Resampling in Data Envelopment Analysis illustrated by a hospital example

Kaoru Tone

National Graduate Institute for Policy Studies,

7-22-1 Roppongi, Minato-ku, Tokyo 106-8677, Japan

tone@grips.ac.jp

Abstract: In this paper, we propose new resampling models in data envelopment analysis (DEA). Input/output values are subject to change for several reasons, e.g., measurement errors, hysteretic factors, arbitrariness and so on. Furthermore, these variations differ in their input/output items and their decision-making units (DMU). Hence, DEA efficiency scores need to be examined by considering these factors. Resampling based on these variations is necessary for gauging the confidence interval of DEA scores. We propose two resampling models. The first model utilizes historical data, e.g., past-present, for estimating data variations, imposing chronological order weights which are supplied by Lucas series (a variant of Fibonacci series). The second one deals with future prospects. This model aims at forecasting the future efficiency score and its confidence interval for each DMU. We applied our models to dataset composed of Japanese municipal hospitals.

Keywords: Data variation; resampling; confidence interval; past-present-future DEA; hospital

1 Introduction

The treatment of data variations by statistical methods has taken a variety of forms in DEA. Banker [1] and Banker and Natarajan [2] show that DEA provides a consistent estimator of arbitrary monotone and concave production functions when the (one-sided) deviations from such a production are degraded as stochastic variations in technical inefficiency.

Several authors developed the sensitivity analysis of DEA scores, e. g. Charnes and Neralić [6], Neralić [11] and Zhu [18]. In Neralić [11], the case of the additive changes of inputs and outputs of an efficient DMU preserving its efficiency is studied first. Then, the cases of the proportionate changes inputs or/and outputs with two coefficients of proportionality are studied. This study utilizes mainly the sensitivity analysis regarding the basis matrix changes of linear programming. In Zhu [18], he proposed a sensitivity analysis of DEA models by using various super-efficiency DEA models in which a test DMU is not included in the reference set. This sensitivity analysis approach simultaneously considers the data perturbations in all DMUs, namely, the change of the test DMU and the remaining DMUs. However, the above researches did not explicitly deal with resampling problems and confidence interval issues, rather than the range analysis.

Simar and Wilson [11, 12] turn to "bootstrap methods" which are modifications of Efron [9]. Their bootstrapping utilizes a subset of original DMUs and data variations and measurement errors are not accounted in their model.

Barnum et al. [3] applied the statistical Panel Data Analysis (PDA) for estimating confidence intervals of DEA score for individual DMUs. They succeeded in complementing Simar and Wilson's bootstrapping by using the panel data and PDA methodologies. They developed a statistical method for estimating DEA score confidence intervals for individual organizations or other entities. This might be a pinnacle by statistical

approaches.

In this paper, we follow the principles stated in Cook, Tone and Zhu [7] and believe that DEA performance measures are relative, not absolute, and frontiers-dependent. DEA scores undergo a change depending on the choice of inputs, outputs, DMUs and models. They are evaluated by corresponding mathematical programming methods. Our approach deals with measurement variations or errors in input and output data directly and resamples data depending on historical data. Thus, the production possible set for the entire DMUs differs at every resample. We solve the frontier problem using the non-oriented slacks-based super-efficiency model. Hence, our approach presents another direction for resampling DEA scores than Barnum et al.'s statistical one.

Throughout this paper, we assume that the dataset is free from outliers,[1] homogenous in the kind of DMUs (e.g. hospitals, banks or universities in the same category) and not volatile, as otherwise the results are unreliable.

This paper unfolds as follows. Section 2 explains the dataset we used in this study and shows preliminary results. Section 3 (Model 1) deals with historical data for estimating the distribution of input/output data and thus we learn the distribution of input/output values from history. We resample data using the discrete distribution with Lucas number weights to past-present data. In section 4 (Model 2), we extend the approach presented in section 3 to future forecast data and resample future data depending on the past-present-future inter-temporal distribution. For forecasting, we utilize the trend, the weighted average or the average of the trend and weighted average provided by past-present data. In all cases, we apply Fisher's z-transformation to check the resampled data, and we utilize the non-oriented super-efficiency model and obtain the confidence intervals. Section 5 concludes the paper.

2 The data and preliminary results

In this section, we introduce the dataset utilized in this study and demonstrate preliminary results.

