An Ordinal Logistic Regression Model to Estimate Cumulative Odds of Length-of- Stay for Efficient Hospitals in DEA Setting: The Case of Oman

Parakramaweera Sunil Dharmapala
Dept. of Operations Management & Business Statistics
College of Commerce & Economics, Sultan Qaboos University
Sultanate of Oman

Abstract

In this paper, we propose a methodology that comprises data envelopment analysis (DEA) and logistic regression to estimate the cumulative odds of length-of-stay for efficient hospitals. Here we evaluate technical efficiency of 45 hospitals in the Sultanate of Oman, using both Charnes-Cooper-Rhodes (CCR) and Banker-Charnes-Cooper (BCC) DEA models with 3 inputs and 2 outputs. Then, using ordinal logistic regression we estimate the cumulative odds of 3 categories of length-of-stay for 3 classes of hospitals, in the presence of efficiency and 2 other explanatory variables. Finally, we compare the cumulative odds of length-of-stay for efficient hospitals under CCR and BCC models.

1. Introduction

The modeling of length-of-stay (LOS) for inpatients in hospitals has important implications in various aspects of health care management (Ng et al., 2003). For many health care management systems, it is important to develop a comprehensive analysis of LOS and to identify hospital-related and patient-related characteristics influencing LOS variations (Xiao et al., 1997). In health care research, LOS is divided into 3 categories: long-term, medium-term, and short-term (Nguyen et al., 2005). In this paper, we propose a statistical modeling procedure to estimate the cumulative odds of 'LOS per Discharge' for the 3 categories, using ordinal logistic regression (OLR) and data envelopment analysis (DEA).

2. Methodology

In the first stage of our analysis, we use BCC and CCR models to compute DEA efficiency scores of 45 hospitals in the Sultanate of Oman, in the year 2002. The hospitals are divided into 3 categories: Local, County, and Regional. The inputs considered are, recurrent expenditure, bed capacity, total staff, and the outputs are, total LOS and live discharges. In the second stage, with 'LOS per Discharge' divided into 3 categories (ordinal classes) as the response variable, we run 2 OLR models: one with CCR efficiency scores, and the other with BCC efficiency scores, as explanatory variables. Both OLR models will have 2 other explanatory variables, namely, bed occupancy rate and average inpatients per day. Finally, we establish the relationship between the cumulative odds of LOS categories under CCR and BCC technical efficiency, using the OLR coefficient estimates.

3. Preliminaries of DEA

DEA is a mathematical programming methodology that provides non-parametric measures of optimal relative efficiency. It identifies the Decision Making Units (DMUs) on the 'efficient productivity frontier' as efficient firms and DMUs that are interior to that frontier as inefficient firms. Many outputs and inputs can be analyzed simultaneously for an arbitrary number of observations. Relative efficiency measurements are computed DMU-by-DMU across all of the DMUs under consideration, for the same inputs and outputs of data. DMUc denotes the selected DMU for comparison. The input/output data entries must be non-negative, with zero entries allowed. (Charnes et al., 1978)

A DEA data domain consists of n DMUs. The selected DMUc (c = 1, 2,…, n) is characterized by an input vector Xc and an output vector Yc . he matrices X and Y contain input and output vectors respectively, for all DMUs. U, the output multiplier and V, the input multiplier, are unknowns that need to be determined by solving the respective linear programming (LP) model stated below. (All vectors are column vectors, and [...]T stands for transpose).

As a modification to the original input-oriented CCR (Charnes et al., 1978) and BCC (Banker et al., 1984) models, Thompson et al. (1993) introduced the following LP formulations (Archimedean form) of the CCR and BCC model for DMUc.

CCR Model (Dual)

Max z = UT YC

YCAT (j)	Y (days)	Cut-off point (T_j)
Short-term (1)	$0 \le Y \le 2$	$T_1 = 2$ days
Medium-term (2)	$2 < Y \le 5$	$T_2 = 5$ days
Long-term (3)	$5 < Y < \infty$	$T_3 = \infty$

s.t.
VT XC = 1 (1)
UT Y - VT X 0
. U 0, V 0

BCC Model (Dual)

Max z = UT YC + u*
s.t.
VT XC = 1 (2)
UT Y - VT X + u* I 0
u* unrestricted
U 0, V 0

The efficiency value of DMUc in CCR or BCC model is denoted by the optimal value c*.

