GRIPS Discussion Paper 13-23

Resampling in DEA

Kaoru Tone

December 2013

GRIPS
NATIONAL GRADUATE INSTITUTE
FOR POLICY STUDIES

National Graduate Institute for Policy Studies
7-22-1 Roppongi, Minato-ku,
Tokyo, Japan 106-8677

Resampling in DEA

Kaoru Tone

National Graduate Institute for Policy Studies, Japan

tone@grips.ac.jp

2013-12-06

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 three resampling models. The first one assumes downside and upside measurement error rates for each input/output, which are common to all DMUs. We resample data following the triangular distribution that the downside and upside errors indicate around the observed data. The second 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 last one deals with future prospects. This model aims at forecasting the future efficiency score and its confidence interval for each DMU.

Keywords: Data error; resampling; triangular distribution; confidence interval; past-present-future intertemporal DEA

1. Introduction

The treatment of data variations by statistical methods has taken a variety of forms in DEA. Banker [1] and Banker and Natarasan [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. Simar and Wilson [7,8] turn to “bootstrap methods” which enable them to deal with the case of multiple inputs and outputs. In this manner, the sensitivity of the efficiency score obtained by the variable returns-to-scale model can be tested by repeatedly sampling from the original samples. A sampling distribution of the efficiency score is then obtained, from which confidence intervals may be derived and statistical tests of significance developed. Tziogkidis [10] points out that “Bootstrap DEA is a significant development of the past decade; however, some of its assumptions and properties are still quite unclear, which may lead to mistakes in implementation and hypothesis testing”, and proposes a hypothesis testing procedure.

However, as far as the author knows, most researches on this subject are modifications of Efron [6]. This ignores the characteristics of inputs and outputs, which differ from DMU to DMU. Originally, DEA stems from the individualization of DMUs which have different values in the specified input and output factors for each DMU. Mathematical programming approaches have succeeded in measuring efficiency within this individualization framework. Hence, even resampling should inherit this merit, which could not be explored and expected by using statistical methods as represented by bootstrapping. In DEA, the data may suffer from measurement errors. There are several researches on measurement errors. Gauss (1777-1852) was the first to demonstrate that the distribution of measurement errors follows the Gaussian distribution (the normal distribution). In recent years, an OR method PERT (program evaluation and review technique) utilizes three-point estimates; pessimistic, most likely and optimistic, for each activity in the concerned project.

This paper deals with measurement errors in inputs and outputs and resamples data depending on the empirical distribution of errors.

This paper unfolds as follows. Section 2 assumes that measurement error rates for inputs and outputs are common to all DMUs and errors follow the triangular distribution. Section 3 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 ages. In section 4, we extend the approach presented in section 3 to future forecast data and resample future data depending on the past-present-future intertemporal 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 utilize the super-efficiency model and obtain the confidence intervals. Section 5 concludes the paper.

2. The common upside and downside measurement errors case

In this section, we introduce three estimations of input/output values and propose a resampling model based on the triangular distribution.

2.1 The triangular distribution

In this section, we assume that the data are bounded by upside and downside limits having a single mode. Like PERT, we denote the downside limit, the mode and the upside limit by a, m and b. The observed input and output values represent the mode m. We employ the triangular distribution for data as exhibited in Figure 1.

