Inefficiency in Childcare Production Evidence from Italian Microdata

Luigi Brighi, Paolo Silvestri

CAPP*aper* n. 135
marzo 2016

Università di Modena e Reggio Emilia Facoltà di Economia Marco Biagi
Università di Bologna Dipartimento di Scienze Economiche

CAPP - Centro di Analisi delle Politiche Pubbliche
Dipartimento di Economia Politica - Università di Modena e Reggio Emilia
Ufficio 54 - Ala Ovest
Viale Berengario, 51 41100 Modena - ITALY
phone: +39 059 2056854 fax: +39 059 2056947
email capp@unimo.it

Inefficiency in Childcare Production. Evidence from Italian Microdata

Luigi Brighi[1] Paolo Silvestri[2]

March 2016

Abstract

The purpose of the paper is to study inefficiency in the production technology of the childcare service and to carry out a comparative analysis of public and private day-care centres. An empirical analysis on cross-section micro-data from a region of northern Italy has been conducted by using an input-distance function with a translog specification. Estimates of the multi-output production technology and input-oriented technical inefficiency are obtained in a stochastic frontier model with a half-normally distributed one-sided error. Heteroscedasticity has been modelled to investigate the determinants of inefficiency and estimate their marginal effects. We find that production exhibits increasing returns with an estimated elasticity of scale of 1.21. Separability between inputs and outputs is rejected at a 5% level of significance. The average estimate of technical inefficiency is 10% and public centres are more inefficient than private centres by 4.1 percentage points. The proportion of part-time children and the presence of mixed-age classrooms are significant determinants of inefficiency which equally affect both public and private centres.

JEL Classification Number: D.24, H.44, J.13

Keywords: Childcare, Day-Care Center, Technical Inefficiency, Stochastic Frontier Analysis, Input-Distance Function

[1]Dipartimento di Economia "Marco Biagi" and RECent, Università di Modena e Reggio Emilia, via Berengario 51, 41121 Modena, Italy, e-mail: luigi.brighi@unimore.it

[2]Dipartimento di Economia "Marco Biagi" and CAPP, Università di Modena e Reggio Emilia, via Berengario 51, 41121 Modena, Italy, e-mail: paolo.silvestri@unimore.it

1 Introduction

Childcare services may have relevant effects on several socio economic problems. Childcare may improve female labour market participation and gender equality by reconciling work and family life; it may contrast the negative trend of fertility rates by making a child less costly in terms of income and career opportunities and, from a socio-pedagogical perspective, it may contribute to child development and socio-economic integration. In the last decades the importance of childcare has been increasingly recognized and most of the developed countries, including Italy, have taken initiatives to increase the availability of the service.

The growth of childcare, which were supplied almost exclusively by the public sector in the past, has been achieved by promoting the private provision of the service. In fact, given budgetary constraints to public funding, the expansion of childcare has been accomplished at lower costs by increasing the presence of private providers which are able to benefit from less costly labour contracts than those of the public sector. An interesting question is to ask whether the opening of the public service to competition from the private sector, besides attaining cost savings due to lower input prices, has also stimulated efficiency gains in production. It would be also interesting to know whether the competitive pressure has reduced the differences in the productive performance between public and private providers and removed the major sources of inefficiency. Our paper is a first attempt to address this issues by providing a comparative analysis of inefficiency in the production of the childcare service between public and private providers.

This paper deals with childcare services for young children defined as *infants*, from 3 to 12 months of age and *toddlers*, from 12 to 36 months. The analysis refers to infant and toddler day-care centres in Emilia Romagna, one of the most developed regions of northern Italy which is also known for the high share of children served and the quality of childcare.[1] Emilia Romagna is the first region in Italy as for the territorial coverage of the service, which is available in more than 80% of the municipalities, and

[1] In Emilia Romagna is active *Reggio Children* which is a worldwide recognized centre promoting innovative pedagogical projects for early childhood. The Reggio Children approach has been carried forth in infant and toddler centres in Reggio Emilia since the 90's and it is shared by many centres

it is the only Italian region to meet the Barcelona 2002 target of providing childcare for more than 33% of children under 3 years of age.

Childcare is regulated by regional authorities in order to guarantee the quality of the service. The main quality requirements are concerned with the proportion of children to teachers, the group size of classrooms, the education of teachers as well as the space for green and covered area. Structural quality standards are enforced by regional legislation and apply to both public and private facilities. The regional childcare system consists of a number of formal arrangements including day-care centres, micro-centres and age-integrated institutions offered by both public and private subjects. For the age group 0-2 years, public childcare is the most common form of provision of the service and day-care centres are the most widespread formal arrangement. Public childcare is administered at the municipal level. The 'owner' of the service is the municipality, which means that the access to the day-care service is managed by the local public authority. The provision of the service is either *public*, when the center is directly operated by the municipality or *private*, when the center is operated by private subjects such as cooperatives or other social organizations.[2]

The focus of our work is the empirical analysis of the childcare service for infant and toddlers provided by day-care centres. Our objectives are to estimate production technology and inefficiency and to carry out a comparative analysis of public and private centres by identifying the major determinants of inefficiency and measuring their marginal effects. Our analysis will be addressed to the production technology of the service and no account will be taken of possible quality differences. Although quality may somewhat differ according to the private or public nature of the provider, quality differences in the provision of a regulated service do not seem to be particularly relevant, as far as the production technology is concerned, and can be neglected as a first approximation.[3] Moreover, we shall consider only *technical* inefficiency and exclude from the analysis inefficiencies arising from the choice of non optimal

in other countries.

[2]To simplify, we will use the term public center if the center is owned and operated by the municipality and private center if the center is owned by the municipality and operated by a private subject.

[3]This remark seems to be consistent with our data, as shown in Section 3.

allocations of resources given prices, which means that our estimate is to be considered as a lower bound to cost inefficiency.

Infant-toddler centres have a production technology with multiple inputs and outputs. The most important types of inputs are the labour services provided by the teaching staff and the service personnel, i.e. the workers involved in the provision of auxiliary services such as food, janitorial and laundry services. Capital input mainly consists of the services of physical buildings and green areas in addition to variable capital such as energy and heating. Two separate outputs, one for infant service and one for toddler service, are considered since the production of childcare services for infant and toddlers has different requirements in terms of the type and the quantity of labour inputs and in terms of food service arrangements.

There are a few important modes of operation of the day-care service that are worth considering. The service can be organized by age group or by mixed-age group classrooms, where children of different ages, usually toddlers, are cared for together. The legislation places requirements on the number of children per teacher differentiated by age category and, specifically, it requires that there must be 5 infants per teacher and 7 toddlers (or 10 older toddlers) per teacher. Mixed-age group classrooms must comply with the more stringent requirement. Another organizational choice is whether to have classrooms arranged on a full-time basis and/or classrooms arranged on a part-time basis. The daily opening hours of full-time service are at least 8 hours, while part-time hours must be no less than 6. Centres with no part-time classes may as well provide part-time service accommodating children in full-time classes, although this involves a suboptimal use of resources. In order to meet the needs of working parents in the first or the final hours of the day, centres may either decide to extend the standard hours of full-time or part-time service, or to provide extra opening hours. Usually, children attendance in the extra hours is limited and the service can be arranged with a smaller number of classes.

The production of the day-care service also involves other auxiliary activities such as food preparation, laundry, janitorial activities and pedagogical coordination. These activities can either be carried out in-house or can be externalized. Usually, outsourcing the food service means buying meals outside, while the outsourcing of the other auxiliary services is carried out by hiring external workers who give their labour services internally.

The analysis of technical efficiency carried out in this paper is input-oriented, i.e. it is focused on the potential savings of inputs given outputs. The quantities of output are not under complete control of centre administrators. Indeed, the number of enrolled infants and toddlers depends, to a large extent, on the decision of municipalities and, similarly, the number of children served on a part-time or full-time basis is the result of parents' choices. On the other hand, the centre administrators have relatively greater control over input choices. For example, they have some discretion about the number, the working hours and the qualification of workers to employ in teaching activities and in auxiliary services. They decide whether or not to organize mixed-age classrooms and whether or not some classrooms are arranged on a part-time basis. They also decide whether and to what extent externalize auxiliary services. Thus, it seems sensible to assume that the objective of the centre administrator, who is usually under a tight budget constraint, is to economize on the use of inputs, so that the analysis of technical efficiency is best conducted in terms of input-oriented measures, i.e. by looking at the potential saving of physical inputs to produce a given combination of physical outputs.

The empirical analysis refers to a sample of 482 infant-toddler centres of Emilia Romagna. The regional survey was carried out between 2007 and 2008 and contains very detailed information about the characteristics and the organization of the service at the level of the single day-care center. To model the multi-output technology of production, an input-distance function approach has been used with a translog functional form specification. The empirical model is a stochastic frontier model with a one-sided error specification which represents technical inefficiency. The one-sided error is assumed to follow a half-normal distribution. Three nested models have been estimated by using the Maximum likelihood method. The first two models are used for comparative purposes, while the third provides the final results. In the final model, heteroscedasticity of all the error terms is parametrized by a number of exogenous variables which are plausible determinants of inefficiency. Our empirical study provides estimates of production technology and of technical efficiency and inefficiency of the day-care service. We also identify the most important determinants of inefficiency, which include the public or private nature of the provider, and evaluate the magnitude of the marginal effects of these determinants.