2.1 The data

Throughout this study we utilize the dataset concerning nineteen municipal hospitals from 2007 to 2009 in Japan. There are approximately 1,000 municipal hospitals in Japan and there is large heterogeneity among them. We selected nineteen municipal hospitals with more than 400 beds. Therefore, this sample may represent larger acute-care hospitals with the homogeneous functions owned by Japanese municipals. The data were collected from the Annual Databook of Local Public Enterprises published by the Ministry of Internal Affairs and Communications. We chose two inputs Doctor ((I)Doc) and Nurse ((I)Nur), and two outputs Inpatient ((O)In) and Outpatient ((O)Out) for this study. Table 1 exhibits the data, while Table 2 shows main statistics. They are all average numbers per year. We have no daily or monthly data. The Japanese government's fiscal year begins on April 1 and ends on March 31. The data are the yearly average of the fiscal year data.

[1] For outlier detection, see Yang et al. [17] and references therein.

Table 1: The data

	2007				2008				2009			
DMU	(I)Doc	(I)Nur	(O)In	(O)Out	(I)Doc	(I)Nur	(O)In	(O)Out	(I)Doc	(I)Nur	(O)In	(O)Out
H1	108	433	606	1,239	114	453	617	1,244	116	545	603	1,295
H2	125	448	642	1,363	133	499	638	1,310	136	482	618	1,300
H3	118	567	585	1,072	121	600	569	1,051	125	616	561	1,071
H4	138	541	699	1,210	138	531	704	1,194	140	554	679	1,182
H5	138	613	653	1,195	142	616	644	1,147	137	633	622	1,147
H6	99	569	716	1,533	106	592	701	1,478	109	613	651	1,457
H7	94	498	540	1,065	103	494	551	1,067	101	491	540	1,067
H8	106	461	496	1,051	118	490	504	1,033	133	479	505	1,081
H9	109	450	483	851	119	483	487	877	121	501	486	904
H10	102	540	581	1,268	106	558	565	1,278	148	611	586	1,321
H11	92	495	490	1,217	101	497	501	1,146	102	501	479	1,113
H12	148	721	771	1,637	147	710	723	1,657	158	737	743	1,714
H13	103	593	679	2,011	106	673	642	1,883	120	697	634	1,872
H14	101	500	613	1,868	110	519	617	1,894	116	517	623	2,009
H15	159	793	964	2,224	160	801	906	2,148	166	817	877	2,155
H16	77	354	410	1,047	68	359	391	916	81	378	406	897
H17	111	663	717	1,674	112	645	702	1,774	112	663	709	1,733
H18	62	388	480	913	64	385	467	907	63	381	463	872
H19	98	323	508	1,192	95	314	483	1,018	95	320	490	1,034

Table 2: Main statistics

	2007				2008				2009			
	(I)Doc	(I)Nur	(O)In	(O)Out	(I)Doc	(I)Nur	(O)In	(O)Out	(I)Doc	(I)Nur	(O)In	(O)Out
Avg	110	524	612	1,349	114	538	601	1,317	120	555	593	1,328
Max	159	793	964	2,224	160	801	906	2,148	166	817	877	2,155
Min	62	323	410	851	64	314	391	877	63	320	406	872
StDev	23.75	120.41	130.51	378.24	24.15	121.43	119.57	380.07	25.58	126.78	113.05	389.49

2.2 The model

In this paper, we utilize the non-oriented super slacks-based measure model (Tone [15]) under the constant returns-to-scale (CRS) assumption for evaluating the relative efficiency. This model is an extension of the SBM (slacks-based measure; Tone [14]).

The reasons why we utilize this model are as follows, although we can apply other models, e.g. radial or oriented, as well.

a CRS assumption

Hospitals in this study are located in the urban districts of their municipal area and have similar functions as hospital. Hence, we can compare them under the constant returns-to-scale assumption. However, we can apply the variable returns-to-scale model if scale merits or demerits are identified.

b SBM model

As a non-radial model, the slacks-based measure (SBM) is appropriate for taking account of input and output slacks which affect efficiency scores directly,

whereas the radial models mainly concern with the proportional changes of inputs or outputs. Thus, SBM scores are more sensitive to data variations than the radial models. Furthermore, the non-oriented SBM can deal with input-surpluses and output-shortfalls within the same scheme.

c Super-SBM model

Most DEA scores are bounded by unity (≤1, or ≥1). This encounters difficulties in comparing efficient DMUs. The super-efficiency models can compare them by removing the bound.