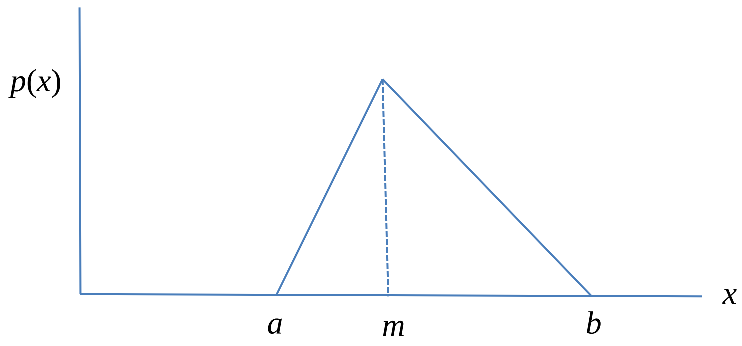

Figure 1: Triangular distribution

We assume that the three estimates differ in their input and output. The reason why we employ the triangular distribution is that, although the Beta distribution is the representative one among the distributions having unimodal and bounded characters, this distribution requires four parameters to be decided and it is difficult to determine them for the input and output data for each individual DMU. Hence, we utilize the triangular distribution represented by three parameters a, m and b. In this section, we assume that a and b can be expressed by the error rates α and β as follows:

$$\begin{aligned} a &= (1-\alpha)m \quad (0 \le \alpha \le 1) \\ b &= (1+\beta)m \quad (\beta \ge 0). \end{aligned} \tag{1}$$

The error rates, α and β, are decided externally, and differ in their input and output factors, but are common to all DMUs.

2.2 Data generation process

The triangular distribution has the following distribution function:

$$\begin{aligned} P(x) &= \frac{(x-a)^2}{(m-a)(b-a)} \quad (a \le x < m) \\ &= 1 - \frac{(b-x)^2}{(b-m)(b-a)} \quad (m \le x \le b). \end{aligned} \tag{2}$$

Hence, using a uniform random number r $(0 \le r \le 1)$, we can obtain an input/output value z as follows:

$$\begin{aligned} &\text{If } r \le \frac{m-a}{b-a}, \text{ then } z = a + \sqrt{r(m-a)(b-a)}, \\ &\text{If } r > \frac{m-a}{b-a}, \text{ then } z = b - \sqrt{(1-r)(b-m)(b-a)}. \end{aligned} \tag{3}$$

2.3 How to determine α and β

There are several methods for estimating α and β, among which we point to the following:

(1) Expert knowledge can be applied. For some instances, data are intentionally underestimated or overestimated, for example, in the accounting statements. From experience, experts in the concerned areas can estimate them.

(2) If historical data are available, we can apply the following procedure. Let the past T periods data for a certain input (output) i be $z_t (t = 1, \ldots, T)$ where T is the current (latest) period. Comparing with z_T, we evaluate downside (upside) error variation rates α_i^t (β_i^t) for the period t. From the distribution of $\{\alpha_i^t\}$ and $\{\beta_i^t\}$ for all DMUs, we can decide their median or average as α_i and β_i.

2.4 An example

Table 1 shows nineteen hospitals each having two inputs (doctor and nurse) and two outputs (inpatient and outpatient).

Table 1: Hospital data

	(I)Doctor	(I)Nurse	(O)Inpatient	(O)Outpatient
H1	116	545	603	1,295
H2	136	482	618	1,300
H3	125	616	561	1,071
H4	140	554	679	1,182
H5	137	633	622	1,147
H6	109	613	651	1,457
H7	101	491	540	1,067
H8	133	479	505	1,081
H9	121	501	486	904
H10	148	611	586	1,321
H11	102	501	479	1,113
H12	158	737	743	1,714
H13	120	697	634	1,872
H14	116	517	623	2,009
H15	166	817	877	2,155
H16	81	378	406	897
H17	112	663	709	1,733
H18	63	381	463	872
H19	95	320	490	1,034

Table 4 reports the estimated error rates in percentage for each input and output which are obtained using past records in the manner described in (2) of section 2.3. Actually, we estimated the error rates for inputs (outputs) using the 2008(Past)-2009(Current) data in Table 6 and decided α_i and β_i by their medians. For example, the 2008-2009 data of the input "Doctor" and its variation rates are exhibited in Table 2 while Table 3 shows downside, upside variations and their medians.