There are not many empirical studies investigating the characteristics of production technology of the day-care service. Most of these studies are based on costs data and recover the properties of technology indirectly from the dual relationship between production and costs. The evidence is not univocal. Using a sample of 182 centres in 35 states of the U.S. conducted in 1989, Powell and Cosgrove (1992) estimate a cost function with a single output by controlling for quality differential as well as other centre characteristics.The empirical results show that technology exhibits increasing returns to scale and that private for profit centres are more cost efficient (by 9.1 percentage points) than other centres, once quality has been controlled for. Mukerjee and Witte (1993) compare the costs of Non For Profit (NFP) and For Profit (FP) centres in Massachusetts using cross-section data collected in 1987 and 1988. They conclude that there are not significant differences as for the method of operation of the two types of centres and that differences of costs are due to differences in wages paid by the two types of employers. A similar conclusion is reached by Preston (1993), who uses a large data set on costs from a National survey in U.S. in 1976-77. Differences in costs between FP and NFP centres are associated with differences in the quality of the service, however there is no evidence that the two types of centres have different level of efficiency. Mocan (1997) studies a multi-product cost function and estimates several properties of the production technology. Mocan's empirical analysis is based on data from a sample of 400 observations collected in different states of the U.S. in 1993. The main results are that there are no significant differences either in the quality of the service or in the level of efficiency between FP and NFP centres. Finally, a different approach is used in Bjurek *et al.* (1992) who study productive efficiency in public day-care centres in Sweden by using the non parametric method of the deterministic production frontier, or Data Envelopment Analysis (DEA). They estimate a multi-output production frontier with a data set of 194 observations collected in the years 1988 and 1989 and obtain measures of (output-oriented) technical efficiency of 0.89-0.91, meaning that actual day-care services in the public centres could have been increased by 9-11%.

Our empirical analysis departs from previously published work under many respects. First, our analysis is based on production rather than cost data. As it is well known, analyses of technology based on dual properties of cost functions rest on the assumption of cost minimization, which may not be fully appropriate for the day-care sector where most of the services are supplied by public providers. On the other hand, in our model a multi-output technology of childcare is estimated directly by using a distance function and without resorting to costs. Second, we estimate a stochastic rather than a deterministic production frontier. In deterministic frontier analysis little or no account is usually taken of measurement errors or other sources of statistical noise so that all the deviations from the frontier are considered as the result of inefficiency. By contrast, stochastic frontier models do not incur the risk to overestimate inefficiency. Third, heteroscedasticity of the error terms has been explicitly considered and incorporated in our research. It must be noticed that the inefficiency measures are based on the residuals of the frontier estimates which are particularly sensitive to specification errors associated with the presence of heteroscedasticity. Indeed, in such a case estimators not only are no longer efficient but are also biased, as shown by Kumbhakar and Lovell (2000), and efficiency measures are seriously affected. Fourth, our study seems to be the first piece of research focusing on infant and toddler day-care sector, which uses one of the largest and most systematic data set covering a population of centres with a high degree of homogeneity in the production technology and regulation.

Finally, our research has also similarities with other work. For instance, the multi-output representation of technology is similar to the model used by Coelli and Perelman (2000) in their study on the railways sector and we apply a single-step procedure for investigating the determinants of inefficiency as in Caudill *et al.* (1995) and Hadri (1999).

The rest of the paper is organized as follows. The next section introduces the notions of technical inefficiency, input-distance function and specifies the empirical models. Section 3 illustrates the main characteristics of the sample by comparing public and private centres and provides the definition of the variables as well as some descriptive statistics. Section 4 presents and discusses the empirical results and finally section 5 provides a summary and conclusions.

2 Inefficiency and model specification

A production technology is a set of feasible plans of production, i.e. input-output combinations, (x, y), where the positive vector of input x can be transformed into the positive vector of output, y. A production plan is technically efficient if there are not any other input-output combinations with higher amounts of some output and lower amounts of some input. The set of all the efficient plans is the *frontier* of production, i.e. the set of input-output combinations lying on the boundary of the technology set. Feasible production plans which do not belong to the frontier are technically inefficient, meaning that either a higher level of output can be obtained with the same input or a lower quantity of inputs can be used to produce the same output or both.

The analysis of efficiency conducted in the present investigation is *input-oriented*, because it refers to the quantities of inputs needed to produce a given vector of output. In order to define a measure of inefficiency let us introduce the notion of input requirement set. For any given output vector y the *input requirement set* $V(y)$ is the set of input vectors that can produce the output y. The boundary of $V(y)$ is the isoquant of y, i.e. the set of efficient input combinations for producing y. Figure 1 shows the input requirement set of a feasible production plan (x, y) with two inputs and the isoquant associated with y. The input vector $\hat{x}$ in Figure 1 is proportional to x and lies on the isoquant associated with y, i.e. $\hat{x}$ is an efficient input vector for producing the output vector y obtained by a proportional reduction of all the inputs. The distance between x and $\hat{x}$ is the basis for the measure of technical inefficiency of the production plan (x, y).

If (x, y) is a feasible plan, i.e. $x \in V(y)$, the vector $\hat{x}$ is a scalar multiple of x and the coefficient of proportionality (which is not greater than 1) indicates the proportion of inputs needed to efficiently produce the given output vector. Graphically, the

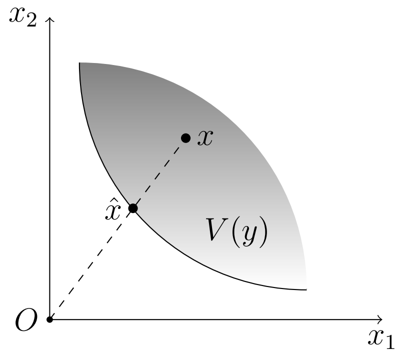

Figure 1: Input-oriented technical inefficiency and efficiency

coefficient is given by the ratio of the distance $0\hat{x}$ to the distance $0x$ in Figure 1. Writing $\hat{x}$ as $\hat{x} = e^{-u}x$, the coefficient e^{-u} can be taken as a measure of efficiency and the scalar u as a measure of inefficiency. Indeed, if x lies on the isoquant, the plan (x, y) is efficient, $e^{-u} = 1$ so that $u = 0$. i.e. efficiency is equal to 1 and inefficiency is equal to zero. On the other hand, if the plan (x, y) is inefficient, the input vector is inside $V(y)$ and $e^{-u} < 1$ implies $u > 0$, i.e. efficiency is less than 1 and inefficiency is strictly positive.

Let us then define[4] the *Input-Oriented Technical Efficiency* associated with the feasible production plan (x, y) by

$$TE = e^{-u} \tag{1}$$

where e^{-u} is the coefficient of the input vector contraction lying on the isoquant associated with y. TE ranges from 0 to 1 and its value indicates the proportion of the observed inputs needed to efficiently produce the given level of output.

On the other hand a measure of technical inefficiency is given by the proportion of inputs used in excess of the efficient vector $\hat{x}$, which is $(1 - e^{-u})$ and is represented in Figure 1 by the ratio of the distance $x\hat{x}$ to the distance $0x$. Therefore, the *Input-Oriented Technical Inefficiency* associated with the feasible production plan (x, y) is

[4]The measure of technical efficiency widely adopted in the literature and used in the present work refers to the early contributions of Farrel (1957). For more details on efficiency measures see also Coelly *et all.* (2006).

defined by

$$TI = 1 - e^{-u} \approx u \tag{2}$$

where we used the first order approximation at $u = 0$, $e^{-u} = 1 - u$. The scalar u is non negative and its value indicates the rate at which all inputs can be proportionally reduced without reducing output. It must be noticed that u is a measure of *technical* inefficiency because it does not take account of changes in input proportions required by cost minimization behaviour. The same caveat also apply to the measure of technical efficiency e^{-u}.

A convenient way to model inefficiency in multiple output production technologies is by means of the *input-distance function* developed by Shephard (1970). The (input) distance function is defined for all the production plans and is given by

$$D(x,y) = \max\{\delta \in \mathbb{R}_{++} \mid \frac{1}{\delta}x \in V(y)\} \tag{3}$$

The distance $D(x,y)$ is the reciprocal of the proportional contractions or expansions of inputs needed to reach the isoquant associated with y. If x lies on the isoquant, no contraction of inputs is needed so that $D(x,y) = 1$. When x is an interior point of $V(q)$, as depicted in Figure 1, a contraction of inputs is required to reach the boundary and $D(x,y) > 1$. If (x,y) is not a feasible plan, then greater quantities of inputs are needed to produce y and an expansion of the input vector is required to reach the boundary of $V(y)$, thus $D(x,y) < 1$.