A DMUc with c* = 1 is said to be 'scale efficient.' The class of scale efficient DMUs can be partitioned into 3 sub-classes; (1) DEA-extreme-efficient DMUs in class E, which are at the vertices on the frontier, (2) DEA-non-extreme-efficient DMUs in class E', which are on the frontier between vertices, and (3) DEA-inefficient DMUs in class F which are on the extended frontier. DMUc with 0 < c* < 1 is said to be 'scale inefficient' and in class N (Charnes et al. 1991).

Scale efficient classes E and E' are also called 'technically efficient'. Class F is scale efficient but not technically efficient because the optimal slacks are present, and class N is both scale and technically inefficient. All these classes are mutually exclusive. In our terminology, $E \; Y \; E'$ form the 'technically efficient' class, and $F \; Y \; N$ form the 'technically inefficient' class.

4. Ordinal logistic regression model

First, we define the continuous response variable Y, 'LOS per Discharge', as the ratio of LOS to Total Discharges (including deaths). Then, we divide Y into 3 ordinal categories, short-term, medium-term, and long-term, and name it as YCAT.

The Cumulative Odds are computed from the following OLR model:

$$Log_e\left[\frac{P(YCAT \le j)}{1 - P(YCAT \le j)}\right] = \alpha_j + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \;;\; j = 1, 2$$

. (Kutner et al., 2004)

For j =1, we have, Odds in favor of short-term Y against medium or long term Y

For j =2, we have, Odds in favor of short or medium-term Y against long-term Y

An important feature of the above model is the following:

For categories j = 1 and 2, only the intercept αj is different, but the slope coefficient βi of Xi is the same. This makes the interpretation of βi simpler.

We also notice that, (i) P(YCAT ≤ 3) = 1, for j =3, and

(ii) 1– P(YCAT ≤ j) = P(YCAT > j), for j = 1, 2.

4.1 Interpretation of βi

For 1 unit change in Xi, the corresponding Cumulative Odds Ratio (COR) is defined as, Odds(1)/Odds(0).

Here, $Odds(1) = \dfrac{P(YCAT_1 \le j)}{P(YCAT_1 > j)}$, and $Odds(0) = \dfrac{P(YCAT_0 \le j)}{P(YCAT_0 > j)}$, where YCAT0 is the

Y-category before 1 unit change in Xi, YCAT1 is the Y-category after 1 unit change.

Then, COR = Odds(1)/Odds(0) = exp(βi) ; i = 1, 2, 3.

5. Computational results

5.1 DEA results

DEA provided the following efficiency evaluation of 45 hospitals.

	CCR model	BCC model
Efficient hospitals	13	19
Inefficient hospitals	32	26

5.2 OLR results

The OLR estimates of αj and βi are, a1, a2, b1, b2, and b3, respectively. We obtain these estimates and test results, by running the following two OLR models in Minitab.

OLR-1: Response = YCAT; X1 = Bed occupancy rate (%), X2 = Inpatients per day
X3 = CCR efficiency (0- inefficient, 1- efficient)

OLR-2: Response = YCAT; X1 = Bed occupancy rate (%), X2 = Inpatients per day
X3 = BCC efficiency (0- inefficient, 1- efficient)

Table 1: OLR test results

	CCR / OLR-1	CCR / OLR-1-2	BCC / OLR-2	BCC / OLR-2-2
Const-1, α1	p-value=0.888	p-value=0.563	p-value=0.583	p-value=0.904
Const-2, α2	p-value=0.000	p-value=0.000	p-value=0.000	p-value=0.000
BedOccupRt (β1)	p-value=0.039	p-value=0.002	p-value=0.013	p-value=0.001
InPatients (β2)	p-value=0.301	Dropped **	p-value=0.163	Dropped **
Efficiency (β3)	p-value=0.228	p-value=0.108 (α = 0.11)	p-value=0.01	p-value=0.008
Likelihood Ratio Test				
H0: All slope coefficients = 0	p-value=0.004	p-value=0.002	p-value=0.000	p-value=0.000

** InPatients is dropped from the model in the 2nd run

The test results above show that OLR models with BCC or CCR efficiency and without InPatients can be used to estimate the Cumulative Odds given by,

$$\frac{P(YCAT \le j)}{P(YCAT > j)} = \exp(a_j + b_1 X_1 + b_3 X_3)\ ;\ j = 1, 2\ .$$

Then, estimated COR = exp(bi) ; i = 1, 3.