Table 2: Variation rate of Doctor from 2008 to 2009

	(I)Doc(2008)	(I)Doc(2009)	Variation rate
H1	**114**	**116**	**-0.017**
H2	**133**	**136**	**-0.022**
H3	**121**	**125**	**-0.032**
H4	**138**	**140**	**-0.014**
H5	**142**	**137**	**0.036**
H6	**106**	**109**	**-0.028**
H7	**103**	**101**	**0.020**
H8	**118**	**133**	**-0.113**
H9	**119**	**121**	**-0.017**
H10	**106**	**148**	**-0.284**
H11	**101**	**102**	**-0.010**
H12	**147**	**158**	**-0.070**
H13	**106**	**120**	**-0.117**
H14	**110**	**116**	**-0.052**
H15	**160**	**166**	**-0.036**
H16	**68**	**81**	**-0.160**
H17	**112**	**112**	**0.000**
H18	**64**	**63**	**0.016**
H19	**95**	**95**	**0.000**

Table 3: Downside, upside variations and median for Doctor

	Downside		Upside
H1	0.017	**H5**	0.036
H2	0.022	**H7**	0.020
H3	0.032	**H18**	0.016
H4	0.014	Median(β)	0.020
H6	0.028		
H8	0.113		
H9	0.017		
H10	0.284		
H11	0.010		
H12	0.070		

H13	0.117
H14	0.052
H15	0.036
H16	0.160
Median(α)	0.034

Table 4: The downside and upside error rates in percentage

	Downside error rate α	Upside error rate β
Doctor	3.4%	2%
Nurse	3.4%	2.8%
Inpatient	1.4%	2.7%
Outpatient	3.2%	1.8%

We resampled the data using the data generation process described in section 2.2 and evaluated the efficiency of hospitals by the input-oriented super SBM model under the constant-returns-to-scale condition (Tone [9], Cooper et al. [5]). We repeated this process 500 times. Table 5 shows the results, where the column DEA is the super-efficiency score of the original data in Table 1.

Table 5: Results of 500 replicas

DMU	DEA	97.5%	75%	50%	25%	2.5%
H1	0.8353	0.8603	0.8442	0.8351	0.8263	0.8148
H2	0.8568	0.8775	0.8624	0.8535	0.8459	0.8296
H3	0.6823	0.7047	0.6896	0.6821	0.6758	0.6619
H4	0.8357	0.8699	0.8477	0.8364	0.8239	0.8051
H5	0.7132	0.7374	0.723	0.7133	0.7043	0.6885
H6	0.8839	0.9106	0.8916	0.8824	0.874	0.8595
H7	0.8258	0.8488	0.834	0.8256	0.8179	0.8015
H8	0.7111	0.727	0.7152	0.7087	0.7019	0.6896
H9	0.6724	0.6971	0.6806	0.6726	0.6646	0.6507

H10	0.6924	0.7104	0.6982	0.6915	0.686	0.6754
H11	0.7526	0.7749	0.7591	0.752	0.745	0.7339
H12	0.7731	0.7946	0.7795	0.7724	0.7664	0.7521
H13	0.8418	0.866	0.8508	0.8429	0.8336	0.8171
H14	1.2785	1.3212	1.293	1.2762	1.2603	1.2308
H15	0.8593	0.8841	0.8675	0.8589	0.8502	0.8365
H16	0.8144	0.8371	0.822	0.8139	0.8062	0.7941
H17	0.9386	1.0305	1.0169	0.9412	0.9303	0.9149
H18	1.133	1.1631	1.1434	1.1328	1.1229	1.1052
H19	1.1089	1.1303	1.1164	1.1088	1.1003	1.0862

Figure 2 exhibits the 95% confidence interval and the original DEA score. The average of the confidence interval for all hospitals is 0.05 which is small in this case. Relatively small downside and upside error rates in Table 4 result in this number.