The distance function has several properties which derive from standard assumption on technology. First, $D(x,y) \geq 1$ characterizes the production set and specifically the set of solutions of $D(x,y) = 1$ is the frontier of production. Moreover, the distance function is linearly homogeneous in the input x, i.e. $D(tx,y) = tD(x,y)$ for $t > 0$, it is increasing and concave in x and decreasing and quasi-convex in y.[5] However, one of the most useful aspects of the distance function is its relationship with technical efficiency. As can be seen from (3), the distance function is the reciprocal of technical efficiency, $e^{-u} = 1/D(x,y)$, thus taking the logs, it turns out that technical inefficiency is equal to the log of the distance, i.e.

$$u = \log D(x,y) \tag{4}$$

[5]For a detailed analysis of the properties of $D(x,y)$ see Fare and Primont (1995).

In order to specify an empirical model of inefficiency a parametric functional form of the input-distance function must be chosen. The translog functional form is particularly convenient because it is linear in the parameters and provides a local second order approximation to any arbitrary function. Therefore, we assume that the distance function of the production technology of infant-toddler centres with two outputs and three inputs is the following[6]

$$\begin{aligned}\log D(x, y) \quad = \quad & \alpha_0 + \sum_{h=1}^{3} \alpha_h \log x_h + \frac{1}{2} \sum_{h=1}^{3} \sum_{k=1}^{3} \alpha_{hk} \log x_h \log x_k + \sum_{m=1}^{2} \beta_m \log y_m + \\ & + \frac{1}{2} \sum_{m=1}^{2} \sum_{n=1}^{2} \beta_{mn} \log y_m \log y_n + \sum_{h=1}^{3} \sum_{m=1}^{2} \gamma_{hm} \log x_h \log y_m \qquad (5)\end{aligned}$$

The properties of distance functions require α_h to be positive and β_m to be negative. Moreover, the parameters must satisfy the symmetry restrictions, $\alpha_{hk} = \alpha_{kh}$ for $h, k = 1, 2, 3$ and $\beta_{mn} = \beta_{nm}$ for $m, n = 1, 2$. Homogeneity of degree one in inputs imposes the following restrictions

$$\alpha_1 + \alpha_2 + \alpha_3 = 1 \qquad (6)$$

$$\alpha_{h1} + \alpha_{h2} + \alpha_{h3} = 0 \qquad \text{for} \quad h = 1, 2, 3 \qquad (7)$$

$$\gamma_{1m} + \gamma_{2m} + \gamma_{3m} = 0 \qquad \text{for} \quad m = 1, 2 \qquad (8)$$

Homogeneity restrictions are more easily placed on (5) if one normalizes inputs by x_3, which gives $x_3^{-1} D(x, y) = D(\tilde{x}, y)$, where $\tilde{x} = (x_1/x_3, x_2/x_3, 1)$. Taking the logs and rearranging terms yields

$$-\log x_3 = \log D(\tilde{x}, y) - u \qquad (9)$$

where, by (4), the inefficiency measure u has been substituted for $\log D$. Adding in (9) a stochastic, zero mean, symmetric error term v, which accounts for measurement

[6] We choose as a point of approximation of the distance function the sample mean of input and output variables. Each variable is divided by its mean so that the first-order parameter can be interpreted as elasticity at the mean value.

errors and other sources of statistical noise, yields the following *stochastic production frontier* model

$$-\log x_3 \;=\; \log D(\tilde{x}, y) + v - u \tag{10}$$

where $u \geq 0$ and $\log D(\tilde{x}, y)$ is given by

$$\begin{aligned}\log D(\tilde{x}, y) \;=\;& \alpha_0 + \sum_{h=1}^{2} \alpha_h \log \tilde{x}_h + \frac{1}{2} \sum_{h=1}^{2} \sum_{k=1}^{2} \alpha_{hk} \log \tilde{x}_h \log \tilde{x}_h + \sum_{m=1}^{2} \beta_m \log y_m + \\ & + \frac{1}{2} \sum_{m=1}^{2} \sum_{n=1}^{2} \beta_{mn} \log y_m \log y_n + \sum_{h=1}^{2} \sum_{m=1}^{2} \gamma_{hm} \log \tilde{x}_h \log y_m \end{aligned} \tag{11}$$

Once the parameters of (11) are estimated the remaining parameters of (5) are recovered from the symmetry and homogeneity restrictions (6), (7) and (8).

To estimate the parameters of the model, further assumptions about the unobserved terms v and u are introduced. For any observation i, the inefficiency term u_i is treated as a random variable and specifically as a one-sided error term with a distribution in the non negative domain, i.e. with $u_i \geq 0$. The most common distributional assumption in the literature starting from the original contribution of Aigner *et al.* (1977) is that u_i is distributed as a *half-normal*. The half-normal is the non negative truncation of a zero mean normal distribution, it is characterized by a single parameter and thus it is relatively easy to estimate. The half-normal distribution is denoted by $N^+(0, \sigma_u^2)$, where σ_u^2 is the variance of the normal distribution before truncation.[7] We assume that the inefficiency terms are independently distributed across observations and have a half-normal distribution,

$$u_i \;\sim\; N^+(0, \sigma_u^2). \tag{12}$$

The random errors v_i are supposed to be independently and normally distributed with zero mean, i.e.

$$v_i \;\sim\; N(0, \sigma_v^2). \tag{13}$$

[7] Notice that, although the distribution has zero mode, the mean of u_i is different from zero and $\mathrm{Var}(u_i)$ is not equal to σ_u^2. Indeed, $E(u_i) = \sigma_u \sqrt{2/\pi}$ and $\mathrm{Var}(u_i) = \sigma_u^2(\pi - 2)/\pi$. For more details on the half-normal distribution see Kumbhakar *et al.* (2015) and the references therein.

For each observation i, the random variables v_i and u_i are independent and uncorrelated with the explanatory variables.

The stochastic frontier model (5) and the distributional assumptions on the unobservable terms, (13) and (12), are referred to as the *half-normal model*. Our aim is to estimate the production frontier coefficients α, β and γ, the variance σ_v^2 and, in particular, the parameter σ_u^2 of the inefficiency distribution. Indeed, the estimate of σ_u^2 together with the distributional hypothesis on u_i yield observation-specific estimates of technical efficiency and inefficiency, which is the focus of our analysis.

The parameters of the the half-normal model are estimated by using the Maximum Likelihood (ML) method. The assumptions about u_i and v_i are used to derive the distribution of the *composed error* term, $\epsilon_i = v_i - u_i$, and thus the log-likelihood for each observation, which is[8]

$$L_i = -\log\left(\frac{1}{2}\right) - \left(\frac{1}{2}\right)\log(\sigma_v^2 + \sigma_u^2) + \log\phi\left(\frac{\epsilon_i}{\sqrt{\sigma_v^2 + \sigma_u^2}}\right) + \log\Phi\left(\frac{\mu_{*i}}{\sigma_*}\right)$$

where ϕ and Φ are, respectively, the probability density and the probability distribution functions of the standard normal and

$$\mu_{*i} = \frac{-\sigma_u^2\epsilon_i}{\sigma_v^2 + \sigma_u^2} \tag{14}$$

$$\sigma_* = \frac{\sigma_u^2\sigma_v^2}{\sigma_v^2 + \sigma_u^2} \tag{15}$$

The ML estimates are obtained by numerical optimization of the sum of the log-likelihood of each observation. Two cases are considered. The first case is the estimate of the half-normal model with *homoscedastic* error terms, i.e. with constant σ_v^2 and σ_u^2. To guarantee that the variance estimates are positive, the parametrization by means of an exponential function is used, i.e.

$$\sigma_u^2 = \exp(\delta_0) \tag{16}$$

$$\sigma_v^2 = \exp(\theta_0) \tag{17}$$

where δ_0 and θ_0 are unrestricted scalars. The values of the variance estimates are recovered by substituting the parameter estimates into the above formulae.

[8] See Stevenson (1980) or Kumbhakar *et al.* (2015).

In the second case, a half-normal model with *heteroscedasticity* in the error terms is considered. As shown by Caudill and Ford (1993) and Kumbhakar and Lovell (2000), ignoring heteroscedasticity may severely bias the estimate of the frontier and by this way the estimates of technical inefficiency. Following the literature[9] we let the variance v_i and the pre-truncated variance of u_i to depend on exogenous observable variables. Therefore, σ_v^2 and σ_u^2 are parametrized by using exponential functions and are given by

$$\sigma_{ui}^2 = \exp(\delta_0 + \delta z_i) \tag{18}$$

$$\sigma_{vi}^2 = \exp(\theta_0 + \theta w_i) \tag{19}$$

where z_i and w_i are vectors of observable variables and δ and θ are the associated vectors of parameters.