2.3 Preliminary results: By Year

We solved the data year by year and obtained the super-efficiency scores in Table 3. As can be seen, the scores fluctuate by year. This suggests need for analysis of data variation.

Table 3: Super-SBM score

	2007	2008	2009
H1	0.883	0.905	0.754
H2	0.875	0.801	0.779
H3	0.623	0.615	0.592
H4	0.700	0.765	0.680
H5	0.619	0.620	0.604
H6	1.004	0.942	0.848
H7	0.719	0.732	0.725
H8	0.676	0.651	0.631
H9	0.588	0.583	0.568
H10	0.758	0.764	0.631
H11	0.757	0.740	0.698
H12	0.711	0.741	0.714
H13	1.034	1.025	0.831
H14	1.039	1.107	1.145
H15	0.858	0.857	0.811
H16	0.831	0.847	0.742
H17	0.847	0.948	0.937
H18	1.034	1.050	1.074
H19	1.071	1.072	1.100
Avg	0.822	0.830	0.782

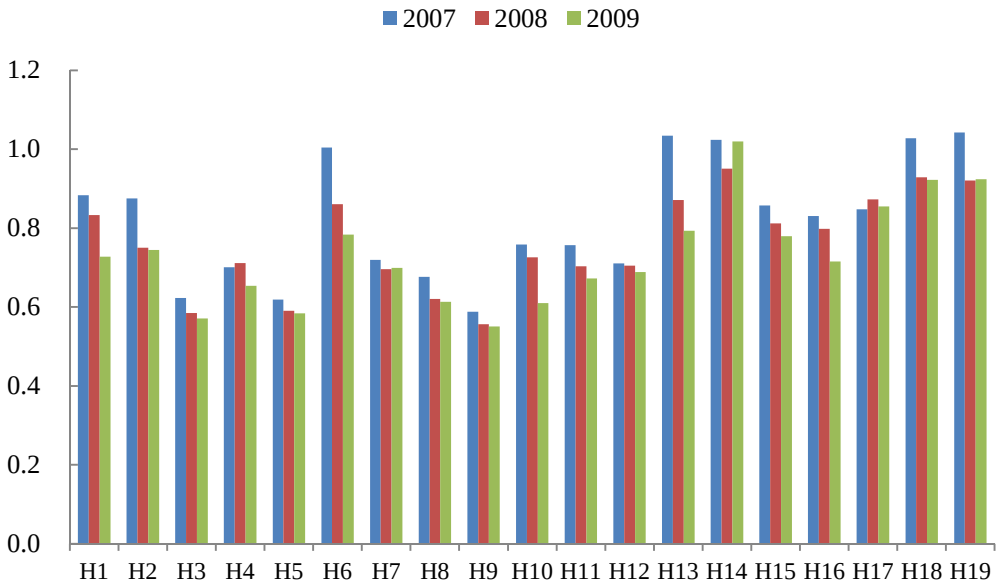

Figure 2: Panel data results

2.4 Preliminary results: Panel

We merged the dataset and evaluated the efficiency scores relative to 57 ($=19\times3$) DMUs as exhibited in Figure 2.Comparing the averages of these three years, we found that the average 0.820 of year 2007 is better than 2008 (0.763) and 2009 (0.732). We checked the non-parametric Wilcoxon rank-sum test. The results

indicate that the null hypothesis 2007 and 2008 have the same distribution of efficiency scores is rejected at the significance level 1%. 2007 outperforms 2008. Similarly, 2007 outperforms 2009. However, we cannot see significant difference between 2008 and 2009.

3 Use of historical data for estimating data variations: Model 1

In this section, we make use of historical data for resampling purposes.