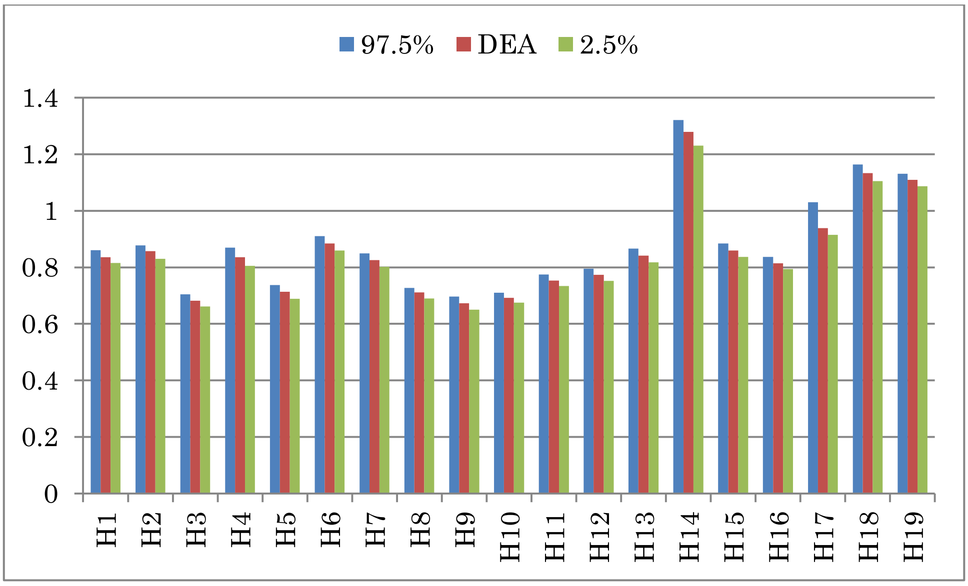

Figure 2: 95% confidence interval and DEA score

3. Use of historical data for estimating data variations

In the previous section, we applied the triangular distribution for simulating measurement errors. 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

$$l_{t+2} = l_t + l_{t+1}\ (t = 1, \ldots, T-2; l_1 = 1, l_2 = 2). \qquad (4)$$

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

$$w_t = l_t / L \quad (t = 1, \ldots, T). \qquad (5)$$

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

$$W_t = \sum_{i=1}^{t} w_i \quad (t = 1, \ldots, T). \tag{6}$$

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 An example of historical data and resampling

Table 6 displays the historical data of nineteen hospitals for the three years 2007-2009.

Table 6: Historical data

	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
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 7 exhibits results obtained by 500 replicas where the column DEA is the last period's (2009) efficiency score.

Table 7: DEA score and confidence interval with 500 replicas

DMU	DEA(2009)	97.5%	75%	50%	25%	2.5%
H1	0.8353	0.9939	0.9509	0.9194	0.8396	0.8159
H2	0.8568	0.9324	0.8904	0.8706	0.8514	0.8245
H3	0.6823	0.7399	0.7165	0.7012	0.6893	0.6702
H4	0.8357	0.9107	0.8801	0.8663	0.8439	0.817
H5	0.7132	0.7654	0.7411	0.7255	0.7132	0.6916
H6	0.8839	1.0125	0.9526	0.9257	0.8972	0.8698
H7	0.8258	0.8586	0.8342	0.8247	0.8145	0.7975
H8	0.7111	0.7797	0.7419	0.7169	0.6974	0.6785
H9	0.6724	0.7406	0.7009	0.6856	0.6731	0.6533
H10	0.6924	0.8619	0.82	0.768	0.6977	0.6617
H11	0.7526	0.8213	0.7863	0.7723	0.7577	0.7356
H12	0.7731	0.8205	0.7965	0.7814	0.7657	0.7441
H13	0.8418	1.1023	1.067	0.9309	0.8591	0.8048
H14	1.2785	1.3282	1.2332	1.1915	1.1426	1.0728
H15	0.8593	0.947	0.8987	0.8807	0.8652	0.8382
H16	0.8144	0.9344	0.8819	0.8423	0.8165	0.7861
H17	0.9386	1.0238	0.9543	0.9382	0.9237	0.9029
H18	1.133	1.1674	1.1283	1.1073	1.0796	1.0277
H19	1.1089	1.135	1.106	1.09	1.0722	1.0438

Figure 3 shows the 95% confidence intervals for the last period's (2009) DEA scores. The average of the 95% confidence interval for all hospitals is 0.13 which is larger than the average in Figure 2. This reflects large variations in the past data.