The measure of technical inefficiency for each observation i is computed as the expected value of u_i conditional on the composed error ϵ_i, according to the formula provided by Jondrow *et al.* (1982)

$$E(u_i \mid \epsilon_i) = \frac{\sigma_* \, \phi\left(\frac{\mu_{*i}}{\sigma_*}\right)}{\Phi\left(\frac{\mu_{*i}}{\sigma_*}\right)} + \mu_{*i} \tag{20}$$

where μ_{*i} and σ_* are as in (14) and (15). Similarly the observation-specific technical efficiency, computed as in Battese e Coelli (1995), is given by

$$E(e^{-u_i} \mid \epsilon_i) = \frac{\Phi\left(\frac{\mu_{*i}}{\sigma_*} - \sigma_*\right)}{\Phi\left(\frac{\mu_{*i}}{\sigma_*}\right)} \, e^{-\mu_{*i} + \frac{1}{2}\sigma_*^2} \tag{21}$$

In the heteroscedastic half-normal model, the examination of the exogenous determinants of inefficiency is conducted through the analysis of the variables z_i affecting the pre-truncated variance of u_i in (18). The *marginal effect* of each exogenous variable z_{ik} on the expected value of observation-specific inefficiency, is given by the derivative

$$\frac{\partial E(u_i)}{\partial z_{ik}} = \delta_k \sigma_{ui} \phi(0) \tag{22}$$

[9] See, for example, Caudill, Ford and Gropper (1995) and Hadri (1999).

Since $\phi(0) > 0$, the sign of the marginal effect is the same as the sign of the δ_k coefficient.

3 Sample and data

In their activity of licensing and monitoring the quality of childcare, regional authorities carry out periodic surveys collecting information at level of detail of day-care centres. The administrator of each day-care facility is required to answer a questionnaire providing information about characteristics of children, teaching staff, service personnel and the outsourcing of auxiliary services. Our empirical analysis is based on cross section micro-data derived from the regional survey carried out in 2007-8.[10]

Different kinds of day-care facilities, such as infant-toddler centres, micro-centres and age-integrated institutions, are distinguished in the survey. From the point of view of the production technology, these kinds of facilities differ in the scale of operation and the output mix as well as the provision of complementary services and the use of the work force. In order to deal with a less uneven technology, we have decided to concentrate our attention on the core service represented by infant-toddler centres, which covers the large majority of public childcare services in the region.

The database contains 626 observations (infant-toddler centres). The centres with access managed by a municipality are 515, three quarter of which are public, i.e. directly operated by the municipality and the remaining quarter are private, i.e. entrusted by the municipality to cooperatives or other private subjects. Dropping non municipal centres, incomplete observations and extreme outliers gave us a sample size of 482 municipal centres, 361 of which are public and 121 are private. The final selected sample retains the same proportion of public to private centres and covers 98.6% of the total number of children enrolled in municipal infant-toddler centres in the region.

[10]Detailed information about the survey is available at the following link: `http://sociale.regione.emilia-romagna.it/infanzia-adolescenza/approfondimenti/osservatorio-infanzia-e-adolescenza/i-dati-e-le-statistiche/i-dati-dei-nidi-dinfanzia`

Some descriptive statistics of the sample composition, which also show the main significant differences between public and private centres, are presented in Table A.1 of the Appendix. Center size is defined with respect to the capacity, i.e. the number of children who can be accommodated. The center is considered small if capacity does not exceed 35, medium if capacity is between 35 and 60 and large if capacity exceeds 60. Medium or large size prevails in public centres (77%), while more than 80% of private centres have medium or small size. More than 60% of private centres provide only toddler service, while more than 70% of public centres serve both infants and toddlers. About 45% of private centres have mixed-age classrooms, which is twice as much the percentage for public centres. Not all centres have children with disability and typically the children with special needs are more concentrated in public facilities. Three-quarters of the centres have only full time classes and more than 70% provides extra daily opening hours of service. This is true for both public and private facilities with no significant differences. Table A.1 also shows that the proportion of public centres producing all auxiliary services in-house is 30%, as compared to only 12% of private centres. Most services are produced in-house by both kinds of centres except for food services which are externalized by 80% of private facilities and by only 40% of public centres.

In short, the main differences can be summarised as follows. Public centres have medium to large size, typically serve both infants and toddlers, most of them do not have mixed-age classrooms and half of them hosts children with disabilities. The large majority prepares meals in-house. On the other hand, private centres have medium to small size, typically serve only toddlers and almost half of them have mixed-age classrooms. The large majority externalizes the food service.

The main average characteristics of the day-care service by type of provider are presented in Table A.2 of the Appendix. Public centres have a significant larger number of children and classrooms than private centres, while average group size per class is the same (17 children). Public centres also have a significantly higher proportion of infants and of children with disabilities, which is matched by a lower child to teaching staff ratio as compared to the private centres.[11] In some studies

[11]The child staff ratio is given by the ratio of children to full-time equivalent teachers.

the child-staff ratio and the group size of classrooms are considered as measures of the structural quality of the service. In our sample there does not seem to be any difference in structural quality between public and private centres. In fact, the group size of classrooms is the same and the higher child to staff ratio for private centres is explained by the higher proportion of toddlers who have lower requirements in terms of teachers.

The average proportion of part-time children is higher in private centres than in their public counterpart, although the difference is not statistically significant. The daily hours of full time service are almost the same for both types of centres, while extra hours of service are higher in private facilities. Another difference between public and private centres is due to differences in labour contracts; in fact, the weekly working time is shorter in the public sector, where it is also easier to get a part-time contract. This difference is also revealed by our data, where we notice that, although 85% of staff in both public and private centres has weekly hours exceeding 24 hours, more than 30% of workers of the private centres work for more than 36 hours, while this figure is only 2% for public centres workers. These differences in working hours are also reflected in a higher number of teachers and a higher fragmentation index, i.e. the ratio of teachers to full-time equivalent teachers, in public centres.

To sum up, the main differences between public and private providers are that public centres have a larger number of children and classes, a higher proportion of infants and of children with disabilities. There are no significant differences as to the proportion of part-time children and the structural quality of the service is similar. Private centres have a larger number of extra hours of service, while public facilities have a higher fragmentation of the labour-force.

We end this section with a description of the variables used in our empirical investigation. Their main descriptive statistics are shown in Table 1.

Input variables

Two labour inputs are considered, the labour of the teaching staff and the labour of the service personnel employed by the administrator of the center or by external contractors. The teaching staff also includes assistant teachers, staff used to extend the opening hours and the staff used to care for children with disabilities. The service personnel includes workers providing auxiliary services such as food, janitorial and laundry services. The two labour input variables are measured by taking the sum of weekly hours of the teaching staff and the sum of weekly hours of the service personnel. Since the labour embodied in the outsourced food service is not directly observed, a figurative amount of service labour input has been imputed to the centres purchasing catered food service. The imputed service labour input was computed by assuming direct proportionality between hours of work and the number of meals served. The two types of labour input, teaching and service, are respectively denoted by `edu` and `ser`. Capital input is mainly given by physical buildings and green areas, furnitures and equipments. Since the survey does not provide measures of physical capital (e.g. square meters), we assume a direct proportionality relationship between the quantities of capital goods and the maximum number of children who can be hosted. Therefore, the capital input is measured by the capacity of the day-care facility and is denoted by `cap`.

Output variables

The infant output and the toddler output are separately measured by the respective total number of weekly child-hours. The daily opening hours are computed as the weighted average of the hours of service of full-time classes and part-time classes and include the extra hours of service. The weekly hours are computed over a five days working week and the outputs are then calculated by multiplying the average weekly opening hours by the number of children in each age group. The infant output and the toddler output are respectively denoted by `yi` and `yt`.

The following variables, representing particular aspects of the organization of the day-care service or relevant attributes of the centres, will be used in our empirical work to capture efficiency differences across centres.

Other variables

`pub` is a dichotomous variable which takes the value of 1 if the center is public, i.e. operated by the municipality. `small` is a dummy indicating the centres with capacity less than 35. `pro` is a dichotomous variable indicating whether the center

Table 1: Input, output and other variables

Variable	**Description**	**Mean**	**Std. Dev.**	**Min.**	**Max.**
yi	weekly infant-hours	225.27	255.93	0	1958.45
yt	weekly toddler-hours	2007.80	865.75	150	4352
cap	capacity	48.89	18.45	12	116
ser	service pers. weekly hours	130.01	59.48	13.21	332.3
edu	staff weekly hours	278.98	126.26	32.5	705.5
pub	Dummy public centre	0.749	0.434		
small	Dummy small centre	0.284	0.452		
pro	Dummy capital of province	0.407	0.492		
mix	Dummy mixed-age classes	0.309	0.463		
disa	Dummy children with disab.	0.355	0.479		
ptr_ft	Part-time ratio (no p-t classes)	0.064	0.120	0	0.625
exh	Extra hours of service	1.130	1.016	0	3.3
food_in	Dummy food in-house	0.508	0.500		
frag	Teachers to f-t-e teachers	1.147	0.153	0.9	2.16

is located in a capital of province, where the survey is likely to have been conducted more accurately. `ptr_ft` is the ratio of part-time children to total enrolled children in centres with no part-time classes; this variable is equal to zero in centres with both part-time and full-time classes. `mix` is a dummy indicating the centres with mixed-age classrooms. `disa` is a dichotomous variable which takes the value of 1 if children with disabilities are enrolled at the centre. `exh` is the number of daily extra hours of service provided by the centre. `food_in` is a dummy variable identifying the centres which do not externalize the food service, i.e. the most important auxiliary service. `frag` is an index of fragmentation of the labour force which is given by the ratio of teachers to full-time equivalent teachers (normalized at 36 hours per week).

