Economies of scale and optimal size of hospitals: Empirical results for Danish public hospitals

Kristensen, Troels

Publication date: 2008

Document version
Publisher's PDF, also known as Version of record

Citation for published version (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain.
• You may freely distribute the URL identifying the publication in the public portal.

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 16. dec., 2018
Preface:
The aim of this Working Paper (WP) is to contribute to the current debate about the future configuration of the Danish hospital sector, in particular the issue of 'optimal' hospital size. From an economic perspective we find that the empirical evidence underpinning the planned hospital structure is ambiguous and partly lacking. The WP contributes to this debate with an empirical study of the question of economies of scale in the Danish hospital sector and estimates of an optimal hospital size.

Our WP is addressed to foreign and Danish economists with an interest in industrial organization and, in particular, the organisation of the (Danish) hospital sector. Furthermore, our WP serves as the back ground paper for an article submitted to a peer-reviewed health economic journal. In addition to the content of the submitted article, the WP contains details such as an appendix and a more detailed discussion of methodological issues and concepts.

Acknowledgments
An earlier version of this WP was presented at the Health Economics’ Study Group Joint UK & Nordic meeting at the Health Economic Research Unit, University of Aberdeen 27-29 August 2008. We thank the discussant, Professor Jørgen T. Lauridsen, for helpful comments and suggestions. We would also like to acknowledge helpful comments from our SDU colleague, associate professor Jacob Nielsen Arendt, and visiting fellow at the Danish Institute for Health Services Research Professor Andrew Street, the University of York. Finally, we would like to thank John Mason of Manuscripts.dk for linguistic revision and secretary Helk Møller Jensen for helping in formatting and compiling the WP.

This project was partly financed by the Danish VAT foundation (Det kommunale moms-fond). All errors, omissions and opinions are the sole responsibility of the authors.

Troels Kristensen, Kim Rose Olsen and Kjeld Møller Pedersen
November 2008
Title: Economies of scale and optimal size of hospitals: Empirical results for Danish public hospitals

Abstract:

Context and aim: The Danish hospital sector is facing a significant rebuilding programme, driven by a political desire to concentrate activity in fewer and larger hospitals. Our aim is to analyse whether the current configuration of Danish hospitals is subject to scale economies that may justify such plans and to estimate an optimal hospital size.

Methods: We estimate cost functions using panel data on total costs, DRG-weighted casemix, and number of beds for three years from 2004-2006. A short-run cost function is used to derive estimates of long-run scale economies by applying the envelope condition.

Results: We identify moderate to significant long-run economies of scale when applying two alternative translog cost functions. However, using a quadratic functional form we identify constant economies of scale for the medium-sized sub-groups and decreasing economies of scale for the largest sub-groups. The optimal number of beds per hospital is estimated to be 275 beds per site. Sensitivity analysis to partial changes in model parameters yields a joint 95% confidence interval in the range 130 – 585 beds per site.

Conclusions: The results indicate that it may be appropriate to consolidate the production of small hospitals (<230 beds) on fewer and larger units. Keywords: Economies of scale, optimal size, hospitals, cost function.


1. Introduction

In Denmark the number of somatic hospitals has decreased from 117 in 1980 to 52 in 2004. Part of this decrease is due to the fact that the concept or idea of a hospital has changed. Until the early 1990’s there was always a one-to-one relationship between a hospital as a management entity and its geographical location. During the past 15 years, however, hospitals at different locations have been merged so that many ‘hospitals’ today consist of several geographically distinct ‘production units’ being managed together. These new entities, consisting of several production units, are called management entities or conglomerate hospitals. This trend towards centralization is not uniquely Danish but is also found in, for instance, England, Norway and Sweden [1,2,3].

Hospital plans from the five Danish regions show that this development is expected to continue in the years to come [4]. They are planning a significant rebuilding program including green field investments at 5 new sites, significant extension and reconstruction of several existing hospitals and mergers or closures of several small hospitals¹.

Whether the increasing centralization of hospitals is to be seen as an advantage depends on a) whether there are economies of scale i.e. lower average costs and b) whether bigger hospitals lead to improved clinical outcomes [5]. Exploiting economies of scale may help to limit costs of health care outputs without compromising their quality and volume. On the other hand, hospitals may become so large that the cost of treatment will be higher due to diseconomies of scale. Furthermore, plans to concentrate further assume that the ‘optimal’ hospital size is bigger than in the current configuration.

From an economic perspective the evidence base underpinning centralization is weak, that is to say that there is a conspicuous absence of research and discussion of economies of scale in Danish hospital production. No econometric studies of economies of scale in hospitals have ever been undertaken in Denmark. In Europe, unlike in the U.S. where more literature exists on economies of scale, the economics of this trend towards larger hospitals have not been sufficiently analyzed. Notable exceptions are [6] and [7] along with a survey [5]. Evidence of the ‘optimal’ hospital size is important at a time when the hospital sector is facing major restructuring. Therefore, the aim of this study is to assess whether there are unexploited economies of scale in the current configuration and to estimate an ‘optimal’ hospital size. The present study

¹ The association of Danish Regions has published the rebuilding program for each region, see http://www.godtsygehusbyggeri.dk.
is limited to assessing economies of scale and 'optimal' hospital size for Danish hospitals in the period 2004-2006 from a hybrid econometric cost function perspective [8].