3.1 Historical data and weights

Let the historical set of input and output matrix be $(\mathbf{X}^t, \mathbf{Y}^t)$ $(t = 1, \ldots, T)$ where $t=1$ is the first period and $t=T$ is the last period with $\mathbf{X}^t = (\mathbf{x}_1^t, \cdots, \mathbf{x}_n^t)$ and $\mathbf{Y}^t = (\mathbf{y}_1^t, \cdots, \mathbf{y}_n^t)$. The number of the DMU is n and, $\mathbf{x}_j^t \in R^m$ and $\mathbf{y}_j^t \in R^s$ are respectively input and output vectors of DMU_j.

a Super-efficiency scores of $(\mathbf{X}^T, \mathbf{Y}^T)$

First we evaluate the super-efficiency scores of the last period's DMUs. Then we gauge their confidence interval using replicas from $(\mathbf{X}^t, \mathbf{Y}^t)$ $(t = 1, \ldots, T)$ as follows.

b Lucas weight

We set the weight w_t to period t and assume the weights are increasing in t. For this purpose, the following Lucas number series $(l_1, \ldots, l_T)$ (a variant of Fibonacci series) is a candidate where we have

$$l_{t+2} = l_t + l_{t+1} \ (t = 1, \ldots, T-2; l_1 = 1, l_2 = 2). \quad (1)$$

Let the sum be $L = \sum_{t=1} l_t$ and we define weight w_t by

$$w_t = l_t / L \quad (t = 1, \ldots, T). \quad (2)$$

If T=5, we have w_1 = 0.0526, w_2 = 0.1053, w_3 = 0.1579, w_4 = 0.2631, w_5 = 0.4211. Thus, the influence of the past period fades away gradually.

3.2 Cumulative weight and random sampling

We regard the historical data $(\mathbf{X}^t, \mathbf{Y}^t)$ $(t = 1, \ldots, T)$ as discrete events with probability w_t and with the cumulative probability

$$W_t = \sum_{i=1}^{t} w_i \quad (t = 1, \ldots, T). \quad (3)$$

[Data Generation Process]

Using a uniform random number r $(0 \le r \le 1)$, we resample $(\mathbf{X}^t, \mathbf{Y}^t)$ if $W_{t-1} < r \le W_t$, where we define $W_0 = 0$. We evaluate the efficiency score of each DMU by using the super-SBM model. We repeat this process for the designated times.

3.3 Use of Fisher's z transformation

We compute the correlation coefficient of two inputs (outputs, input vs. output) items of the last period data over the all DMUs. Then, we calculate its ζ % confidence interval, e.g., 95%, using Fisher's z transformation [10]. If the corresponding correlation of a resampled data is out of range of this interval, we discard this resample data. We execute the same process between all pair of inputs, outputs and input vs. output. Thus, inappropriate samples with unbalanced inputs and outputs relative to the inputs and outputs of the last period are excluded from resampling. The above noted 95% confidence interval is not compulsory. The narrower the interval, the closer the resample will be to the last period data.

3.4 An example of historical data and resampling

We applied the above procedure to the historical data of nineteen hospitals for the two years 2008-2009 in Table 1. We excluded the year 2007 data, because they belong to a different population than 2009 as explained in Preliminary results (Panel) in section 2.4.

Table 5 shows the correlation matrix of the observed 2009 year data in Table 1 and Fisher 95% confidence intervals are exhibited in Table 6.

Table 5: Correlation matrix

	Doc	Nurse	Inpatient	Outpatient
Doc	1	0.7453	0.7372	0.5178
Nurse	0.7453	1	0.8610	0.7387
Inpatient	0.7372	0.8610	1	0.8264
Outpatient	0.5178	0.7387	0.8264	1

Table 6: Fisher 95% confidence lower/upper bounds for correlation matrix

		Lower bounds			
		Doc	Nurse	Inpatient	Outpatient
Upper bounds	Doc		0.4400	0.4255	0.0832
	Nurse	0.8961		0.6681	0.4281
	Inpatient	0.8926	0.9455		0.5959
	Outpatient	0.7869	0.8932	0.9311	

Table 7: Fisher 20% confidence lower/upper bounds for correlation matrix

		Lower bounds			
		Doc	Nurse	Inpatient	Outpatient
Upper bounds	Doc		0.71578	0.70695	0.46998
	Nurse	0.77214		0.8437	0.70854
	Inpatient	0.76482	0.87652		0.80525
	Outpatient	0.56266	0.76614	0.84547	

For example, the correlation coefficient between Doc and Outpatient is 0.5178 and its 95% lower/upper bounds are respectively 0.0832 and 0.7869.

In addition, we report Fisher 20% confidence lower/upper bounds in Table 7. The intervals are considerably narrowed down compared with Fisher 95% case.