4 Empirical results

The day-care service technology is estimated using the data described in Section 3. The model is defined with three input variables, $x_1 = \texttt{cap}$, $x_2 = \texttt{ser}$ and $x_3 = \texttt{edu}$, and two output variables, $y_1 = \texttt{yi}$ and $y_2 = \texttt{yt}$. Each variable is divided by its mean, inputs are normalized by `edu` and finally the natural logs are taken. We denote by `nledu` the negative of the log of teaching staff hours, by `lcap` the normalized log of capacity and by `lser` the normalized log of service personnel hours. `lyi` and `lyt` are respectively the logs of the infant output and toddler output.

The ML estimates of parameters for three different nested models are presented in Table 2.[12] As a benchmark we first estimated the half-normal model defined by (10), (11), (12) and (13), under the assumption of no one-sided error, i.e. with the restriction $\sigma_u^2 = 0$. This model, which is denoted by OLS, reduces to the standard linear regression model with a symmetric, normally distributed error term and can be estimated by the OLS method. The estimated coefficients are shown in the third column of Table 2. The second column shows the estimates of the unrestricted half-normal model, where σ_v^2 and σ_u^2 are parametrized as in (16) and (17). This model, which is denoted by HN, allows us to conduct a first check of the correct specification of the Stochastic Frontier model with technical inefficiency. The first column of Table 2 presents the estimates of the half-normal model with heteroscedasticity in both the error terms, which is denoted by HNH. The variance of v_i and the pre-truncated variance of u_i are parametrized by the exponential function as in (18) and (19). The models HN and HNH are nested and specifically HN is the HNH model under the restriction that the coefficients δ and θ are equal to zero. The two models are compared to show the effects of heteroscedasticity on the estimated parameters and efficiency measures. The HNH model is also used to examine the determinants of technical inefficiency.

The OLS model fits the data quite well with an R-squared exceeding 94%. All the first-order coefficients have a correct sign and are significantly different from zero

[12]The estimates have been carried out in Stata by using the software code provided by Kumbhakar *et al.* (2015). The complete tables of results are found in the Appendix.

Table 2: ML estimates – Dependent variable `nledu`

	HNH	HN	OLS
Variable	Coefficient	Coefficient	Coefficient
lcap	0.620**	0.693**	0.615**
lser	0.094**	0.096**	0.107**
lcap2	-0.642**	-0.590**	-0.646**
lser2	0.122*	0.159**	0.241**
lcap_lser	0.052	0.031	- 0.019
lyi	-0.093**	-0.090**	- 0.084**
lyt	-0.726**	-0.712**	-0.700**
lyi2	-0.029**	-0.027**	-0.025**
lyt2	-0.118**	-0.163**	-0.145**
lyi_lyt	0.027**	0.029**	0.025**
lcap_lyi	0.014**	0.009	0.009
lcap_lyt	-0.070	-0.163**	0.053
lser_lyi	0.000	0.007	0.003
lser_lyt	-0.035	-0.063†	-0.014
Intercept	0.033†	0.096**	-0.069**
pub	0.865*		
ptr_ft	3.650**		
exh	-1.341**		
disa	0.317		
mix	0.698**		
food_in	0.174		
frag	1.451*		
Intercept	-6.244**	-3.303**	
pro	-0.522*		
small	1.028**		
Intercept	-5.348**	-5.786**	
Log-likelihood	425.1	331.43	305.73

Significance levels : † : 10% $*$: 5% $**$: 1%

as well as several second-order parameters. However, the OLS residuals exhibit a strongly significant negative skewness which is a symptom of the presence of a one-sided error. In fact, a comparison between the OLS and the HN models reveals that the specification of a stochastic production frontier model with technical inefficiency is supported by empirical evidence. A Likelihood Ratio (LM) test for the null hypothesis of no one-sided error was conducted by comparing the log-likelihood values of the 'restricted' model, OLS, and the 'unrestricted' stochastic frontier HN model.[13] The LR test for the null hypothesis $\sigma_u^2 = 0$ has a mixed χ^2 distribution with 1 degree of freedom and its critical values for hypothesis testing are tabulated in Kodde and Palm (1986). The critical value at the 1% significance level is 5.412. The computed value of LM is 51.396 indicating a strong rejection of the null hypothesis of no one-sided error. Therefore, there is empirical evidence that justifies the use of the stochastic frontier model for the analysis of technical inefficiency in day-care centres.

The estimates of technical inefficiency in the HN model are obtained by using Jondrow *at al.* (1982) formula (20). The average value of inefficiency is 0.15 (with a standard deviation of 0.11), which means that actual inputs can be proportionally reduced by 15% without reducing output. We also obtained the estimates of observation-specific technical efficiency, $E(e^{-u} \mid \epsilon_i)$, by using Battese and Coelli (1995) formula (21). The estimated average technical efficiency is 0.87, with a standard deviation of 0.09, meaning that the day-care service can be provided by employing, on average, only 87% of the actual inputs. The distribution of observation-specific technical efficiency estimates of the HN model is plotted in Figure A.1 in the Appendix.

An unsatisfactory aspect of the estimated HN model is revealed by the comparison of the frontier coefficient estimates between the OLS and HN models, which shows sizeable differences. Since the OLS is known to produce consistent estimates of the slope parameters,[14] the presence of these discrepancies may cast some doubts about the goodness of frontier estimates in the HN model and, as a result, of the ineffi-

[13]The LR statistic is given by $LR = -2(ll_R - ll_U)$, where ll_R and ll_U are, respectively, the log-likelihood values of the restricted model and the unrestricted model.

[14]See, for example, Olson *et al.* (1980)

ciency estimates. Moreover, since neglecting heteroscedasticity may result in biased estimates, as pointed out by Kumbakar and Lovell (2000), we decided to introduce heteroscedasticity of the error terms in the half normal model. Hetheroscedasticity has been modelled by using the exogenous variables listed in Table 1. In the variance equation (19) of the statistical error, two exogenous w_h variables have been considered, `pro` and `small`. In fact we noticed that data measurements are more accurate and systematic in centres located in urban areas and that small size centres exhibit greater variability in the data because they are usually located in rural areas, where the use of capacity is more vulnerable to changes of demographic factors. In the variance equation (18) of the inefficiency term we included as exogenous z_k variables those which are supposed to be the major sources of technical efficiency or inefficiency. For example, the public or private nature of the center and the characteristics of the organization of the service such as the size of part-time when part-time classes are absent, the extra hours of service, the presence of mixed-age classes, the degree of externalization of auxiliary services and the fragmentation of the labour-force. The estimates of the HNH model are shown in column one of Table 2.

We performed a LR test to check whether the model with heteroscedasticity is preferred as compared to the HN model. Since the two models are nested, we tested for the null hypothesis that $\gamma = 0$ and $\delta = 0$. The computed LR test is 187.84 while the critical value at the 1% significance level of the χ^2 distribution with 9 degrees of freedom is 21.67. Therefore, the homoscedastic half-normal model is strongly rejected and we shall focus our analysis on the HNH model.

First of all it can be noticed that the slope coefficients of the production frontier are not noticeably different from those of the OLS model and, in particular, the first-order input coefficients are positive and significant. The estimated coefficient of the teaching staff input is recovered from the homogeneity restriction (6) and is $\alpha_3 = 0.286$. The other parameters in (5) can be similarly recovered from the homogeneity restrictions (6), (7) and (8).