Table 2: OLR estimates

Coefficient	CCR model	BCC model	COR for CCR	COR for BCC
a1	0.40842	-0.08777		
a2	4.5316	4.5042		
b1	-0.07412	-0.08061	0.93	0.92
b3	1.41821	2.12095	4.13	8.34

COR for X3 with BCC efficiency is 8.34, and it says, “Cumulative odds in favor of short or medium LOS (YCAT ≤ 2) for a BCC-efficient hospital is 8.34 times higher than that for a BCC-inefficient hospital.”

COR for X3 with CCR efficiency is 4.13, and it says, “Cumulative odds in favor of short or medium LOS (YCAT ≤ 2) for a CCR-efficient hospital is 4.13 times higher than that for a CCR-inefficient hospital.”

6. Summary and conclusion

The statistical modeling procedure that we proposed in this paper to estimate the cumulative odds of LOS categories comprised data envelopment analysis and ordinal logistic regression. The CCR and BCC efficiency scores in DEA yielded different OLR estimates for cumulative odds of LOS categories.

The regression test results showed that one of the two explanatory variables, inpatients, is not significant in estimating the cumulative odds. The regression estimates showed that the cumulative odds in favor of short or medium LOS for an efficient hospital against an inefficient hospital is higher under BCC-efficiency than under CCR-efficiency by a ratio of 8.34 to 4.13. DEA efficiency evaluation revealed that 6 hospitals that were inefficient under CCR model became efficient under BCC model. This increase in efficiency may have contributed to the increase in cumulative odds under BCC efficiency. Thus, the BCC model outperformed the CCR model by more than twice, in terms of cumulative odds in favor of short or medium LOS for efficient hospitals.

As for future research, this methodology can be extended, by including other explanatory variables that seem relevant to length-of-stay, and also by imposing bounds on the multipliers (input and output) to restrict the solutions of CCR and BCC models.

References

Banker, R.D., A. Charnes and W.W. Cooper (1984), “Models for estimating technical and scale efficiencies in DEA,” Management Science, vol. 30, pp. 1078-1092

Charnes A., W.W. Cooper and E. Rhodes (1978), “Measuring efficiency of decision making units,” European Journal of Operational Research, vol.2, pp. 429-444

Charnes A., W.W. Cooper and R.M. Thrall (1991), “A structure for classifying and characterizing efficiency and inefficiency in Data Envelopment Analysis,” Journal of Productivity Analysis, vol. 2, pp. 197-237

Kutner, M.H., C.J. Nachtsheim, J. Neter (2004), Applied Regression Analysis, 4th ed.

Ng, S.K., K.K.W. Yau and A.H. Lee (2003), “ Modelling inpatient length of stay by a hierarchical mixture regression via the EM algorithm ,” Mathematical and Computer Modelling, vol. 37, pp. 365-375

Nguyen, J.M., P. Six, D. Antonioli, P. Glemain, G. Potel, P. Lombrail and P. Le Beux (2005), “A simple method to optimize hospital beds capacity,” International Journal of Medical Informatics, vol. 74, pp. 39-49

Thompson, R.G., P.S. Dharmapala and R.M. Thrall (1993), “Importance for DEA of zeros in data and multipliers,” Journal of Productivity Analysis, vol. 4, pp. 337-348

Xiao, J., D. Douglas, A. H. Lee and S. R. Vemuri (1997), “ A Delphi evaluation of the factors influencing length of stay in Australian hospitals,” International Journal of Health Planning and Management, vol. 12(3), pp. 207-218