The unit of analysis is the hospital “production unit”, not the hospital management entity. This approach has become increasingly relevant due to the trend towards concentration in secondary healthcare. In relation to the rebuilding programme it is the geographical hospital “production unit” which is the relevant decision unit when deciding to build new hospitals to replace one or more former hospitals. It is not the management entities with satellite production units which may be located far from each other that are the relevant analytical unit. A distance of 30-50 km between units within the same management entity is quite common. Besides, using the production unit as the unit of analysis means that we can interpret the estimated economies of scale and hospital sizes in relation to the actual geographical hospital production units instead of multisited hospital management entities. In the following, consequently, the term hospital is reserved for freestanding “production units” in specific geographical sites rather than “hospital management entities” which consist of several production units at different sites.

Our presentation of earlier studies is restricted to those that use econometric cost functions that resemble the methods used in this study. However, this study differs from the majority of earlier studies in several ways – especially in its estimation of long run cost functions and ‘optimal’ hospital size. So far, this approach has not been used in European studies. The literature search revealed only a single Canadian study that has estimated an 'optimal' hospital size using the envelope condition [9]. All other earlier studies of 'optimal' size are based on scale estimates – excluding specific estimates of 'optimal' hospital size.

2. Earlier results

The empirical literature on economies of scale in hospitals is extensive, if all statistical techniques are included [5]. Despite the fact that the literature reflects different methods and covers many different countries the results are remarkably consistent, according to a recent survey of 103 studies by Aletras [10], i.e. these studies reveal constant economies – or even diseconomies – of scale for the average hospital with about 200-300 beds, see also Aletras et al. [11]. However, studies based on structural or hybrid econometric cost functions only represent about one fifth of these studies. According to [11] economies of scale were evident only for small hospitals with less than 200 beds and the ‘optimal’ size for acute hospitals ranged form 200 to 400 beds (based on the interpretation of scale estimates). For hospitals above 400-600 beds it was concluded that the average cost increases.

Studies after 1997 based on structural or hybrid econometric cost function do not confirm the above-mentioned consistency. In North America the application of panel data has shown economies of scale in Canada [13,9]. A third study based on cross-section data also indicated economies of scale [14]. Moreover, a study of acute care hospitals in California has revealed a minor trend towards economies of scale [15].

In contrast to, for example, [13] and [14], the present writers use casemix-adjusted output measures instead of particularly constructed casemix indexes to adjust for differences in patient mix and severity. Furthermore, this study differs from [9], for example, by including costs shifters to adjust the structural model for cost drivers that are specific to hospitals.

Finally, it is apparent that the studies described do not rely on the latest data. This study applies the latest data and data adapted for managerial decision-making and efficiency-measurement in the Danish hospital sector.

---

This study is inspired by the relatively well known properties of scale and optimal scale size of parametric cost functions. The corresponding properties of the alternative non-parametric deterministic data envelopment analysis (DEA) are less explored [12].
3. Methods

Using econometric assessment of economies of scale and optimal size of hospitals, a number of choices need to be made such as unit of analysis, model of hospital production, model for cost functions, specification of cost and output variables, and estimation technique.

As elsewhere, the number of hospitals in Denmark has declined radically over the past two decades as a result of mergers and closures. This means that many hospitals have merged into management conglomerates of hospitals spread across several geographical sites, often with a degree of division of labour and hence specialization. From a theoretical point of view, therefore, it is appropriate to estimate short-run cost functions. The argument is that this approach allows hospitals to use positive non-optimal levels of the fixed inputs in the short run. Hence, hospitals are only assumed to use cost-minimizing quantities of easily adjustable variable inputs, such as nurses, physicians and materials. Furthermore, cost function estimation by frontier estimators may account for deviations from the cost frontier (non-minimum cost functions).

The advantages of 'flexible' cost functions are that they do not prejudge the existence or degree of economies of scale. Unfortunately, this increased flexibility is obtained at the cost of having more parameters to be estimated than in more restricted functional forms such as the classical Cobb-Douglas cost function. Coble and Kauflinger (1994) showed that this approach allows hospitals to use positive non-optimal levels of the fixed inputs in the short run. Hence, hospitals are only assumed to use cost-minimizing quantities of easily adjustable variable inputs, such as nurses, physicians and materials. Furthermore, cost function estimation by frontier estimators may account for deviations from the cost frontier (non-minimum cost functions).

Cost and output variables

The common cost function

\[ g(C_t) = \sum_{i=1}^{n} \alpha_i g(C_{i,t}) + \sum_{j=1}^{m} \beta_j g(C_{j,t}) + \sum_{h=1}^{l} \gamma_h g(C_{h,t}) + \sum_{j=1}^{m} \delta_j g(C_{j,t}) + \sum_{h=1}^{l} \epsilon_h g(C_{h,t}) + \eta + \zeta, \]

where \( g \) is a real valued function of the total cost for somatic treatment \( C \), \( \alpha_i \), \( \beta_j \), \( \gamma_h \), and \( \delta_j \) are parameters to be estimated, \( \eta \) and \( \zeta \) are random disturbance terms. Firstly, (1) is equal to the quadratic form when \( \alpha_i \) and \( \zeta \) are chosen to be equal to \( \alpha \). Secondly, if \( \alpha_i \) and \( \zeta \) are chosen to be the natural logarithmic function, (1) yields the translog model. Thirdly, the Cobb-Douglas version in (1) is excluded. The exact model specifications are made as a minimal basis. This category of functional forms is now preferred to the more naive structural functional form (1), for instance when applied to examine the sensitivity of results to the chosen functional form.