The first-order coefficients of outputs are significant and negative, as expected, and are used to obtain an estimate of returns to scale. A local measure of returns to scale is the elasticity of scale which is given, at the input and output means, by

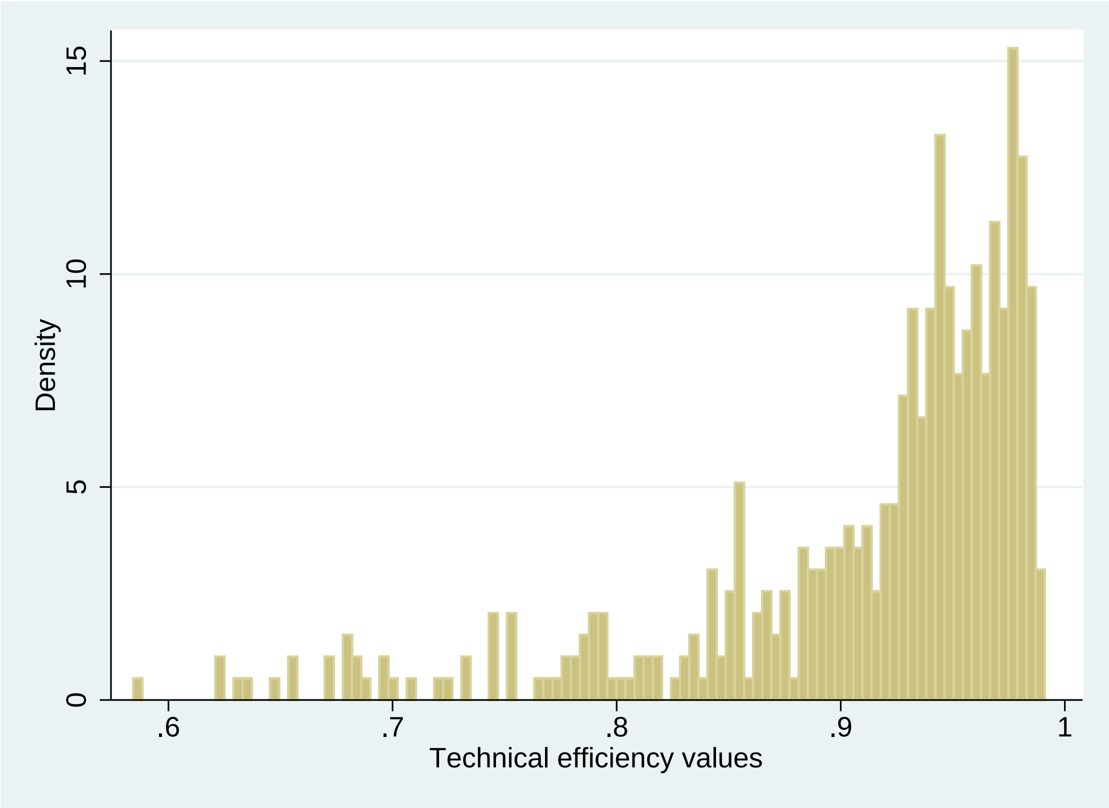

Figure 2: Distribution of technical efficiency in the HNH model

$\eta_s = -1/(\beta_1 + \beta_2)$.[15] The estimated elasticity of scale, obtained by substituting the first-order output coefficients, is 1.21, which means that the day-care production technology exhibits locally increasing returns to scale.[16] A 1% proportional increase of all the inputs may produce a proportional increase of all the outputs by more than 1% and precisely by 1.21%. This conclusion, which conforms with the results of others empirical studies on day-care services,[17] indicates that the average centre size of 47 children is smaller than the minimum efficient scale of production.

The estimated model allows us to perform a test of separability between inputs and outputs. If the marginal rate of transformation between inputs is not affected

[15] See Färe and Primont (1995).

[16] The LM test for the null hypothesis of constant returns against the alternative of increasing returns to scale is rejected at any of the usual levels of significance.

[17] See, for example, Powell and Cosgrove (1992) and Mocan (1997).

by the amount of outputs, then an aggregate measure of outputs can be used in the representation of technology, which can also be described by a single-product technology. Separability obtains if all the cross input-output coefficients γ_{hm} are equal to zero. Although none of these coefficients taken individually is significantly different from zero, the joint hypothesis that all γ parameters are equal to zero can be rejected. Indeed, a LR test for the null hypothesis that $\gamma_{hm} = 0$ for $h = 1, 2$ and $m = 1, 2$ was conducted. The value of the LR test is 9.78 while the critical value at the 5% (1%) level of significance of the χ^2 distribution with 4 degrees of freedom is 9.49 (13.78). Therefore the hypothesis of separability between inputs and outputs is rejected at the 5% level of significance (but not at 1%). Our empirical results seem to suggest that the use of aggregate output and a production function in the analysis of day-care technology is not justified and should be avoided.

The estimates of technical inefficiency and efficiency in the HNH model are computed by using the formulae (20) and (21).[18] The average inefficiency is 0.10, with a standard deviation of 0.09, and the average technical efficiency is 0.91, with a standard deviation of 0.08. As can be noted, there are noticeable differences in the estimated values of efficiency and inefficiency between the HNH and the HN model. The average inefficiency decreases from 14% to 10% while efficiency increases from 87% to 91%. These more favourable measures are to be preferred since are based on more accurate estimates of the production frontier. The distribution of technical efficiency estimates is plotted in Figure 2. As can be seen, the distribution is highly concentrated since almost three-quarters of the centres have technical efficiency above 90%.[19]

Finally, the HNH model has been used to address the issue of the exogenous determinants of inefficiency. In particular, we were interested to see which of the z_h variables that parametrize the variance of u_i can be regarded as the sources of inefficiency and to evaluate the magnitude of their effects. We have considered as explanatory variables the public or private nature of the day-care center and the other variables in Table 1 related to the organization of the service and the labour-force.

[18]The measures are computed by using the observation-specific estimates of $\sigma^2_{v,i}$ and $\sigma^2_{u,i}$, which in turns are obtained from (18) and (19) by substituting for the estimated values of the parameters.

[19]The distribution of estimated technical inefficiency is plotted in Figure A.2 in the Appendix.

Table 3: Marginal effects on technical inefficiency (average values)

Variable:	`pub`	`ptr_ft`	`exh`	`disa`	`mix`	`food_in`	`frag`
	0.041	0.174	-0.064	0.015	0.033	0.008	0.069

The estimates of the δ coefficients in the parametrization equation of $\sigma^2_{u,i}$, (18), are presented in the first column of Table 2. The estimated value of δ_k gives the effect of the z_k variable on the variance of u_i, but it does not directly provide the *marginal effect* of z_k on inefficiency, because the dependence of the unconditional mean $E(u_i)$ on $\sigma^2_{u,i}$ is non linear. Although the sign of the coefficient δ_k reveals the sign of the effect of the variable z_k, the magnitude is given by the value of the marginal effect computed by using (22) and shown in Table 3.

All the estimated δ coefficients have the expected sign and most of them are different from zero at a level of significance below 5%. The coefficient of `pub` is positive and significantly different from zero meaning that public day-care centres are more inefficient. The point estimate of the marginal effect is 0.041, which means that, on average, public centres have a value of technical inefficiency of 4.1 percentage points higher than private centres, all else being equal. Another positive and highly significant effect on inefficiency is given by the presence of part-time children in centres with no classrooms arranged on a part-time basis. The marginal effect of `ptr_ft` is 0.174 meaning that if the ratio of part-time children to total children increases by 10 percentage points the inefficiency increases by about 1.7 percentage points. Since more than three-quarters of centres have only full-time classes and since the average proportion of children served on a part-time basis is about 20%, without significant differences between public and private centres, we conclude that this determinant is a major source of inefficiency and equally affects public and private facilities.

A negative effect on inefficiency and therefore a pro-efficiency effect is provided by extra hours of service, i.e. hours in excess of the standard opening hours. From Table 2 it is seen that `exh` has a negative and highly significant coefficient and that its marginal effect is -0.064. Increasing by 10 minutes the extra hours of service there is the opportunity to obtain a proportional reduction of inputs by about 1.1 percentage points. Centres supplying extra hours of service are about 70% and are equally found between public and private facilities.

The `mix` variable has a positive and significant coefficient. The marginal effect on inefficiency of having mixed age classes is 0.033, which means that this type of organization of the service increases, on average, by 3.3 percentage points the inefficiency of the day-care centre. Indeed, if children of different age groups are put together in the same classroom, regulations require that the child staff ratio of the younger children is met, thus older children are served with more teachers than required. As shown in Table A.1, this factor of inefficiency affects more heavily private rather than public centres.

The `frag` variable has a positive and significant effect on inefficiency, although its marginal effect, which is 0.069, is of limited magnitude. For example, suppose to have 10 full-time teachers and decide to pass to 9 full-time teachers an two half-time teachers, then the `frag` index increases by 0.1 and the average predicted increase of inefficiency is only about 0.7 percentage points. The last two variables, `food_in` and `disa` might have a positive effect on inefficiency, however their coefficients are not significantly different from zero. Therefore, empirical evidence suggests that in-house production of the food service (and the other auxiliary services) is not a determinant of inefficiency. Finally, children with disabilities do not seem to absorb extra resources, may be because their number is relatively small and they are evenly distributed across centres.

5 Summary and conclusions

Our objectives have been to estimate the production technology of the day-care service and to provide measures of (input oriented) technical inefficiency. We also have identified the major determinants of inefficiency and provided estimates of the magnitude of their effects, carrying out a comparative analysis between public and private centres.

The data seem to support the specification of the day-care production technology by means of a stochastic frontier model with one-sided technical inefficiency error. It is also shown that heteroscedasticity of the error terms can not be neglected without introducing serious biases in the estimates of technology and of measures of inefficiency. The technology exhibits locally increasing returns with an estimated elasticity of scale of 1.21. This result indicates that the technically efficient size of the productive units in childcare sector is greater than the average centre size of 47 children. Moreover, the data do not seem to support a technology with separable inputs and outputs, therefore the production of day-care services is best described by a multiple input and output model, while the use of an aggregate output with a production function does not seem to be justified.

