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

Kristensen, Troels

Publication date: 2008

Document version
Final published version

Citation for published version (APA):

Terms of use
This work is brought to you by the University of Southern Denmark through the SDU Research Portal. Unless otherwise specified it has been shared according to the terms for self-archiving. If no other license is stated, these terms apply:

- You may download this work for personal use only.
- 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 this open access version

If you believe that this document breaches copyright please contact us providing details and we will investigate your claim. Please direct all enquiries to puresupport@bib.sdu.dk

Download date: 09. Sep. 2019
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 Helle Møller Jensen for helping in formatting and compiling the WP.

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

Troels Kristensen, Kim Rose Olsen and Kjeld Møller Pedersen
November 2008
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. From an economic perspective, the evidence base underpinning centralization is weak, that is to say that there is a conspicuous absence of studies of economies of scale in hospitals. 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 analysed. Notable exceptions are [6] and [7]. 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 analyse whether there are unexploited economies of scale in the current configuration and to estimate an optimal hospital size. The present study

Keywords: Economies of scale, optimal size, hospitals, cost function.

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 multi-sited 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 from 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.

---

2 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 changed from being an institution located on a single geographical site, often with a degree of division of labour and hence specialization. When estimating economies of scale and "optimal" hospital size this is a challenge. The cost and output data for hospital production sites are aggregated at the conglomerate level. This 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'.

Most studies argue that there is no evidence that hospitals operate in their long-run equilibrium, i.e. hospitals do not adjust all their inputs to their cost-minimizing levels [16]. 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 possible non-optimal levels of the final inputs in the short run. Hence, hospitals are only assumed to use cost-minimizing quantities of each adjustable variable input, 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 translog cost function

The translog cost function is most widely used in the field of hospital production. The translog model is chosen to be the natural logarithmic function. The Cobb-Douglas version is the special nested case of the translog model where squared and interaction terms in (1) are excluded. The exact model specifications are shown in the appendix. Since this study aims to allow \( x_{it} \) to contain cost shifters for hospital \( i \) in time period \( t \), the estimated models also belong to the family of hybrid cost functions [8]. This category of functional forms is now preferred to the more naive structural functional form, see, for instance [6].

Cost shifters controls for the fact that the cost of individual hospitals may be influenced by other external factors to the hospital management than the demand for output such as capital endowment, wages, epidemics and other academic challenges. Additional arguments for the assumption about constant prices are that it is a panel data study with a relatively short time dimension and the literature indicates the difficulty of measuring the input prices for hospitals e.g. the price of capital [6, 8].

Throughout the estimation, this increased flexibility is obtained at the cost of there being more parameters to be estimated than in more restricted functional forms such as the classical Cobb-Douglas cost function. One consequence is that the estimation of 'flexible' cost functions often results in multicollinearity problems, see, e.g. [6, 19], especially if attempts are made to disaggregate the outputs into more subgroups. This was also the case in this study. In other words, the closest approximation of an unknown functional form [7].
The estimated coefficients became insignificant and unstable, and the signs changed in such hospitals, for instance, different managerial abilities and severity of illness. Therefore it is less likely that FE models differ from random-effects models in some cases in cross-sectional estimation issues. FE models tend to be overidentified. Finally, the small number of observations, even in panel data, is an important issue in all small countries. These denominators make the choice of the constraints expensive. We have merged the multiple hospital outputs into two aggregated outputs two important tasks. The 'optimal' size of a hospital can be calculated from the short-run cost function by applying the envelope condition [9].

\[ K = \begin{vmatrix} a_1 & a_2 & \cdots & a_n \end{vmatrix} \]

Given the cost function (1), this calculation can be expressed as:

\[ K = \begin{vmatrix} a_1 & a_2 & \cdots & a_n \end{vmatrix} \]

Calculation of the optimal hospital size

The optimal 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 bed and output as defined in (1). 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 cost function and the long run quadratic cost function respectively.

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

The long-run cost function

We have merged the multiple hospital outputs into two aggregated outputs two important tasks. The Danish DRG system adjusts each output for casemix and to a certain extent for severity through the DRG cost weights attached to each discharge. Therefore, this study uses the DRG value to measure aggregated hospital output. In case the estimated short-run cost function is a translog form, yields a linear relationship between outputs and included cost shifters (if cost shifters are interacted with outputs) [9, 27]. Given the cost function (1), this calculation can be expressed as:

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

The short-run cost function

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 bed and output as defined in (1). 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 cost function and the long run quadratic cost function respectively.