The study is based on data for hospital production sites to reveal knowledge relevant for Danish hospital sector. If the management conglomerate sites had been used as the sole unit of analysis, then estimates could have been conducted only in relation to an 'optimal management unit'. This was also the case in this study. In other words, the arguments for the judgment of results to the chosen functional form are not significant compared to the judgment of results to the chosen functional form. The advantages of 'flexible' cost functions are that they do not prejudge the existence or degree of economies of scale. Unfortunately, this increased flexibility is obtained at the cost of having more parameters to be estimated than in more restricted functional forms such as the classical Cobb-Douglas cost function. This was also the case in this study. In other words, the arguments for the judgment of results to the chosen functional form are not significant compared to the judgment of results to the chosen functional form.
words.  The estimated coefficients became insignificant and unstable, and the signs changed in such a way that the results did not make sense from an economic point of view. The reason is likely that FE models suffer from omitted variables bias as is often the case in cross-sectional studies.

The long-run cost function has been estimated in two different ways. The first approach uses the estimated short-run cost function and the envelope condition to calculate the long-run cost function. This means that the first order condition of the short-run cost function set equal to zero defines the optimal relationship between inputs and outputs as defined in (2). Substituting (2) into the short-run cost function (1) yields the long-run cost function. Since (1) is a second order Taylor approximation, this calculation yields a second order equation for the long-run marginal cost function and the long-run quadratic cost function respectively.

Calculation of the optimal hospital size

The optimal size of a hospital can be calculated from the short-run cost function by application of the envelope condition [9, 27]. Given the cost function (1), this calculation can be expressed as:

\[ K^* = \frac{\partial C}{\partial K} = 0 \]

The Danish DRG system adjusts each output (discharge) for casemix and to a certain extent for severity through the DRG cost weights attached to each discharge. Therefore, this study uses DRG values to measure aggregate hospital output, as related Danish studies, e.g. [21, 22]. Outputs are defined as the total DRG value per year per hospital. In this study, the number of patients is used as a proxy for the fixed cost (ownership) and hospital size. This approach makes it feasible to estimate the optimal number of beds (optimal capital stock) by application of a flexible functional form and the envelope condition [9].

Sensitivity analysis was conducted for the optimal hospital size by substituting the estimated short-run cost function and the envelope condition into the long-run cost function.

5. While it is possible to conduct the above-mentioned calculations for the flexible functional form, this is not feasible for the Cobb-Douglas model. This is due to the squared term lacking in the Cobb-Douglas functional form, which prohibits the calculation of an optimal hospital size (K*) from (2) since K is not twice differentiable with respect to K.

 Outputs are defined as the total DRG value per year per hospital. In this study, the number of patients is used as a proxy for the fixed cost (ownership) and hospital size. This approach makes it feasible to estimate the optimal number of beds (optimal capital stock) by application of a flexible functional form and the envelope condition [9].

Sensitivity analysis was conducted for the optimal hospital size by substituting the estimated short-run cost function and the envelope condition into the long-run cost function.

While it is possible to conduct the above-mentioned calculations for the flexible functional form, this is not feasible for the Cobb-Douglas model. This is due to the squared term lacking in the Cobb-Douglas functional form, which prohibits the calculation of an optimal hospital size (K*) from (2) since K is not twice differentiable with respect to K.
The second approach applies an alternative way, where the cost function is estimated directly without use of the envelope condition \([6]\). This approach, the ‘direct approach’, has been achieved by omitting the number of beds in the estimation of the cost function. Hence, in contrast to the ‘envelope condition approach’, the direct approach assumes that hospitals use an optimal amount of capital in terms of beds \((K)\) in the short and long run, see for instance, \([16]\). In other words, the direct approach assumes fixed cost to become variable in the long run, in other words as a function of, for example, outputs. Besides, it assumes that fixed cost varies with output across the data set and that the hospitals are always endowed with an optimal amount of capital. The latter is a relatively restrictive assumption to be discussed later. The degrees of freedom gained from dropping the beds variable \((K)\) are used to include two output measures, DRG value of inpatients and outpatients instead of the total DRG value per hospital.

Derivation of economies of scale estimates

In accordance with \([28]\), economies of scale are estimated in a way that shows the relative rise in costs when output is increased proportionally. Since the translog and Cobb-Douglas models are logged in all variables and the quadratic forms are unlogged this yields (3) and (4):

\[
SE_1 = \sum \frac{\partial g(Q)}{\partial (Q_m^m)} \quad (3)
\]

\[
SE_2 = \sum \left( \frac{2C}{\partial Q_m} \right) \left( \frac{C}{Q_m} \right) \quad (4)
\]

\(SE_1\) in (3) expresses the sum of first order partial derivatives of the cost function \((1)\) with respect to each output \(Q_m\) in logs. The logarithmic transformations imply that each of these derivatives is an estimate of cost elasticities for each \(Q_m\).