Table 8 exhibits results obtained by 500 replicas where the column DEA is the last period's (2009) efficiency score and Average indicates the average score over 500 replicas. The column Rank is the ranking of average scores. We applied Fisher 95% threshold and found no out-of-range samples.

Figure 3 shows the 95% confidence intervals for the last period's (2009) DEA scores along with Average scores. The average of the 95% confidence interval for all hospitals is 0.10.

Table 8: DEA score and confidence interval with 500 replicas

	97.50%	DEA (2009)	Average	2.50%	Rank (Avg)
HH1	0.9228	0.7540	0.8047	0.7240	8
H2	0.8279	0.7787	0.7865	0.7415	9
H3	0.6285	0.5918	0.5999	0.5730	18
H4	0.7574	0.6802	0.7090	0.6694	14
H5	0.6375	0.6042	0.6088	0.5792	17
H6	0.9384	0.8475	0.8758	0.8159	6
H7	0.7620	0.725	0.7284	0.6998	11
H8	0.6902	0.6311	0.6365	0.6002	16
H9	0.6030	0.5681	0.5732	0.5452	19
H10	0.7963	0.6308	0.6818	0.6032	15
H11	0.7433	0.6985	0.7116	0.6808	13
H12	0.7684	0.7140	0.7237	0.6849	12
H13	1.0465	0.831	0.8978	0.8081	5
H14	1.1564	1.1448	1.1329	1.1037	1
H15	0.8692	0.8107	0.8277	0.7886	7
H16	0.8792	0.7418	0.7782	0.7140	10
H17	1.0142	0.9368	0.9542	0.9076	4
H18	1.0837	1.0745	1.0708	1.0497	3
H19	1.1194	1.0996	1.0897	1.0618	2

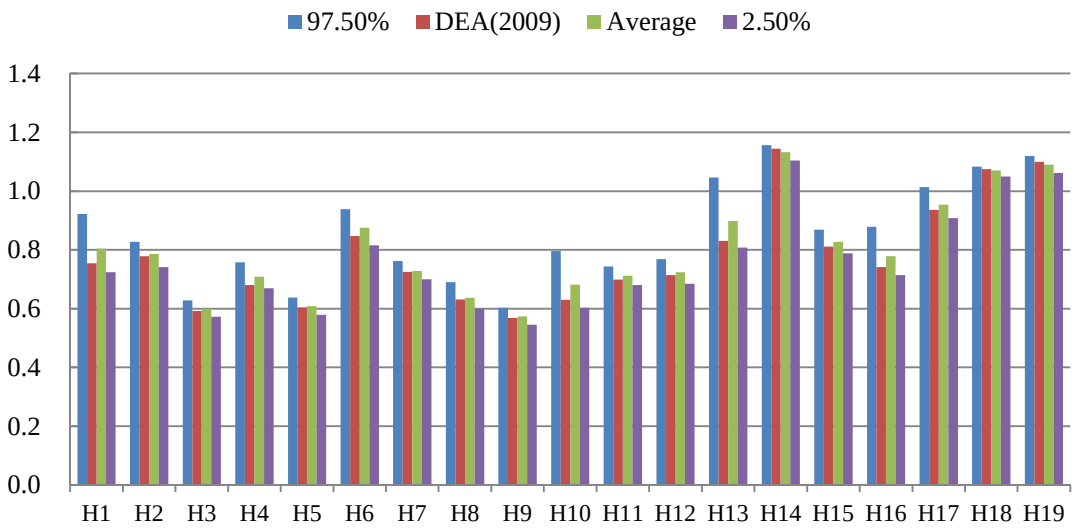

Figure 3: 95% confidence interval

3.5 Observations

3.5.1 Historical data

As pointed in section 3.4, we excluded 2007 data from the past data in this case. Historical data may suffer from accidental or exceptional events, for example, oil shock, earthquake, financial crisis, environmental system change and so forth. We must exclude these from the data. If some data are under age depreciation, we must adjust them properly.

3.5.2 Lucas weight

In this study, we put Lucas weights for past and present data. However, we can use other weight, e.g. exponential, as well.

3.5.3 Fisher's threshold

In the Fisher 95% (ζ95) case, we found no discarded samples, whereas in the Fisher 20% (ζ20) case, 1945 samples are discarded before getting 500 replicas. However, we cannot see significant differences between scores calculated by both thresholds.