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

The short-run cost function

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 bed and output as defined in (1). 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 cost function and the long run quadratic cost function respectively.

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

The short-run cost function

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 bed and output as defined in (1). 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 cost function and the long run quadratic cost function respectively.

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

The short-run cost function

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 bed and output as defined in (1). 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 cost function and the long run quadratic cost function respectively.

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

The short-run cost function

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 bed and output as defined in (1). 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 cost function and the long run quadratic cost function respectively.

\[ \frac{\partial C}{\partial K} = 0 \]
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):

\[ SE1 = \sum \frac{\partial g(Q)}{\partial (Q_m)} \]

\[ SE2 = \sum \left( \frac{\partial C}{\partial Q_m} / \frac{C}{Q_m} \right) \]

\[ SE1 \] 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 \( Q_m \).

\[ SE2 \] in (4) measures the sum of cost elasticities with respect to output. Each of the cost elasticities in \( SE2 \) 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 (1\textsuperscript{st} quartile) consists of the 25% of hospitals, which has the smallest number of beds, while the other size groups, 2\textsuperscript{nd} quartile, 3\textsuperscript{rd} quartile and 4\textsuperscript{th} quartile, include hospitals with a size in the respective quartiles. Both \( SE1 \) and \( SE2 \) 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.\(^\text{6}\) 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.

\[^{6}\text{The National Board of Health calculates adjusted actual operating costs by deducting from total reported operating costs, whenever relevant. This applies, for instance, to the cost of psychiatric services, laboratory services for general practitioners, the cost of medicines provided for outpatients, adjustments for differences in accounting practice and unpaid services between hospitals. This gives the figure for 'the adjusted operational costs' which is used in the present study.}\]
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>C</td>
<td></td>
<td>Adjusted operational costs</td>
<td>in 1000 DKK (Danish currency)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2004</td>
<td>-</td>
<td>530,934</td>
<td>648,465</td>
<td>21,581</td>
<td>1,591,319</td>
<td></td>
</tr>
<tr>
<td>2005</td>
<td>-</td>
<td>616,490</td>
<td>784,055</td>
<td>19,035</td>
<td>4,337,614</td>
<td></td>
</tr>
<tr>
<td>2006</td>
<td>-</td>
<td>439,160</td>
<td>449,149</td>
<td>16,761</td>
<td>1,890,084</td>
<td></td>
</tr>
<tr>
<td>Q</td>
<td></td>
<td>DRG value inpatient</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q&lt;sub&gt;1&lt;/sub&gt;</td>
<td>2004</td>
<td>328,514</td>
<td>379,452</td>
<td>0</td>
<td>2,123,694</td>
<td></td>
</tr>
<tr>
<td>2005</td>
<td>-</td>
<td>361,908</td>
<td>477,284</td>
<td>0</td>
<td>2,516,725</td>
<td></td>
</tr>
<tr>
<td>2006</td>
<td>-</td>
<td>232,824</td>
<td>231,748</td>
<td>3,265</td>
<td>898,218</td>
<td></td>
</tr>
<tr>
<td>Q&lt;sub&gt;2&lt;/sub&gt;</td>
<td>2004</td>
<td>209,637</td>
<td>270,439</td>
<td>4,909</td>
<td>1,180,949</td>
<td></td>
</tr>
<tr>
<td>2005</td>
<td>-</td>
<td>274,984</td>
<td>335,966</td>
<td>5,555</td>
<td>1,572,589</td>
<td></td>
</tr>
<tr>
<td>2006</td>
<td>-</td>
<td>226,344</td>
<td>247,555</td>
<td>15,163</td>
<td>1,898,966</td>
<td></td>
</tr>
<tr>
<td>Q&lt;sub&gt;T&lt;/sub&gt;</td>
<td>2004</td>
<td>538,152</td>
<td>592,734</td>
<td>22,968</td>
<td>3,304,644</td>
<td></td>
</tr>
<tr>
<td>2005</td>
<td>-</td>
<td>538,152</td>
<td>592,734</td>
<td>22,968</td>
<td>3,304,644</td>
<td></td>
</tr>
<tr>
<td>2006</td>
<td>-</td>
<td>459,168</td>
<td>468,859</td>
<td>15,163</td>
<td>1,898,966</td>
<td></td>
</tr>
<tr>
<td>K</td>
<td></td>
<td>Average number of staffed beds</td>
<td>in no. of beds and percentage of hospitals</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2004</td>
<td>-</td>
<td>281.6</td>
<td>250.9</td>
<td>25.6</td>
<td>1107.1</td>
<td></td>
</tr>
<tr>
<td>2005</td>
<td>-</td>
<td>265.1</td>
<td>259.5</td>
<td>9</td>
<td>1136.7</td>
<td></td>
</tr>
<tr>
<td>2006</td>
<td>-</td>
<td>176.0</td>
<td>152.1</td>
<td>9</td>
<td>517</td>
<td></td>
</tr>
</tbody>
</table>