\(SE_2\) in (4) measures the sum of cost elasticities with respect to output. Each of the cost elasticities in \(SE_2\) is calculated using the standard (unlogged) approach, because the quadratic form is in cost levels. In the translog and quadratic models, in which scale estimates by definition are flexible, the sub-group median hospital was used to calculate scale estimates for each of the defined size groups. The size groups were defined by quartiles. The smallest size group (1st quartile) consists of the 25% of hospitals, which has the smallest number of beds, while the other size groups, 2nd quartile, 3rd quartile and 4th quartile, include hospitals with a size in the respective quartiles. Both \(SE_1\) and \(SE_2\) express the multi-product analog of marginal cost divided by average cost. The exact model specifications are shown in the Appendix.

In equations (3) and (4), \(SE\) values less than 1 indicate economies of scale corresponding to cost increases, which are smaller than the proportional output increase. \(SE\) values larger than 1 show diseconomies of scale.

Data

The data comes from a national cost database developed by the National Board of Health \([29]\). The cost database is based on patient activity and cost information from most public hospitals and is also used to calculate Danish DRG tariffs. Total hospital costs are actual costs incurred in respective years adjusted for costs from shared facilities with other hospitals, such as laundry. They are used as the best available proxy for the total cost for somatic treatment. DRG values, or in other words the reimbursement received by hospitals, give the most appropriate picture of the value of hospital production.

There may be some inconsistencies for the DRG values for the three years 2004 to 2006, because the DRG grouper used for 2005 and 2006 was different from that used for 2004 (giving different input prices). This means that 2004 data is based on 2007 input prices while 2005 and 2006 data is based on real 2008 input prices. We assume, however, that the effect of this is negligible due to a low inflationary level. Variables used and descriptive statistics are shown in table 1.

The National Board of Health calculates adjusted actual operating costs by deducting from total reported operating costs, whenever relevant. This applies, for instance, to the cost of psychiatric services, laboratory services for general practitioners, the cost of medicines provided for outpatients, adjustments for differences in accounting practice and unpaid services between hospitals. This gives the figure for ‘the adjusted operational costs’ which is used in the present study.
Table 1 Descriptive statistics for Danish hospitals in the years 2004-2006

<table>
<thead>
<tr>
<th>Variable</th>
<th>Year</th>
<th>Description</th>
<th>Average</th>
<th>Std. dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dependent:</td>
<td></td>
<td>in 1000 DKK (Danish currency)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Adjusted operational costs</td>
<td>530,934</td>
<td>648,465</td>
<td>21,581</td>
<td>3,591,319</td>
</tr>
<tr>
<td></td>
<td>2004</td>
<td></td>
<td>616,490</td>
<td>784,055</td>
<td>19,035</td>
<td>4,337,614</td>
</tr>
<tr>
<td></td>
<td>2005</td>
<td></td>
<td>439,160</td>
<td>449,149</td>
<td>16,761</td>
<td>1,890,084</td>
</tr>
<tr>
<td></td>
<td>2006</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Independent:</td>
<td></td>
<td>DRG value inpatient</td>
<td>328,514</td>
<td>379,452</td>
<td>0</td>
<td>2,123,694</td>
</tr>
<tr>
<td></td>
<td>2004</td>
<td></td>
<td>361,908</td>
<td>477,284</td>
<td>0</td>
<td>2,516,725</td>
</tr>
<tr>
<td></td>
<td>2005</td>
<td></td>
<td>252,824</td>
<td>231,748</td>
<td>3,265</td>
<td>898,218</td>
</tr>
<tr>
<td></td>
<td>2006</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>DRG value outpatient(b)</td>
<td>209,637</td>
<td>270,439</td>
<td>4,909</td>
<td>1,180,949</td>
</tr>
<tr>
<td></td>
<td>2004</td>
<td></td>
<td>274,984</td>
<td>335,966</td>
<td>2,514</td>
<td>1,572,89</td>
</tr>
<tr>
<td></td>
<td>2005</td>
<td></td>
<td>226,344</td>
<td>247,555</td>
<td>3,555</td>
<td>1,000,748</td>
</tr>
<tr>
<td></td>
<td>2006</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total DRG value (Q(1) + Q(2))</td>
<td>538,152</td>
<td>592,734</td>
<td>22,968</td>
<td>3,304,644</td>
</tr>
<tr>
<td></td>
<td>2004</td>
<td></td>
<td>636,892</td>
<td>788,952</td>
<td>12,098</td>
<td>4,089,314</td>
</tr>
<tr>
<td></td>
<td>2005</td>
<td></td>
<td>459,168</td>
<td>668,859</td>
<td>15,163</td>
<td>1,898,966</td>
</tr>
<tr>
<td></td>
<td>2006</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Independent cost shifters: in no. of beds and percentage of hospitals</td>
<td></td>
<td>Average number of staffed beds</td>
<td>281.6</td>
<td>250.9</td>
<td>25.6</td>
<td>1107.1</td>
</tr>
<tr>
<td></td>
<td>2004</td>
<td></td>
<td>265.1</td>
<td>259.5</td>
<td>9</td>
<td>1136.7</td>
</tr>
<tr>
<td></td>
<td>2005</td>
<td></td>
<td>176.0</td>
<td>152.1</td>
<td>9</td>
<td>517</td>
</tr>
</tbody>
</table>

\(a\) Unbalanced due to missing data for 2006. The numbers of observations in 2004-06 are 57, 55 & 31 respectively. \(b\) Including the value of grey zone DRG activity.