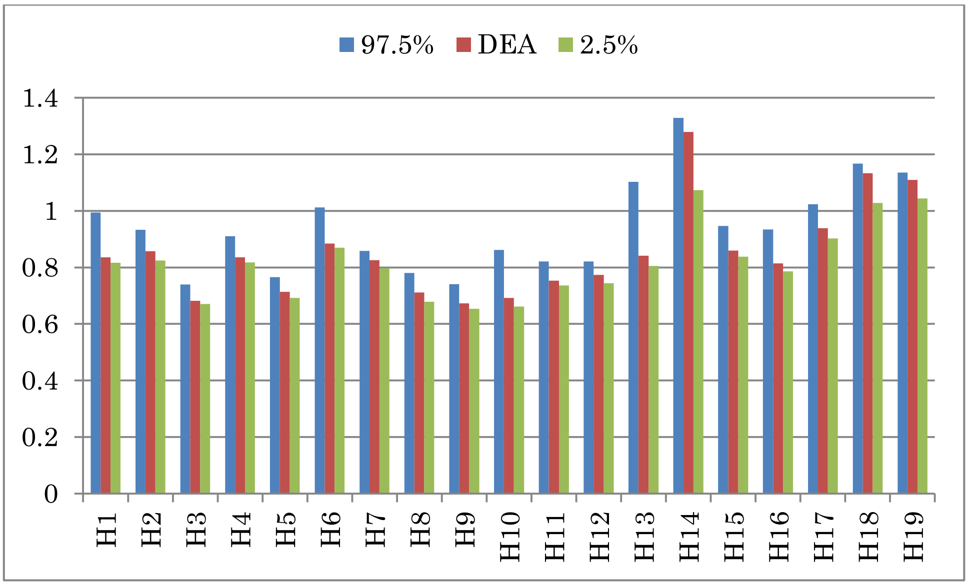

Figure 3: 95% confidence interval

3.4 Remarks on historical data

Historical data may suffer from accidental or exceptional events, for example, oil shock, earthquake, financial crisis and so forth. We must exclude these from the data. If some data are under age depreciation, we must adjust them properly.

4. Resampling with future forecasts

In the previous section, we utilized historical data $(\mathbf{X}^t, \mathbf{Y}^t)\,(t = 1, \ldots, T)$ 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 $(\mathbf{X}^t, \mathbf{Y}^t)\,(t = 1, \ldots, T)$ and evaluate the efficiency scores of the future DMUs with their confidence intervals.

4.1 Forecasting and efficiency score of the forecast DMUs

Let $z^t (t = 1, \ldots, T)$ be the observed historical data for a certain input/output of a DMU. We wish to forecast z^{T+1} from $z^t (t = 1, \ldots, T)$. 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 intertemporal 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 6. In this case we regard 2007-2008 as the past-present and 2009 as the future.

4.3.1 Forecast by trend case

Table 8 reports the forecast 2009 data by trend.

Table 8: Forecast 2009 data: forecast by trend

DMU	(I)Doc	(I)Nurse	(O)Inpatient	(O)Outpatient
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 9 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 9: Forecast DEA score and confidence interval: Forecast by trend