*Unbalanced due to missing data for 2006. The numbers of observations in 2004-06 are 57, 55 & 31 respectively.

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. Frikliniken 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 2 Regression results – short-run cost functions

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></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 of in- and outpatients</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 (Avg. number of beds)</td>
<td>-</td>
<td>-0.0062</td>
<td>0.8131***</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>
</tbody>
</table>

**R**

<table>
<thead>
<tr>
<th></th>
<th>Within</th>
<th>Between</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>F-test (5,78)</td>
<td>28.11***</td>
<td>105.35***</td>
<td>30.10***</td>
</tr>
<tr>
<td>Hausman chi2(5)</td>
<td>8.49**</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

*** P < 0.01, ** P <0.05%, * P <0.10

1 Model did not meet the assumptions of the Hausman test

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 estimate 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 3 Regression and calculated result – long-run cost functions

<table>
<thead>
<tr>
<th></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>Total DRG value</td>
<td>-0.0045</td>
<td>-</td>
<td>0.3524</td>
</tr>
<tr>
<td>Inpatients DRG value</td>
<td>-</td>
<td>0.5388***</td>
<td>-</td>
</tr>
<tr>
<td>Outpatients DRG value</td>
<td>-</td>
<td>0.4100***</td>
<td>-</td>
</tr>
<tr>
<td>Inpatients DRG value</td>
<td>-</td>
<td>0.0879**</td>
<td>-</td>
</tr>
<tr>
<td>Outpatients DRG value</td>
<td>-</td>
<td>0.0994***</td>
<td>-</td>
</tr>
<tr>
<td>Inpatients*Outpatients</td>
<td>-</td>
<td>-0.1483***</td>
<td>-</td>
</tr>
</tbody>
</table>

**R**

<table>
<thead>
<tr>
<th></th>
<th>Within</th>
<th>Between</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>F-test (5,78)</td>
<td>-</td>
<td>27.36***</td>
<td>-</td>
</tr>
<tr>
<td>Hausman-test</td>
<td>12.02**</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

*** P < 0.01, ** P <0.05%, * P <0.10

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

1 In 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 form is due to the logarithmic transformations and the functional form. For example, the negative sign on the interaction term captures elements of cost complementarities between in- and outpatient activity, and the negative intercepts in the translog model will become positive after antilogarithmic transformations.

Table 4 shows estimates of scale for the alternative functional forms when applying the short-run cost 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.

All scale estimates are calculated for four size groups measured by the number of beds to obtain information on the shape of the cost curve. The policy implication of economies of scale for all size groups is an L-shaped average cost curve where average cost decreases when hospital output increases. This means the cheapest way of operating a hospital system would be to build hospitals that are as large as possible. In the extreme, we would plan one super hospital per region or a single hospital for the entire country. Diseconomies of scale imply that the average cost curve must be U-shaped because average costs increase as a function of output [30].

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

<table>
<thead>
<tr>
<th>Groups of hospitals</th>
<th>Short-run</th>
<th>Long-run</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cobb-Douglas</td>
<td>Translog</td>
</tr>
<tr>
<td>All hospitals</td>
<td>0.7160</td>
<td>0.7086</td>
</tr>
<tr>
<td>1st quartile</td>
<td>0.7160</td>
<td>0.6235</td>
</tr>
<tr>
<td>2nd quartile</td>
<td>0.7160</td>
<td>0.6688</td>
</tr>
<tr>
<td>3rd quartile</td>
<td>0.7160</td>
<td>0.6972</td>
</tr>
<tr>
<td>4th quartile</td>
<td>0.7160</td>
<td>0.7206</td>
</tr>
</tbody>
</table>

The short- and long-run estimates in table 4 indicate that results are dependent on the functional forms. 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.

