, the demand can be affected by single hospital marketing strategies. As one can see in Table 6 of Appendix A, substantial changes in case-mix index cause the raise of different elasticity patterns.

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

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.
References


Ryan, D.L. and Wales, T.J. (2000). Imposing local concavity in the translog and gen-

Table 4: Estimated elasticities computed with respect to each input variable and scale elasticity
Table 5: Estimated elasticities computed with respect to each input variable and scale elasticity
Table 6: Estimated elasticities computed with respect to each input by given values of case-mix index. Negative values of case-mix index corresponds 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.