Data in table 1 shows that hospital production units on average had operating costs in the range DKK 530 to 616 million. The DRG values are measured in local currency, DKK. The total value of DRG production for each hospital is divided into two output categories: 1) the production value of inpatients and 2) the production value of outpatients, including both so called grey zone patients and emergency patients.

Grey zone patients are patients that the hospital staff both can choose to treat as outpatient or as inpatient (in connection with hospitalization). To avoid distortion of this substitution choice, a special grey zone DRG rate is used. The grey zone DRG rate is calculated as the weighted average between what it costs to perform same-day surgery or outpatient treatment, and the corresponding price for similar inpatient treatment.

The average number of beds per hospital production unit is in the range 265 to 281, but this average covers wide variation between production units (e.g. min. 9, max. 1136 in 2005). The average number of disposable beds per hospital is used as a proxy for the size of hospitals and fixed inputs.

Table 1 also shows that the percentage of public hospitals that were university hospitals was on average approximately 21% in the period 2004 to 2005. Finally, it should be noted that the data for 2006 is generally sparser than the data for the previous year. This is due to data being missing for some of the large units in 2006, i.e. unbalanced panel data. Psychiatric hospitals are excluded from this study. Danish psychiatric hospitals do not use the DRG system. In special hospitals, e.g. Friklinikken in Brædstrup and Hammel Neurocenter, the production process is considered to be atypical. Therefore, six hospitals were excluded.
Results

Table 2 shows the results for the short-run cost models in the Cobb-Douglas, the translog and the quadratic model specification.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Cobb-Douglas</th>
<th>Translog</th>
<th>Quadratic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-0.1423**</td>
<td>-0.1966**</td>
<td>-0.0297</td>
</tr>
<tr>
<td>Inpatients (DRG value)</td>
<td>0.4425***</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Outpatients (DRG value)</td>
<td>0.2736***</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Total DRG value</td>
<td>-0.6921***</td>
<td>1.4800***</td>
<td></td>
</tr>
<tr>
<td>Avg. number of beds</td>
<td>0.0511</td>
<td>-0.0403</td>
<td>-1.2360**</td>
</tr>
<tr>
<td>(Total DRG value)</td>
<td>-</td>
<td>-0.0262</td>
<td>0.1266**</td>
</tr>
<tr>
<td>(Avg. number of beds)</td>
<td>-</td>
<td>-0.0967</td>
<td>0.8131***</td>
</tr>
<tr>
<td>Total DRG value*Avg.</td>
<td>-0.0938</td>
<td>-0.5377***</td>
<td></td>
</tr>
<tr>
<td>number of beds</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of observations</td>
<td>143</td>
<td>143</td>
<td>143</td>
</tr>
<tr>
<td>Number of hospitals</td>
<td>54</td>
<td>60</td>
<td>60</td>
</tr>
<tr>
<td>R²</td>
<td>0.6028</td>
<td>0.7177</td>
<td>0.7657</td>
</tr>
<tr>
<td>Between</td>
<td>0.9387</td>
<td>0.9778</td>
<td>0.9517</td>
</tr>
<tr>
<td>Overall</td>
<td>0.9763</td>
<td>0.9663</td>
<td>0.9574</td>
</tr>
<tr>
<td>F-test (5,78)</td>
<td>28.11***</td>
<td>105.35***</td>
<td>30.10***</td>
</tr>
</tbody>
</table>

Table 2 Regression results – short-run cost functions

For the Cobb-Douglas and translog specification the beta estimates should be interpreted as elasticities, while in the quadratic form they indicate the absolute increase in costs due to an increase in one unit of output.

The Cobb-Douglas and the translog model show that elasticities for inpatients are higher than for outpatients and that the results are quite similar for cross-section and panel data specification except for the outpatient elasticity being higher in 2006 than in 2004 and 2005. This deviation is probably due to missing data in 2006 as mentioned in the data description.

The beta estimates for the average number of beds changes sign and significance across the model specifications leaving the effect ambiguous. The university hospital dummy in table 1 was eliminated in the fixed effect model 2.

The regression results in table 2 are used to estimate the long-run cost function based on the envelope condition, shown together with the direct approach to long-run cost function in table 3. The results in table 2 are also used to estimate the scale elasticities shown in table 4.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Translog &amp; Envelope condition</th>
<th>Translog, FE (without beds)</th>
<th>Quadratic &amp; Envelope condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-0.1950</td>
<td>-0.1709***</td>
<td>1.4980</td>
</tr>
<tr>
<td>Total DRG value</td>
<td>-0.6845</td>
<td>-</td>
<td>2.8092</td>
</tr>
<tr>
<td>Inpatients DRG value</td>
<td>-0.0045</td>
<td>-</td>
<td>0.3524</td>
</tr>
<tr>
<td>Outpatients DRG value</td>
<td>-0.5388***</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Inpatients DRG value</td>
<td>-0.4100***</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Outpatients DRG value</td>
<td>-0.0879**</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Inpatients*Outpatients</td>
<td>-0.1483***</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>within</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between</td>
<td>-0.6028</td>
<td>0.6880</td>
<td>-</td>
</tr>
<tr>
<td>Overall</td>
<td>0.9387</td>
<td>0.9834</td>
<td>-</td>
</tr>
<tr>
<td>F-test (5,78)</td>
<td>28.11***</td>
<td>105.35***</td>
<td>30.10***</td>
</tr>
</tbody>
</table>

Table 3 shows the three long-run cost functions. The first version of the translog function and the quadratic function are calculated from the short-run cost functions in table 2 by substitution of equation (2) into equation (1).