In the short-run, the translog model expresses decreasing economies of scale and the quadratic model expresses decreasing diseconomies of scale as the size of hospitals are increased.

Overall the models suggest that results of the flexible models depend on the functional form used, while the nested Cobb-Douglas and the translog models yield similar 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 hospitals become larger.

However, in view of the major restructuring and centralisation that are to take place in Denmark over the next decade, it is less meaningful to base decisions on short-run scale estimates. Therefore, in the following, our focus is on the long-run scale estimates. The long-run scale estimates in table 4 show scale estimates based on (3) for the three alternative long-run cost curves, taking into account the fact that hospitals do not necessarily use the ‘optimal’ capital in the short run and that hospitals can change the amount of capital according to the activity level in the long run. Long-run results and short-run results are similar in the sense that the results are sensitive to the functional form. Besides, the directly estimated translog model indicates that results are sensitive to the two alternative long-run estimation approaches.

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 hospitals).

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 term. Thus, there is nothing in the translog models to indicate that the hospitals experience diseconomies of scale in the long run, as the scale estimates are not significantly above the value 1.

Overall, we identify significant to moderate 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.
Figure 1 shows the long-run (LR) scale estimates as a function of hospital size for the three different LR model specifications. The figure shows that LR scale estimates for the translog model using the envelope condition lie below 1 for all hospital sizes, whereas they start to exceed 1 for hospitals above around 400 beds in the direct translog LR model. The estimates based on the quadratic form show less correlation between hospital size and LR scale estimates even though a positive trend can be detected with increasing size of hospitals. This increased level of noise probably stems from the lack of compression of outliers in the unlogged model.

In contrast to table 3, which showed scale estimates for the median hospitals data in each size group (representative units), figure 1 shows the short-run scale characteristics for all observed hospitals. The smallest quartile has estimates in the range of 50.9-229.0 beds, while the following three size groups (2nd quartile, 3rd quartile and 4th quartile) have observations in the intervals 229.0-356.6 and 356.6-1136.7 beds.

Finally, figure 1 shows that there are three outlier observations that we did not find any arguments to exclude and that there are relatively few observations among the largest hospital production sites.

**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. Another example is a recent Canadian study [33]. Still, due to the fact that input prices were assumed to be constant, this study was limited to adopting a second-best practice where it was impossible to test all cost function forms for the nature of the flexible forms imposes through squared and interactions terms. Besides, the unit of analysis has been defined as 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.

### Table 5: Sensitivity analysis for estimate of optimal hospital size

<table>
<thead>
<tr>
<th>Parameter</th>
<th>95% Confidence Interval</th>
<th>Quadratic model (275 beds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average number of beds</td>
<td>130.0 – 419.2</td>
<td>275 beds</td>
</tr>
<tr>
<td>Total DRG value * Average number of beds</td>
<td>235.2 – 282.3</td>
<td>275 beds</td>
</tr>
</tbody>
</table>

4. Discussion

The study shows that parametric estimates of economies of scale and of the optimal size of public hospitals in Denmark are in line with other studies and the chosen model specification. Theory makes assumptions about second-order elasticities of cost functions, i.e. marginal cost elasticity and the envelope condition. This study is based on the relatively well-known properties of parametric flexible cost functions guided by neoclassical production theory [33]. Still, due to the fact that input prices were assumed to be constant, this study was limited to adopting a second-best practice where it was impossible to test all cost function forms for
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 enough to include variables to describe hospital products in 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].

From another point of view, it can be argued that this is exactly what is necessary 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 functional form. From one point of view, this may be preferred to situations where only a relatively small number of degrees of freedom are available as in the small country case. The Cobb-Douglas model implies constant returns to scale, which is not consistent with the data. The Cobb-Douglas model may also be used as a supplementary measure of the degree of flexibility. A more flexible model would allow the degree of freedom and the risk of misspecification to be estimated. As mentioned earlier, in the literature the number of beds per hospital has become a standard method to measure hospital size. However, it could be claimed that this is exactly what is necessary to avoid bias due to omitted variables.