Technical inefficiency in the day-care service is about 10% indicating that, on average, centres can proportionally reduce inputs by that amount without reducing the amount of service provided. Also, 10% can be considered as a measure of the potential reduction in the cost of providing the service. On average, public centres are more inefficient than their private counterparts by 4.1 percentage points. In fact, technical inefficiency is about 11% in public centres and about 6 to 7% in private centres. This discrepancy does not seem to be due to differences in the structural quality of service between public and private facilities.

Other important determinants of inefficiency are related to specific modes of organization of the day-care service. Centres with no classes arranged on a part-time basis and serving part-time children are more inefficient; specifically, increasing the proportion of part-time children by 10 percentage points raises inefficiency by 1.7 percentage points. On the other hand, centres offering more extra hours of service are more efficient, because they economize on the hours of standard full-time service; 10 more minutes of extra service are associated with an efficiency gain of 1.1 percentage points. Centres with mixed-age classrooms are, on average, more inefficient by 3.3 percentage points. The fragmentation of the labour force increases inefficiency, although the effect is of limited magnitude, and the choice of outsourcing or producing in-house auxiliary services, such as food, does not seem to have any effect on efficiency. Finally, most of the above determinants do not seem to affect differently public and private facilities, because they are evenly spread over the two types of providers. For example, the practice of supplying part-time service without classes organized on a part-time basis, which is one of the major sources of inefficiency, is equally shared by more than 70% of both public and private centres. A notable exception, however, is the presence of mixed-age classes, which is an important source of inefficiency prevailing in private centres.

A policy implication of our analysis is that there is room for improving efficiency in the day-care sector either by further opening the service to the competition of the private sector or by internal reform of the public centres. Increasing the extent of privatization, however, does not remove some of the major sources of inefficiency which are shared by both public and private providers. As a final remark we may notice that technical measures do not take into account the additional effect of allocative inefficiency originating from non optimal combinations of inputs given prices. As a result, the technical inefficiency measure underestimates overall (or economic) inefficiency and our estimate of 10% represents a lower bound to overall inefficiency. Taking into account that the day-care service is supplied by a regulated sector where three quarters of centres are public and that the main objective of public centres is not necessarily the minimization of costs, a value between 6% and 10% may be considered as a sensible target of the average potential cost reduction which could be achieved by improving the efficiency in the provision of the day-care service.

APPENDIX

Table A.1 Sample composition by attribute and by type of provider, private or public

Variable	**Public**	**Private**	**All**
	N=361	N=121	N=482
Small size centres	23.0%*	44.6%	28.4%
Medium size centres	46.0%	38.8%	44.2%
Large size centres	31.0%	16.5%	27.4%
Centres in capital of province	43.2%	33.1%	40.7%
Centres with only toddlers	29.6%*	61.2%	37.6%
Centres with mixed-age classes	26.3%*	44.6%	30.9%
Centres with children with disabilities	41.8%*	16.5%	35.5%
Centres with only full time classes	77.6%	77.7%	77.6%
Centres with extra hours	72.9%	70.2%	72.2%
All in-house	32.1%*	12.4%	27.2%
In-house food service	60.9%*	20.7%	50.8%
In-house janitorial service	77.3%	82.6%	76.8%
In-house teaching	71.5%*	89.3%	75.9%

(*) indicates that the percentage values between public and private centres are different at the 5% level of significance.

Table A.2 Average characteristics of the service by type of provider, public or private

Variable	**Public**	**Private**	**All**
Children per centre	50.1*	39.2	47.4
Classrooms per centre	3.0*	2.3	2.8
Group size per classroom	17.0	17.1	17.0
Infants (%)	9.7*	5.5	8.6
Children with disabilities (%)	1.1*	.6	1.0
Part-time children (%)	17.7	22.9	19.0
Child to staff ratio (per fte teacher)	6.20*	6.96	6.39
Full-time opening hours (daily)	8.81	8.65	8.77
Extra hours of service (daily)	1.48*	1.84	1.57
Teachers for centre	9.6*	6.4	8.8
Fragmentation (teachers/fte teachers)	1.16*	1.11	1.15
Teaching staff weekly hours ($24 \leq h \leq 36$)	82.1%*	58.7%	76.2%
Teaching staff weekly hours ($h > 36$)	1.7%*	30.1%	8.8%

(*) indicates that the means between public and private centres are different at the 5% level of significance.

Table A.3 OLS – Dependent variable nledu

Variable	Coefficient (Std. Err.)
nlr_cap	0.615**
	(0.040)
nlr_ser	0.107**
	(0.025)
nlr_cap2	-0.646**
	(0.125)
nlr_ser2	0.241**
	(0.070)
nlr_cap_ser	-0.019
	(0.084)
lr_yi	-0.084**
	(0.009)
lr_yt	-0.700**
	(0.020)
lr_yi2	-0.025**
	(0.004)
lr_yt2	-0.145**
	(0.043)
nlr_yi_yt	0.025**
	(0.006)
nlr_cap_yi	0.009
	(0.013)

(continued) OLS

Variable	Coefficient (Std. Err.)
nlr_cap_yt	0.053
	(0.058)
nlr_ser_yi	0.003
	(0.008)
nlr_ser_yt	-0.014
	(0.043)
Intercept	-0.069**
	(0.014)
N	482
R^2	0.944
$F_{(14,467)}$	564.422

Table A.4 HN – Dependent variable nledu

Variable	Coefficient (Std. Err.)
Equation 1 : frontier	
nlr_cap	0.693**
	(0.032)
nlr_ser	0.096**
	(0.022)
nlr_cap2	-0.590**
	(0.114)
nlr_ser2	0.159**
	(0.058)
nlr_cap_ser	0.031
	(0.068)
lr_yi	-0.090**
	(0.009)
lr_yt	-0.712**
	(0.017)
lr_yi2	-0.027**
	(0.004)
lr_yt2	-0.163**
	(0.039)
nlr_yi_yt	0.029**
	(0.005)

(continued) HN

Variable	Coefficient (Std. Err.)
nlr_cap_yi	0.009
	(0.011)
nlr_cap_yt	0.163**
	(0.054)
nlr_ser_yi	0.007
	(0.007)
nlr_ser_yt	-0.063†
	(0.037)
Intercept	0.096**
	(0.014)
Equation 2 : usigmas	
Intercept	-3.303**
	(0.107)
Equation 3 : vsigmas	
Intercept	-5.786**
	(0.233)
N	482
Log-likelihood	331.43
$\chi^2_{(14)}$	9762.895

Table A.5 HNH – Dependent variable nledu

Variable	Coefficient (Std. Err.)
Equation 1 : frontier	
nlr_cap	0.620**
	(0.033)
nlr_ser	0.094**
	(0.020)
nlr_cap2	-0.642**
	(0.108)
nlr_ser2	0.122*
	(0.054)
nlr_cap_ser	0.052
	(0.069)
lr_yi	-0.093**
	(0.006)
lr_yt	-0.726**
	(0.018)
lr_yi2	-0.029**
	(0.003)
lr_yt2	-0.118**
	(0.042)
nlr_yi_yt	0.027**
	(0.005)
nlr_cap_yi	0.014
	(0.009)
nlr_cap_yt	0.070
	(0.046)
nlr_ser_yi	0.000
	(0.006)
nlr_ser_yt	-0.035
	(0.036)

(continued) HNH

Variable	Coefficient (Std. Err.)
Intercept	0.033†
	(0.018)
Equation 2 : usigmas	
pub	0.865*
	(0.347)
ptr_ft	3.650**
	(0.785)
v20	-1.341**
	(0.253)
disa	0.317
	(0.217)
mix	0.698**
	(0.243)
v_mensa	0.174
	(0.252)
frag	1.451*
	(0.628)
Intercept	-6.244**
	(0.859)
Equation 3 : vsigmas	
size1	1.028**
	(0.235)
pro	-0.522*
	(0.228)
Intercept	-5.348**
	(0.226)
N	482
Log-likelihood	425.1
$\chi^2_{(14)}$	6974.916