DMU	Forecast	97.50%	75%	50%	25%	2.50%	Actual
H1	0.9633	1.0355	0.9959	0.9641	0.9359	0.8898	0.835319
H2	0.8118	1.0102	0.882	0.8412	0.812	0.7816	0.856758
H3	0.6917	0.7566	0.7282	0.7075	0.6879	0.6618	0.682268
H4	0.96	0.977	0.9459	0.926	0.9057	0.8695	0.835736
H5	0.7477	0.774	0.7555	0.7424	0.728	0.7027	0.713184
H6	0.9223	1.028	0.974	0.9439	0.9263	0.8921	0.883851
H7	0.8473	0.902	0.851	0.8303	0.8101	0.7773	0.825768
H8	0.7013	0.7896	0.741	0.7148	0.692	0.6589	0.71106
H9	0.677	0.759	0.709	0.6876	0.665	0.6354	0.672391
H10	0.8035	0.8609	0.8246	0.8104	0.794	0.7657	0.692355
H11	0.7825	0.8516	0.8079	0.7818	0.7629	0.7348	0.752595
H12	0.7952	0.8537	0.8089	0.7824	0.7646	0.7381	0.773051
H13	0.8501	1.1226	1.0809	1.0412	0.8865	0.8055	0.841793
H14	1.0987	1.2584	1.1706	1.1354	1.1036	1.0361	1.278508
H15	0.878	0.9647	0.9145	0.8876	0.8638	0.8343	0.859348
H16	0.9046	1.0407	0.9417	0.9008	0.8613	0.7904	0.814377
H17	1.0348	1.0387	1.0147	0.9777	0.9457	0.9093	0.938581
H18	1.0963	1.1381	1.0938	1.072	1.05	0.9931	1.132974

H19	1.0779	1.152	1.1126	1.0877	1.0672	1.0262	1.108945

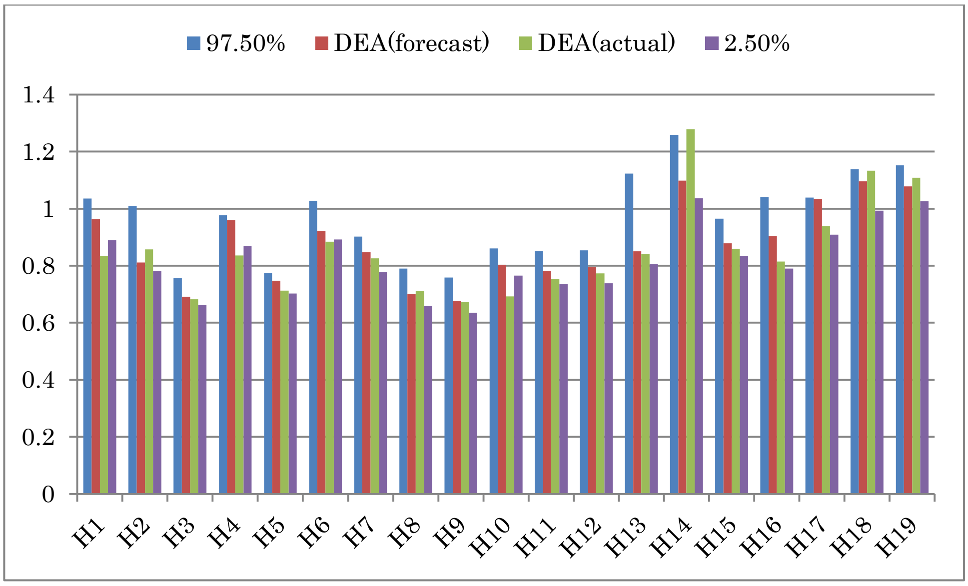

Figure 4: Confidence interval, forecast score and actual 2009 score: Forecast by trend

It is observed that, of the nineteen hospitals, the actual 2009 scores of fourteen are included in the 95% confidence interval. The average of $|\text{Actual} - \text{Forecast}| / \text{Actual}$ over the nineteen hospitals was 0.062 (6.2%).

4.3.2 Forecast by Lucas weighted average case

Table 10 reports forecast 2009 data by Lucas weight and Table 11 shows forecast 2009 scores, confidence intervals.