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 functional form. From one point of view, this may be preferred to situations where only a relatively small number of degrees of freedom are available as in the small country case. The Cobb-Douglas model implies constant returns to scale, which is not consistent with the data. The Cobb-Douglas model may also be used as a supplementary measure of the degree of flexibility. A more flexible model would allow the degree of freedom and the risk of misspecification to be estimated. As mentioned earlier, in the literature the number of beds per hospital has become a standard method to measure hospital size. However, it could be claimed that this is exactly what is necessary to avoid bias due to omitted variables.

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 functional form. From one point of view, this may be preferred to situations where only a relatively small number of degrees of freedom are available as in the small country case. The Cobb-Douglas model implies constant returns to scale, which is not consistent with the data. The Cobb-Douglas model may also be used as a supplementary measure of the degree of flexibility. A more flexible model would allow the degree of freedom and the risk of misspecification to be estimated. As mentioned earlier, in the literature the number of beds per hospital has become a standard method to measure hospital size. However, it could be claimed that this is exactly what is necessary to avoid bias due to omitted variables.
It is not realistic to assume that hospitals can adjust all their inputs quickly as was done in the present study. Individual effects were not used to calculate technical inefficiencies. The estimated scale economies and optimal hospital size are solely based on the interpretation of the long-run cost function in (1). On the other hand, Mester [49] among others has concluded that the quality of the estimation of economies of scale.

Overall, this study shows that when applying alternative specifications of translog cost functions, there are significant to moderate long-run economies of scale for all size groups of Danish hospitals in 2004-2006. These results indicate a U-shaped unit cost curve. However, Danish hospitals in the smallest groups with hospitals in the medium-sized groups. In both cases, it is necessary to take into account the technical inefficiencies that may not be accounted for in the analysis of hospital production, such as the need for local emergency facilities, transport costs and opportunity costs through increased travel time. Furthermore, the optimal number of beds per hospital is estimated to be 275 beds per site with a 95% confidence interval between 130 to 585 beds per hospital. This is roughly in line with international results.

The analysis prescribes drawing conclusion about the consolidation of hospitals leading to hospital size exceeding 1200 beds, because it is outside the range of data used here. In other words, it is not known whether the unit cost will decline (be U-shaped) or it will be L-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 in the Danish hospital system than was the case in 2004-2006. 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, for example, through the use of other supplementary data, analysis of cost functions, and estimation of potential efficiency gains from consolidation.

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, data envelopment analysis (DEA), and alternative estimation methods such as data envelopment, analysis specification, may be used for technical efficiency analysis. The FE approach correctly accounts for technical inefficiencies since, for instance, the unobserved firm-specific cross-sectional heterogeneity may be assumed away [48].
Appendix I

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

\[ C(Q_t, K_t) = \beta_0 + \beta_1 Q_t + \frac{1}{2} \beta_2 Q_t^2 + \beta_3 K_t + \frac{1}{2} \beta_4 K_t^2 + \beta_5 Q_t K_t \]  
\[ \ln C(Q_t, K_t) = \beta_0 + \beta_1 \ln Q_t + \frac{1}{2} \beta_2 \ln Q_t^2 + \beta_3 \ln K_t + \frac{1}{2} \beta_4 \ln K_t^2 + \beta_5 \ln Q_t \ln K_t \]  
\[ \ln C(Q_t, K_t) = \beta_0 + \beta_1 \ln Q_t + \beta_2 \ln Q_t^2 + \beta_3 \ln K_t + \beta_5 \ln K_t \]  

(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:

\[ \frac{\partial C(Q_t, K_t)}{\partial K_t} = \beta_1 + \beta_3 K_t + \beta_5 Q_t = 0 \Rightarrow K = \frac{-\beta_1 - \beta_5 Q_t}{\beta_3} \]  
\[ \frac{\partial \ln C(Q_t, K_t)}{\partial K_t} = \beta_1 + \beta_3 \ln K_t + \beta_5 \ln Q_t = 0 \Rightarrow \ln K = \frac{-\beta_1 - \beta_5 \ln Q_t}{\beta_3} \Rightarrow K = e^{\frac{-\beta_1 - \beta_5 \ln Q_t}{\beta_3}} \]  
\[ \frac{\partial \ln C(Q_t, K_t)}{\partial K_t} = \beta_1 = 0 \Rightarrow \text{unfeasible} \]  