3.5.4 Number of replicas

One resample produces one efficiency score for each DMU. We compared 500 and 5000 replicas. The difference was negligible small. 500 replicas may be acceptable in this case. However, number of replicas depends on the numbers of inputs, outputs and DMUs. Hence, we need to check the variations of scores by increasing the number of replicas.

4 Resampling with future forecasts: Model 2

In the previous section, we utilized historical data $(\mathbf{X}^t, \mathbf{Y}^t)$ $(t = 1,\dots,T)$ to gauge the confidence interval of the last period's scores. In this section, we forecast "future" $(\mathbf{X}^{T+1}, \mathbf{Y}^{T+1})$ by using "past-present" data $(\mathbf{X}^t, \mathbf{Y}^t)$ $(t = 1,\dots,T)$ and forecast the efficiency scores of the future DMUs with their confidence intervals.

4.1 Forecasting future data

Let $h^t(t = 1,\dots,T)$ be the observed historical data for a certain input/output of a DMU. We wish to forecast h^{T+1} from $h^t(t = 1,\dots,T)$. There are several forecasting engines available for this purpose. We must choose one or try several for deciding which

one is best suited for the problem at hand. As candidates, we choose the following three scenarios:

(a) Trend analysis: a simple linear least square regression,

(b) Weighted average: weight by Lucas number,

(c) Average of trend and weighted average.

By applying a forecasting model, we obtain the data set $(\mathbf{X}^{T+1}, \mathbf{Y}^{T+1})$. We evaluate the super-efficiency of the "future" DMU $(\mathbf{X}^{T+1}, \mathbf{Y}^{T+1})$.

4.2 Resampling by using past-present-future data

We have the past-present-future inter-temporal data set $(\mathbf{X}^t, \mathbf{Y}^t)$ $(t = 1, \ldots, T+1)$. Thus, we can apply the resampling scheme in the previous section and obtain confidence intervals.

4.3 An example of past-present-future DEA

In this section, we apply our scheme for the dataset displayed in Table 1. In this case we regard 2007-2008 as the past-present and 2009 as the future.

4.3.1 Forecast by Trend case

Table 11 reports the forecast 2009 data by Trend.

Table 12 shows the forecast DEA score and confidence interval along with the actual super-SBM score for 2009. Figure 4 exhibits 97.5% percent, 2.5% percent, forecast score and actual score.

Table 11: Forecast 2009 data: forecast by Trend

DMU	(I)Doc	(I)Nurse	(O)In-patient	(O)Out-patient
H1	120	473	628	1249
H2	141	550	634	1257
H3	124	633	553	1030
H4	138	521	709	1178
H5	146	619	635	1099
H6	113	615	686	1423
H7	112	490	562	1069
H8	130	519	512	1015
H9	129	516	491	903
H10	110	576	549	1288
H11	110	499	512	1075
H12	146	699	675	1677
H13	109	753	605	1755
H14	119	538	621	1920
H15	161	809	848	2072
H16	59	364	372	785
H17	113	627	687	1874
H18	66	382	454	901
H19	92	305	458	844

Table 12: Forecast DEA score, actual (2009) score and confidence interval: Forecast by Trend

DMU	(I)Doc	(I)Nurse	(O)In-patient	(O)Out-patient
H1	1.0237	0.9338	0.7540	0.8245
H2	1.0027	0.7870	0.7787	0.7220
H3	0.6649	0.6148	0.5918	0.5641
H4	0.8816	0.8581	0.6802	0.7319
H5	0.6814	0.6421	0.6042	0.5771
H6	1.0213	0.8768	0.8475	0.8062
H7	0.8292	0.7586	0.7250	0.6945
H8	0.7641	0.6725	0.6311	0.6066
H9	0.6983	0.6213	0.5681	0.5390
H10	0.8422	0.7781	0.6308	0.7111
H11	0.8425	0.7206	0.6985	0.6679
H12	0.8136	0.7716	0.7140	0.7068
H13	1.0814	1	0.8310	0.8276
H14	1.1575	1.0909	1.1448	1.0281
H15	0.9467	0.8541	0.8107	0.7902
H16	1.0376	0.9444	0.7418	0.7258
H17	1.0387	1.0348	0.9368	0.8982
H18	1.0899	1.0537	1.0745	0.9692
H19	1.1354	1.0594	1.0996	1.0113

It is observed that, of the nineteen hospitals, the actual 2009 scores of sixteen are included in the 95% confidence interval. The average of Forecast-Actual over the nineteen hospitals was 0.063 (6.3%).