Table 10: Forecast 2009 data: Forecast by Lucas weight

DMU	(I)Doc	(I)Nurse	(O)Inpatient	(O)Outpatient
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 11: Forecast DEA score and confidence interval: Forecast by Lucas weight

DMU	Forecast	97.50%	75%	50%	25%	2.50%	Actual
H1	0.9556	1.0002	0.9703	0.9578	0.9448	0.9215	0.835319
H2	0.8903	0.952	0.9025	0.8887	0.8745	0.8549	0.856758
H3	0.7208	0.7453	0.7297	0.7205	0.7135	0.6968	0.682268
H4	0.8939	0.9127	0.9023	0.895	0.8867	0.8734	0.835736
H5	0.7398	0.7553	0.746	0.7412	0.7357	0.7214	0.713184
H6	0.9763	1.0119	0.989	0.9758	0.9654	0.9476	0.883851
H7	0.8222	0.8519	0.8301	0.8223	0.8141	0.7991	0.825768
H8	0.7348	0.7657	0.7445	0.7336	0.7238	0.7108	0.71106
H9	0.7034	0.7366	0.712	0.702	0.6935	0.6773	0.672391
H10	0.8201	0.8441	0.8261	0.8182	0.8115	0.7999	0.692355
H11	0.7912	0.824	0.7992	0.7895	0.7817	0.7669	0.752595
H12	0.7895	0.8188	0.7954	0.787	0.7796	0.7672	0.773051
H13	1.0628	1.0887	1.0715	1.0618	1.0522	0.9004	0.841793

H14	1.1123	1.1726	1.1379	1.1177	1.0906	1.0434	1.278508
H15	0.9093	0.9448	0.9152	0.9055	0.8966	0.8843	0.859348
H16	0.8819	0.9297	0.8999	0.8842	0.8683	0.8304	0.814377
H17	0.9323	0.9565	0.9413	0.9322	0.9232	0.907	0.938581
H18	1.0517	1.0903	1.064	1.0543	1.0438	1.0178	1.132974
H19	1.0875	1.1277	1.1021	1.0891	1.0773	1.045	1.108945

In this case, only four hospitals are included in the 95% confidence interval. The average of $\left|\text{Actual} - \text{Forecast}\right| / \text{Actual}$ over the nineteen hospitals was 0.075 (7.5%).

4.3.3 Comparisons

We compare the correlation coefficients between the forecast 2009 scores and the actual 2009 scores. We have results as exhibited in Table 12. "Trend" gives a better correlation than "Lucas weight" in this case. Although we did not report the average of the trend and Lucas weight case in detail, this case gives the worst correlation.

Table 12: Correlations between forecasted and actual scores

	Trend	Lucas weight	Average of Trend and Lucas
Correlation	0.900421	0.874473	0.868957

5. Conclusion

DEA is a non-parametric mathematical programming model that deals directly with input/output data. Using the data, DEA can evaluate the relative efficiency of DMUs and propose a plan to improve the inputs/outputs of inefficient DMUs. This function is difficult to achieve with similar models in statistics, e.g., stochastic frontier analysis. This is a great contribution of Charnes and Cooper (Charnes et al. [4])).

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 three models. The first model assumes downside and upside error rates for each input and output which are common to all DMUs, and utilizes the triangular distribution for the data generation process of resampling. Other data generation processes may be possible. This is a future research subject. The second 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, confidence intervals will prove to be very wide, even when the Lucas weights are decreasing depending on age. The choice of the length of historical span is a future research subject. Monte Carlo simulation will be useful for deciding the span. The third 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 [11] will be useful for the selection of forecasting models, and Chang et al. [3] will provide useful information on the estimation of the pessimistic and optimistic probabilities of the forecast future input/output values.

We did not compare our resampling models with the bootstrapping models by Simar and Wilson, because the underlining concepts are different between the two models. However, comparative studies in theory and applications are interesting future subjects.

References

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

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

[3] Chang TS, Tone K, Wu CH. Past-present-future intertemporal DEA models. Journal of the Operational Research Society, forthcoming.

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

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

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

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

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

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

[10] Tziogkidis P. Bootstrap DEA and hypothesis testing. Cardiff Business School, Working paper No. E2012/18.

[11] 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.