The second version of the translog model is a directly estimated, fixed-effect, long-run cost function. In this model, the total output vector has been divided into two output measures – inpatient and outpatient DRG value – and no cost shifters have been included to avoid collinearity.

The beta estimates of the Cobb-Douglas and translog cost functions are elasticities, whereas the betas of the quadratic model show the absolute increases in costs. The difference in

1In an earlier cross-section analysis the university hospital dummy was positively significant for each of the years 2004-2006. This indicates, as expected, that university hospitals incur higher cost, see the method section.
signs between the translog models and quadratic for m is due to the logarithmic transformations and the functional form used, while the nested Cobb-Douglas and the translog models depend on the results. However, an important difference is that the translog and quadratic models allow us to make a non-constant scale estimate. The results indicate that scale estimates are increasing with the size of the hospitals, which should be interpreted as an indication of declining economies of scale as the size of the hospital becomes larger.

Table 4 shows estimates of economies of scale for the alternative functional forms when applying the short-run and long-run cost functions respectively. Both short-run and long-run economies of scale are measured by conventional ray scale economies, which are the elasticity of cost taken along a ray that holds product mix constant. SE < 1 implies scale economies and SE > 1 implies diseconomies when outputs are changed proportionately.

Table 4: Short-run and long-run scale estimates for hospital production units in Denmark.

<table>
<thead>
<tr>
<th>Groups of hospitals</th>
<th>Cobb-Douglas (envelope condition)</th>
<th>Translog (envelope condition)</th>
<th>Translog (directly estimated)</th>
<th>Quadratic (envelope condition)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All hospitals</td>
<td>0.7160</td>
<td>0.7086</td>
<td>1.0826</td>
<td>0.6493</td>
</tr>
<tr>
<td>1st quartile</td>
<td>0.7160</td>
<td>0.6235</td>
<td>1.3517</td>
<td>0.6897</td>
</tr>
<tr>
<td>2nd quartile</td>
<td>0.7160</td>
<td>0.6688</td>
<td>1.3277</td>
<td>0.6796</td>
</tr>
<tr>
<td>3rd quartile</td>
<td>0.7160</td>
<td>0.6972</td>
<td>1.2241</td>
<td>0.6716</td>
</tr>
<tr>
<td>4th quartile</td>
<td>0.7160</td>
<td>0.7206</td>
<td>0.9657</td>
<td>0.6674</td>
</tr>
</tbody>
</table>

The short- and long-run estimates in table 4 indicate that results are dependent on the functional forms when applying the short-run and long-run cost functions respectively. The Cobb-Douglas model and the translog model show significant economies of scale for all size groups and the quadratic form shows constant or decreasing economies of scale. The two translog models indicate presence of scale economies, while the quadratic form indicates constant or decreasing returns to scale. However, while the findings of the translog model based on the envelope condition indicate that scale effects have a low variation between hospital size groups (0.69 for the largest hospitals and 0.67 for the smallest), the variation shows up as larger when we use the direct approach (0.70 for the smallest to 1.02 for the largest hospital).

The translog scale estimates for the largest size groups lie around 0.67 or very close to the value 1, equivalent to constant economies of scale in the long run. This means there is nothing sensitive to the functional form. Besides, the directly estimated translog model indicates that results are sensitive to the functional form. Overall, we identify significant to moderate long-run economies of scale, when applying two alternative long-run cost functions. However, using a quadratic functional form we identify constant economies of scale for the medium-size sub-groups and decreasing economies of scale for the largest size groups.
Optimal hospital size

The estimation based on (1) and the calculation of an ‘optimal’ hospital size based on (2) yields 204.9 beds for the median Danish hospital in the translog model and 275.2 beds for the quadratic functional form.

In figure 2 the estimated optimal hospital size is shown as a function of present size. Using the 45 degree line as point of departure, the figure illustrates how the ‘optimal’ size of each hospital deviates from the present size (‘45º’ line). The results of both models indicate that small and medium-sized hospitals with less than 204 or 275 beds are too small, while the larger hospitals are too large. However, it is not evident whether, for example, ‘small is too small’ in the translog model, since optimal and actual sizes are not different, at least not statistically. Both results are in line with the above-mentioned literature review by Aletras & Jones, which points to optimal sizes...
of hospitals as being between 200 and 400 beds. An example is a recent Canadian study, which estimated an 'optimal' value of 179.5 beds [9]. Applying the 'directly' estimated long-run cost function, it is possible to derive an optimal hospital size through (2). However, the long-run cost estimates in table 4 can be analyzed to reveal whether these indicate an optimal hospital size (357-1137 beds), and that the only functional form that is not very frequently used is the quadratic form. 

From an econometric perspective, the quadratic form for estimates of 'optimal' hospital size is preferred to the translog model. This is due to transformation problems in the translog model, which may result in biased cost estimates as well as biased estimates of economies of scale. Therefore, as mentioned above and found in [36], it is not possible to include more than one or two different output measures without getting multicollinearity, which, of course, leaves open critical issues regarding collapsing the multidimensional output vector, see, for instance, Table 5 shows how the estimated 'optimal' size (275 beds) is sensitive to partial changes in each of the estimated parameters included in the envelope condition (2). According to the estimated confidence interval, the 'optimal' number of beds is estimated to be in the interval 130 to 585 beds per hospital. 