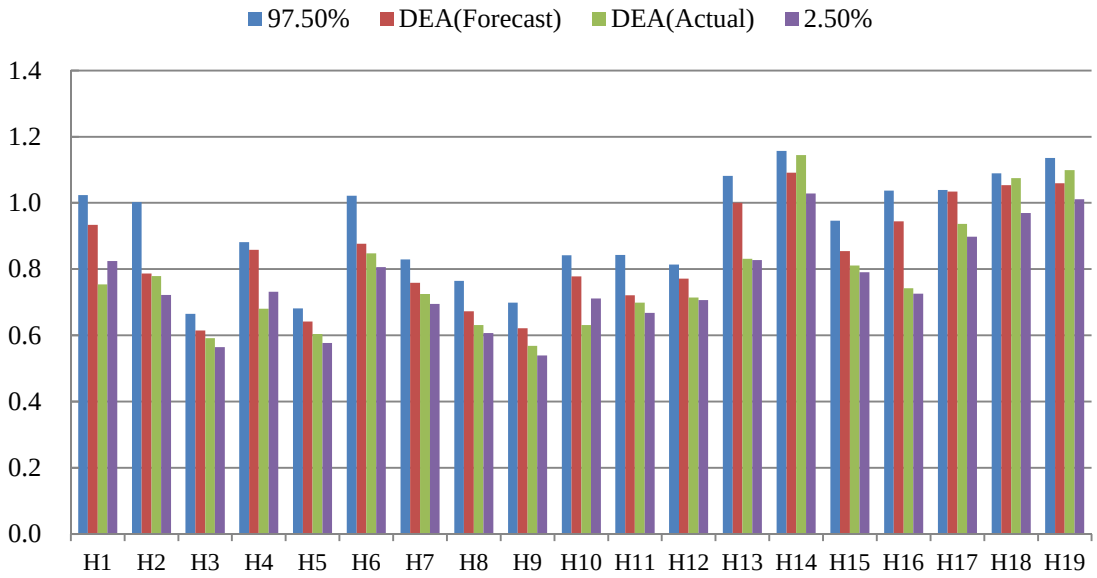

Figure 4: Confidence interval, forecast score and actual 2009 score: Forecast by Trend

4.3.2 Forecast by Lucas case

Table 13 reports forecast 2009 data by Lucas weight and Table 14 shows forecast 2009 scores, confidence intervals.

Table 13: Forecast 2009 data: Forecast by Lucas case

DMU	(I)Doc	(I)Nurse	(O)In-patient	(O)Out-patient
H1	112	446	613	1242
H2	130	482	639	1328
H3	120	589	574	1058
H4	138	534	702	1199
H5	141	615	647	1163
H6	104	584	706	1496
H7	100	495	547	1066
H8	114	480	501	1039
H9	116	472	486	868
H10	105	552	570	1275
H11	98	496	497	1170
H12	147	714	739	1650
H13	105	646	654	1926
H14	107	513	616	1885
H15	160	798	925	2173
H16	71	357	397	960
H17	112	651	707	1741
H18	63	386	471	909
H19	96	317	491	1076

Table 14: Forecast DEA score and confidence interval: Forecast by Lucas case

	97.50%	Forecast (2009)	Actual (2009)	2.50%
H1	1.0001	0.8974	0.7540	0.8469
H2	0.9329	0.8527	0.7787	0.7970
H3	0.6448	0.6218	0.5918	0.5987
H4	0.7855	0.7618	0.6802	0.7303
H5	0.6584	0.6400	0.6042	0.6200
H6	1.0101	0.9604	0.8475	0.9123
H7	0.7813	0.7347	0.7250	0.7006
H8	0.7201	0.6867	0.6311	0.6596
H9	0.6578	0.6177	0.5681	0.5894
H10	0.8109	0.7829	0.6308	0.7441
H11	0.8101	0.7573	0.6985	0.7171
H12	0.7623	0.7336	0.7140	0.712
H13	1.0590	1.0286	0.8310	1.0000
H14	1.1306	1.0868	1.1448	1.0409
H15	0.9120	0.8665	0.8107	0.8263
H16	0.9296	0.8488	0.7418	0.7869
H17	0.9731	0.9427	0.9368	0.8984
H18	1.0686	1.0443	1.0745	1.0115
H19	1.1075	1.0769	1.0996	1.0417

In this case, only four hospitals are included in the 95% confidence interval. The average of Forecast-Actual over the nineteen hospitals was 0.056 (5.6%).