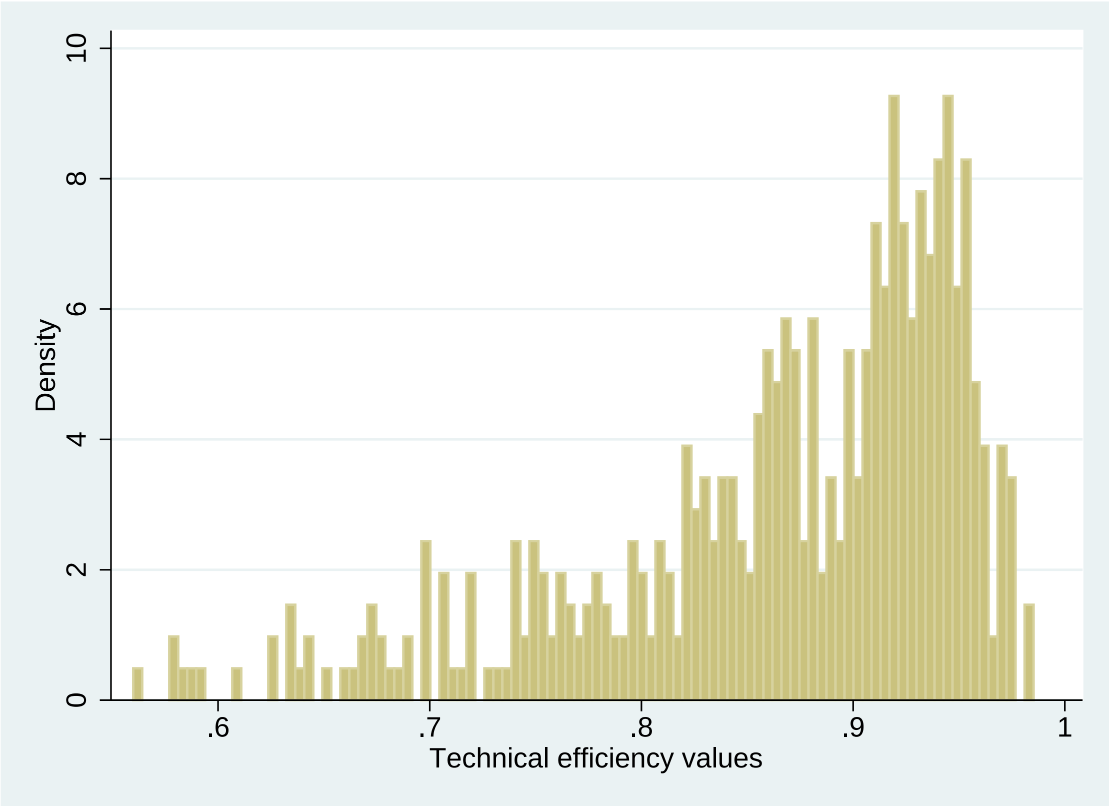

Figure A.1 Distribution of technical efficiency in the HN model

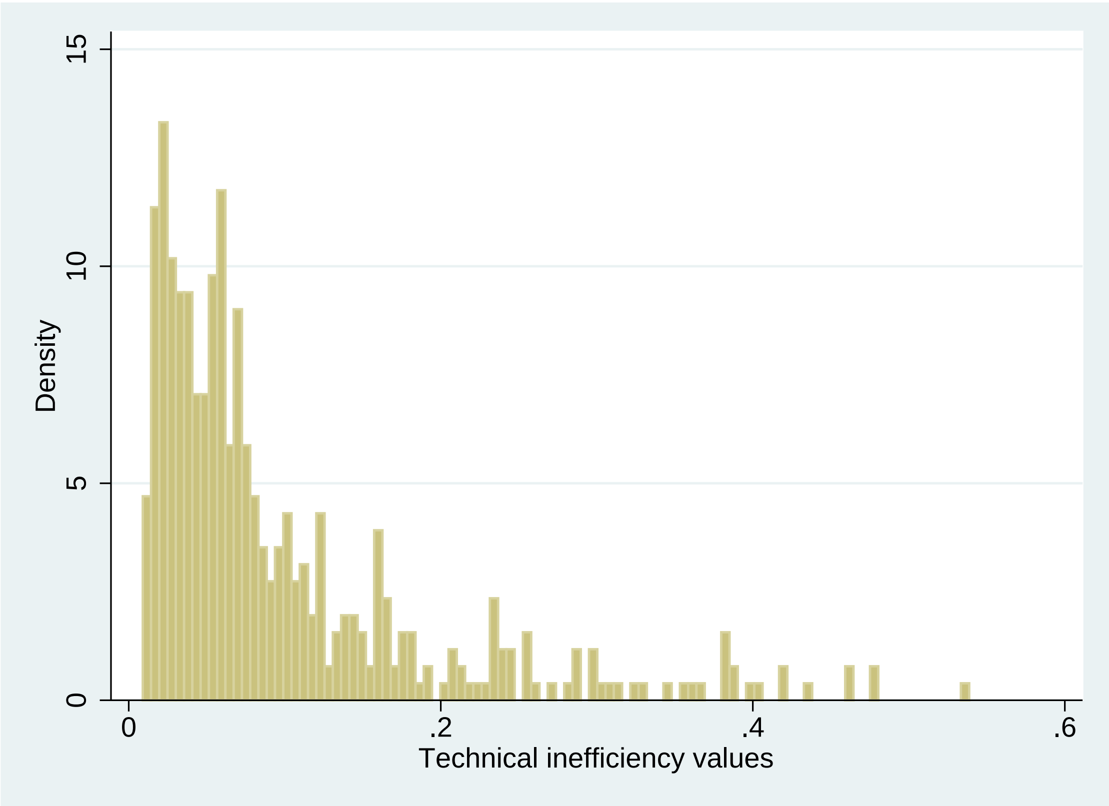

Figure A.2 Distribution of technical inefficiency in the HNH model

Acknowledgements

We wish to acknowledge the contribution of Massimo Baldini who has been very helpful in data analysis and offered valuable advice and suggestions. We also wish to thank Paolo Bosi, Mario Forni and Barbara Pistoresi for reading an earlier draft and providing helpful comments. Of course, none of them is to be held responsible for any remaining errors.

References

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

Battese, G.E. and T.J. Coelli (1995), A Model for Technical Inefficiencies effects in a Stochastic Frontier Production Function for Panel Data, *Empirical Economics*, 20, 325-32.

Bjurek, H. , K. Urban and B. Gustafsson (1992), ... Efficiency, Productivity and Determinants of Inefficiency at Public Day Care Centres in Sweden, *Scandinavian Journal of Economics*, 94, 173-87.

Caudill, S.B. and J.M. Ford (1993) Biases in Frontier Estimation due to Heteroscedasticity, *Economics Letters* , 41, 17-20

Caudill, S.B., J.M. Ford and D.M. Gropper (1995) Frontier Estimation and Firm-Specific Inefficiency Measures in the Presence of Heteroscedasticity, *Journal of Business and Economic Statistics* , 13, 105-111.

Coelli, T.J., and S. Perelman (2000), Technical Efficiency of European Railways: A Distance Function Approach, *Applied Economics*, 32, 1967-76.

Coelli, T.J., D.S. Prasada Rao, C.J. O'Donnell and G. E. Battese (2006) *An Introduction to Efficiency and Productivity Analysis*, Springer, N.Y.

Färe, R. and D. Primont (1995), *Multi-output Production and Duality: Theory and Applications*, Kluwer Academic Publishers

Farrel, M.J. (1957) The Measurement of Productive Efficiency, *Journal of the Royal Statistical Society* Series A (General), 120, 253-90

Hadri, K. (1999) Estimation of a Doubly Heteroscedastic Frontier Cost Function, *Journal of Business and Economic Statistics* , 17, 359-363.

Jondrow, J., C.A.K. Lovell, I.S. Materow and P. Schmidt (1982), On the Estimation of technical efficiency in the Stochastic Frontier production Function Model, *Journal of Econometrics*, 19, 233-8.

Kodde, D.A. and F.C. Palm (1986, Wald Criteria for Jointly Testing Equality and Inequality Restrictions, *Econometrica*, 54. 1243-8.

Kumbhakar, S.C. and C.A.K. Lovell (2000), *Stochastic Frontier Analysis*, Cambridge, Cambridge University Press.

Kumbhakar, S.C., H-J. Wang and A. P Horncastle (2015), *A Practitioner's Guide to Stochastic Frontier Analysis Using Stata*, Cambridge University Press, N.Y.

Mocan, Nacy H. (1997), Cost Functions, Efficiency, and Quality in Day Care Centres, *Journal of Human Resources*, 32, 861-891.

Mukerjee, S. and A.D. Witte (1993), Provision of Child Care: Costs Functions for Profit-Making and Non-For-Profit Day Care Centres, *Journal of Productivity Analysis*, 41, 145-63.

Olson, J., B. Schmidt and D. Waldman (1980), A Monte Carlo Study of Estimators of Stochastic Frontier Production Functions, *Journal of Econometrics*, 13, 67-82

Preston, A. (1993), Efficiency, Quality, and Social Externalities in the Provision of Day Care: Comparisons of Nonprofit and For-Profit Firms *Journal of Productivity Analysis*, 41, 165-82.

Powell, I. and J. Cosgrove (1992), Quality and Cost in Early Childhood Education, *Journal of Human Resources*, 27, 472-84.

Shephard, R.W. (1970), *The Theory of Cost and Production Functions*, Priceton, Priceton University Press

Stevenson, R.E. (1980), Likelihood Function for Generalized Stochastic Frontier Estimation, *Journal of Econometrics*, 13, 57-66