Table 5. Sensitivity analysis for estimate of optimal hospital size

<table>
<thead>
<tr>
<th>Model parameters</th>
<th>Average number of beds</th>
<th>Total DRG value * Average number of beds</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quadratic model (275 beds)</td>
<td>130.0</td>
<td>235.2</td>
</tr>
<tr>
<td>Quadratic model (275 beds)</td>
<td>419.2</td>
<td>845.3</td>
</tr>
</tbody>
</table>

4. Discussion

This study shows that parametric estimates of economies of scale and of the optimal size of public hospitals in Denmark line with other studies. However, it is not possible to include more than one or two different output measures without getting multicollinearity, which, of course, leaves open critical issues regarding collapsing the multidimensional output vector, see, for instance, Table 5. 

The structural model requires all parts of the flexible functional model, including squared and interaction terms to be included, so that the model can capture varying economies of scale. Therefore, as mentioned above and found in [36], it is not possible to include more than one or two different output measures without getting multicollinearity, which, of course, leaves open critical issues regarding collapsing the multidimensional output vector, see, for instance, Table 5. 

In order to minimize these obstacles, the approach used in this study has been to aggregate the output vector to one aggregated output and to use one or two cost shifters in order to have sufficient degrees of freedom for model estimation without getting multicollinearity. Besides, the unit of analysis has been defined at hospital production units instead of conglomerate hospitals. This is done both to define the unit of analysis relevant for policy and to increase the number of observations that are higher for hospital production units than conglomerate hospitals.
The above-mentioned approach, which uses only one output index, is debatable. From one point of view, the multi-dimensional output can only be aggregated if the original dimensions are broad terms of, for example, quality, patient characteristics and institutional conditions in order to avoid bias due to omitted variables. On the other hand, empirical studies show that quality data, for example, can be significant in hospital cost functions, despite the belief in some quarters that better quality comes at a cost [17].

Furthermore, the Cobb-Douglas model has been applied to try to avoid multicollinearity. The lack of flexibility in the Cobb-Douglas model implies that fewer parameters have to be estimated than in the flexible cost function. From one point of view, this may be preferred in situations where only a relatively small number of degrees of freedom is available as in the small country case. A time wise bigger panel would give more degrees of freedom and increase the risk of misspecification. The flexibility of the Cobb-Douglas model also shows that it cannot be used as a supplementary measure as was the case in this study.

The Cobb-Douglas model implies constant scale elasticities, which decrease flexibility and increase the risk of misspecification. The inflexibility of the Cobb-Douglas model also shows that it cannot be used to calculate optimal scale and size as a function of size or activity in a way that can be seen from table 4. A more flexible model such as the translog function can be used as a supplementary measure. The implication is that the Cobb-Douglas model should only be used as a supplementary measure in this study.

In this paper we used only three years of panel data. The reason is that changes in the results compared to an ideally balanced panel. Outliers will simply influence too much. Instead the results of several functional forms are presented to measure sensitivity to different functional forms.

While the present study assumes that the quality of hospital output is homogeneous, the reasoning behind the assumption is a combination of a lack of recognized and objective approaches to recognize different measures. The reasoning is that it is difficult to secure a sufficient amount of degrees of freedom to include more variables to adjust for quality.

Furthermore, the explanation is that it is difficult to secure a sufficient amount of degrees of freedom. The reasoning is that it is difficult to secure a sufficient amount of degrees of freedom to include more variables to adjust for quality.
It is not realistic to assume that hospitals can adjust all their inputs quickly as was the case in the 'direct approach' [44]. Most studies indicate that hospitals cannot adjust all inputs quickly when output levels or factor prices change. Consequently, it is probably more reasonable to assume that hospitals only apply optimal quantities of the most easily adjustable variable inputs (manpower and medical supplies) given likely non-optimal levels of fixed inputs (measured in terms of beds). Therefore, it is more appropriate to estimate short-run cost functions and explain the envelope condition to derive the long-run cost function. A test for long-run equilibrium is assumed to be constant. For the short run Cobb-Douglas model and the directly estimated long-run translog model, the Hausman test confirmed that the fixed effect model is preferred to the random effect model. This indicates that there is a correlation between the individual effects and the covariates. As shown in table 1, it was not feasible to conduct the test for the short-run translog model. Both models failed to meet the assumptions of the Hausman test. Hence, the fixed effect model was preferred. It still has some disadvantages. One example is that time-invariant effects cannot be estimated per se. This prohibits relevant covariates being included directly such as teaching or university status, which are believed to be structural cost drivers in hospitals. However, FE models indirectly correct for individual differences through cross-sectional dimensions going to infinity [25]. It could be argued that the old technology is not representative for the future one time productivity gain from the rebuilding programme, e.g. better logistics and efficient patient pathways in new hospitals. However, FE models indirectly correct for individual differences through time series and cross-sectional dimensions going to infinity [25].