4.3.3 Comparisons

Although we did not report the results by the Average of Trend and Lucas case, the results are similar to the Lucas case. We compare the number of fails for the three forecast models that actual score is out of 97.5% and 2.5 % interval. We have results as exhibited in Table 15. "Trend" gives the best performance among the three in this example.

Table 15: Number of fails

	Trend	Lucas	Average of Trend and Lucas
No. of fails	3	15	15

5 Conclusion

DEA scores are subject to change by data variations. This subject should be discussed from the perspective of the itemized input/output variations. From this point of view, we have proposed two models. The first model utilizes historical data for the data generation process, and hence this model resamples data from a discrete distribution. It is expected that, if the historical data are volatile widely, confidence intervals will prove to be very wide, even when the Lucas weights are decreasing depending on the past-present periods. In such cases, application of the moving-average method is recommended. Rolling simulations will be useful for deciding choice of the length of historical span.

The second model aims to forecast the future efficiency and its confidence interval. For forecasting, we proposed three scenarios; the trend, the weighted average and their average. On this subject, Xu and Ouenniche [16] will be useful for the selection of forecasting models, and Chang et al. [4] will provide useful information on the estimation of the pessimistic and optimistic probabilities of the forecast future input/output values.

References

[1] Banker R. Maximum likelihood, consistency and data envelopment analysis: A statistical foundation. Management Science 1993; 39, 1265-1273.

[2] Banker R, Natarajan R. Statistical test based on DEA efficiency scores. Chapter 11 in Cooper, Seiford, Zhu, eds., Handbook on Data Envelopment Analysis, Springer; 2004.

[3] Barnum DT, Gleason JM, Karlaftis MG, Schumock GT, Shields KL, Tandon S., Walton SM. Estimating DEA confidence intervals with statistical panel data analysis. Journal of Applied Statistics 2011; DOI:10.1080/02664763.2011.620948.

[4] Chang TS, Tone K, Wu CH. Past-present-future intertemporal DEA models. Journal of the Operational Research Society 2014; 214, 73-98.

[5] Charnes A, Cooper WW, Rhodes E. Measuring the efficiency of decision making units. European Journal of Operational Research 1978; 2, 429-444.

[6] Charnes A and Neralić L. Sensitivity analysis of the additive model in data envelopment analysis. European Journal of Operational Research 1990: 48, 332–341.

[7] Cook WD, Tone K, Zhu J. Data envelopment analysis: Prior to choosing a model. Omega 2014, 44, 1-4.

[8] Cooper WW, Seiford LM, Tone K. Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software. Second Edition. Springer; 2007.

[9] Efron B. Bootstrap methods: another look at the jackknife. Annals of Statistics 1979; 7, 1-26

[10] Fisher RA. Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population, Biometrika 1915; 10, 4, 507-521.

[11] Neralić L. Sensitivity analysis in models of data envelopment analysis. Mathematical Communications 1998; 3, 41-59.

[12] Simar L, Wilson P. Sensitivity of efficiency scores: How to bootstrap in nonparametric frontier models. Management Science 1998; 44, 1, 49-61.

[13] Simar L, Wilson P. Statistical inference in nonparametric frontier models: the state of the art. Journal of Productivity Analysis 2000; 13, 49-78.

[14] Tone K. A slacks-based measure of efficiency in data envelopment analysis. European Journal of Operational Research 2001; 130, 498-509.

[15] Tone K. A slacks-based measure of super-efficiency in data envelopment analysis. European Journal of Operational Research 2002; 143, 32-41.

[16] Xu B, Ouenniche J. A data envelopment analysis-based framework for the relative performance evaluation of competing crude oil prices' volatility forecasting model. Energy Economics 2012; 34, 576-583.

[17] Yang M, Wan G, Zheng E. A predictive DEA model for outlier detection. Journal of Management Analytics 2014; 1:1, 20-41, DOI: 10.1080/23270012.2014.889911.

[18] Zhu J. Super-efficiency and DEA sensitivity analysis. European Journal of Operational Research. 2001; 129, 443-455.