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:

\[ C(Q_t) = \beta_0 + \beta_1 Q_t + \frac{1}{2} \beta_2 Q_t^2 + \beta_3 \left( \frac{-\beta_1 - \beta_5 Q_t}{\beta_3} \right) + \frac{1}{2} \beta_4 \left( \frac{-\beta_1 - \beta_5 Q_t}{\beta_3} \right)^2 + \beta_5 Q_t \left( \frac{-\beta_1 - \beta_5 Q_t}{\beta_3} \right) \]

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

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

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

\[ SE1 = \frac{\partial \ln C(Q_t)}{\partial \ln Q_t} = \beta_1 \left( -\frac{2 \beta_2 \beta_3}{\beta_1} + \frac{\beta_3 \beta_5}{\beta_1} \right) Q_t \]  
\[ \frac{\beta_3 \beta_5}{\beta_1} \]  

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

\[ SE1 = \frac{\beta_3 \beta_5}{\beta_1} \left( \frac{-2 \beta_2 \beta_3}{\beta_1} Q_t + \frac{\beta_3 \beta_5}{\beta_1} \right) \]

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

\[ SE2 = \frac{\partial C(Q_t)}{\partial Q_t} = \beta_1 \frac{\beta_3 \beta_5}{\beta_1 \beta_2} + \frac{\beta_3 \beta_5}{\beta_1^2} \left( \frac{-2 \beta_2 \beta_3}{\beta_1} - \frac{\beta_3 \beta_5}{\beta_1} \right) Q_t \]

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

\[ SE2 = \frac{\beta_3 \beta_5}{\beta_1} \left( \frac{-2 \beta_2 \beta_3}{\beta_1} - \frac{\beta_3 \beta_5}{\beta_1} \right) Q_t \]
References


4. Reference to a non-extracted text (e.g., a technical report or a book).

5. Reference to a conference paper or an abstract.

6. Reference to a dataset or a software tool.

7. Reference to a preprint or a working paper.

8. Reference to a software package or a programming library.

9. Reference to an internet resource or a website.

10. Reference to an oral presentation or a talk.

11. Reference to a letter or a comment.

12. Reference to a blog post or a social media commentary.

13. Reference to a review article or a meta-analysis.

14. Reference to a newsletter or a news bulletin.

15. Reference to a policy brief or a brief for policymakers.

16. Reference to a presentation slides or a presentation at a conference.

17. Reference to a government report or a policy document.

18. Reference to a court decision or a legal opinion.

19. Reference to a military operation or a strategic move.

20. Reference to a legal proceeding or a lawsuit.

21. Reference to a dictionary or a thesaurus.

22. Reference to an encyclopedia or an academic textbook.

23. Reference to a dictionary or a glossary.

24. Reference to a dictionary or a thesaurus.

25. Reference to a dictionary or a thesaurus.

26. Reference to a dictionary or a thesaurus.

27. Reference to a dictionary or a thesaurus.

28. Reference to a dictionary or a thesaurus.

29. Reference to a dictionary or a thesaurus.

30. Reference to a dictionary or a thesaurus.

31. Reference to a dictionary or a thesaurus.

32. Reference to a dictionary or a thesaurus.

33. Reference to a dictionary or a thesaurus.

34. Reference to a dictionary or a thesaurus.

35. Reference to a dictionary or a thesaurus.

36. Reference to a dictionary or a thesaurus.

37. Reference to a dictionary or a thesaurus.

38. Reference to a dictionary or a thesaurus.

39. Reference to a dictionary or a thesaurus.

40. Reference to a dictionary or a thesaurus.

41. Reference to a dictionary or a thesaurus.

42. Reference to a dictionary or a thesaurus.

43. Reference to a dictionary or a thesaurus.

44. Reference to a dictionary or a thesaurus.

45. Reference to a dictionary or a thesaurus.

46. Reference to a dictionary or a thesaurus.

47. Reference to a dictionary or a thesaurus.

48. Reference to a dictionary or a thesaurus.

49. Reference to a dictionary or a thesaurus.

50. Reference to a dictionary or a thesaurus.