It is feasible to test whether the short-run scale estimates in table 4 are significantly different from one using the delta method or bootstrapping, see for instance, [46,47]. Since we are focusing on the long-run scale estimates we have omitted these tests. Since the FE approach is a version of regression analysis, cost function estimation usually requires that it is assumed that hospitals apply variable inputs in a cost-minimizing way or that they cannot alter the number of beds in the short run. The present approach which assumes constant prices does not take into consideration the possibility of allocative inefficiencies in other words, it is not known whether the unit cost will decline (be L-shaped) or will be U-shaped when hospital size increases above 1200 beds. Overall, this study supports the hypothesis that there may be cost advantages (or no disadvantages) for the smallest sub-group in producing hospital services in large hospitals even though it is outside the range of data used here. In other words, it is not known whether the unit cost will decline (be L-shaped) or will be U-shaped when hospital size increases above 1200 beds. Overall, this study supports the hypothesis that there may be cost advantages (or no disadvantages) for the smallest sub-group in producing hospital services in large hospitals even though it is outside the range of data used here. In other words, it is not known whether the unit cost will decline (be L-shaped) or will be U-shaped when hospital size increases above 1200 beds. Overall, this study supports the hypothesis that there may be cost advantages (or no disadvantages) for the smallest sub-group in producing hospital services in large hospitals even though it is outside the range of data used here. In other words, it is not known whether the unit cost will decline (be L-shaped) or will be U-shaped when hospital size increases above 1200 beds. However, policy conclusions should not be drawn solely on the basis of this study, which is solely based on panel data for the years 2004-2006. The findings reported here should be investigated further, e.g. through the use of alternative estimation methods such as data envelopment analysis (DEA) and estimation of potential efficiency gains from consolidation.
Appendix I

(1) Short run Quadratic (1a), Translog (1b) and Cobb Douglas (1c) cost functions:

\[ C(Q, K) = \beta_0 + \beta_1 Q_t + \beta_2 Q_t^2 + \beta_3 K + \beta_4 K^2 + \beta_5 Q_t K \]  
(1a)
\[ \ln C(Q, K) = \beta_1 Q + \beta_2 \ln Q + \beta_3 \ln Q^2 + \beta_4 \ln K + \beta_5 \ln K^2 + \beta_6 \ln Q \ln K \]  
(1b)
\[ \ln C(Q, K) = \beta_1 Q + \beta_2 \ln Q + \beta_3 \ln Q + \beta_4 \ln K + \beta_5 \ln K \]  
(1c)

(2) The optimal hospital production unit size measured in terms of beds is calculated from the short run cost functions (1a-c) by application of the envelope condition:

\[ \frac{\partial C(Q, K)}{\partial K} = \beta_1 + \beta_3 = 0 \quad \Rightarrow K = -\frac{\beta_3}{\beta_1} \]  
(2a)
\[ \frac{\partial \ln C(Q, K)}{\partial K} = \beta_1 + \beta_3 \ln K + \beta_6 \ln Q = 0 \quad \Rightarrow \ln K = -\frac{\beta_6}{\beta_1} \ln Q \quad \Rightarrow K = e^{-\frac{\beta_6}{\beta_1} \ln Q} \]  
(2b)
\[ \frac{\partial \ln C(Q, K)}{\partial K} = \beta_1 = 0 \quad \Rightarrow \text{unfeasible} \]  
(2c)

Calculation of long run cost function

The long run cost function is calculated from the short run cost function (1a, b) by substitution of the optimal number of beds (2a, 2b respectively) derived by the envelope condition (2). For the long run Quadratic cost function this yields:

\[ C(Q_t) = \beta_0 + \beta_1 Q_t + \beta_2 Q_t^2 + \beta_3 \left( -\frac{\beta_1}{\beta_3} - \beta_2 Q_t \right) + \beta_4 \left( -\frac{\beta_1}{\beta_4} - \beta_2 Q_t \right) + \beta_5 Q_t \left( -\frac{\beta_1}{\beta_5} - \beta_2 Q_t \right) \]

After mathematical reduction the long run cost function can be reduced to:

\[ C(Q_t) = \beta_0 + \beta_1 Q_t + \frac{\beta_2 Q_t^2}{2} \left( \beta_1 + \frac{\beta_2}{2} Q_t \right) \]

We omitted the long run Translog cost function since the only differences from the above mentioned quadratic cost function is that the total DRG-production value \( Q_t \) is replaced by logged levels.

Appendix II

Calculation of long run economies of scale

The expression for long run economies of scale (3a, 4a) is calculated from the long run cost function (2a) and (3, 4 respectively).

(3) Long run economies of scale - Translog cost function

\[ SE_1 = \frac{\partial \ln C(Q_t)}{\partial \ln Q_t} = \beta_1 = \frac{\beta_2}{\beta_1} + \frac{\beta_3}{\beta_2} + \frac{2 \beta_4}{\beta_3} \ln Q_t \]  
(3a)

By mathematical reduction expression (3a) can be reduced to the following:

\[ SE_1 = \frac{\beta_1 - \beta_3 + \beta_2 \ln Q_t}{\beta_1} \]  
(3b)

(4) Long run economies of scale - Quadratic cost function

\[ SE_2 = \left( \frac{\partial C(Q_t)}{\partial Q_t} \right) / \frac{C}{Q_t} = \left( \frac{\beta_2 Q_t^2}{2} + \frac{\beta_3 Q_t^2}{2} + \frac{\beta_4 Q_t^2}{2} \right) / \left( \frac{\beta_2}{2} Q_t^2 \right) \]  
(4a)

By mathematical reduction expression (4a) can be reduced to the following:

\[ SE_2 = \left( \beta_1 - \beta_3 + \beta_2 \ln Q_t \right) / \beta_1 \]  
(4b)
References


