



Estimating the confidence intervals for DEA efficiency scores of Asia-Pacific airlines

Thanh Ngo¹ · Kan Wai Hong Tsui¹

Received: 15 February 2021 / Revised: 20 May 2021 / Accepted: 28 July 2021 /
Published online: 7 August 2021

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

Data envelopment analysis (DEA) is a popular non-parametric approach to examine performance and productivity of airlines; however, it could not provide statistical information such as confidence intervals on the estimated efficiency scores. We combined stochastic frontier analysis and DEA into a single framework to disentangle noise and ‘pure’ inefficiency from the DEA inefficiency scores and accordingly provide confidence intervals for the estimated efficiency scores. Monte-Carlo simulation verified that our novel model is a good alternative for the conventional DEA as well as the bootstrap DEA. Empirical application using Asia-Pacific airlines’ data (2008–2015) shows that after accounting for the ‘pure’ random errors, the sampled Asia-Pacific airlines performed well during the study period but their ‘pure’ efficiency was declining, hence, there is still room for improvement.

Keywords Data envelopment analysis and stochastic · Confidence intervals · Monte Carlo simulation · Bootstrap DEA · Asia-Pacific airlines · Efficiency levels

1 Introduction

Prior to the Covid-19 pandemic, the global economy was significantly boosted up by the aviation sector, with 87.7 million jobs supported and US\$3.5 trillion generated (corresponding to 4.1% of the global gross domestic product in 2018) (Air Transport Action Group 2019). The direct and indirect impacts of the aviation industry have been recognised in business and trade development, infrastructure development and tourism, among others (Acar and Karabulak 2015). These developments, in turn, fuel the demand for air transportation, especially in the Asia-Pacific region

✉ Thanh Ngo
T.Ngo@massey.ac.nz

Kan Wai Hong Tsui
W.H.K.Tsui@massey.ac.nz

¹ The School of Aviation, Massey University, Palmerston North, New Zealand

(Heshmati and Kim 2016). A recent report from the IATA forecast revealed that the net profits of the Asia-Pacific's aviation industry reached US\$6.3 billion in 2016, second to North America. However, the report pointed out that airlines operating in the Asia-Pacific region performed differently, and thus an examination of the efficiency and performance of Asia-Pacific's airlines is justified (IATA 2016).

Researchers in the transportation field have witnessed the rise of the frontier analysis approach in evaluating the performance and efficiency of homogeneous decision-making units (DMUs) such as airports (Assaf et al. 2014; Ngo and Tsui 2020), airlines (Adler et al. 2013; Heshmati and Kim 2016), ports (Tovar and Wall 2017), railways (Lan and Lin 2005; Azadeh et al. 2018) and buses (Viton 1997; Kang et al. 2019). The idea of frontier analysis can be traced back to the possible production frontier, in which DMUs are compared with a benchmark (i.e. the best practice frontier): a DMU is efficient if it operates on the frontier and inefficient otherwise. To estimate the best practice frontier, one can use a parametric approach [e.g. stochastic frontier analysis (SFA)] or a non-parametric approach [e.g. data envelopment analysis (DEA)]. These approaches have their own pros and cons (see, for example, Coelli et al. 2005); for example, DEA is often used in cases of small samples, whereas SFA often requires larger samples.

An important difference between these two approaches is that in the non-parametric model, if a DMU is away from the best practice frontier, the deviation is treated as inefficiency; in the parametric approach, the deviation or the error is broken down into inefficiency and pure error (e.g. measurement errors or luck). The argument for DEA is that if a DMU deviates from the frontier, for whatever reason, it is inefficient. However, statisticians (of SFA) tend to argue that if the reason is exogenous, such as a flood in agriculture, it should be distinguished from inefficiency. Specifically, in SFA, the frontier is first estimated via the ordinary least squares (OLS) model, and the residuals between the observed and estimated data are decomposed via the maximum likelihood estimation (MLE) approach (Aigner et al. 1977; Meeusen and van den Broeck 1977; Bogetoft and Otto 2011; Ngo et al. 2019). These econometric techniques allow one to make statistical inferences about SFA efficiency estimates, such as the confidence intervals. In contrast, the non-parametric approach has been characterised as deterministic (as opposed to the econometric approach) and thus it is deemed to be a non-statistical approach (Gong and Sickles 1992; Banker 1996).

To remedy the non-statistical deficiency of the non-parametric approach of performance analysis, several approaches have been introduced (see, for example, Grosskopf 1996; Olesen and Petersen 2016). The recent review of Olesen and Petersen (2016) summarised these approaches as three directions. The first direction is to extend DEA to be able to handle the estimated deviations as random deviations (Banker 1988, 1996; Banker et al. 2015). The second direction is to extend DEA to be able to handle random errors in data based on the bootstrap technique (Simar and Wilson 1998; Lothgren and Tambour 1999). The last direction is to extend DEA to regard the estimated frontier as a random or stochastic frontier, also based on random variation in the data (Huang and Li 2001; Ruggiero 2004; Kao and Liu 2009; Azadeh et al. 2017). These three directions have their own limitations. The first does not include any estimation of the sampling distributions of the estimated efficiency

scores, implying that no information can be extracted on the confidence intervals for each of the estimated DEA efficiency scores. The second contributes to the issue by resampling the data to get an approximation of these unknown sampling distributions and extract confidence intervals for the estimated DEA efficiency scores (Olesen and Petersen 2016); however, these *pseudo*-bootstrapped data and efficiency scores provide limited implications for managers to improve the performance of the DMUs being studied. The last one departs from the deterministic frontier basis of DEA; from the managerial perspective, it is also not very convincing to argue that the observed inputs and outputs data of the involved DMUs are not 'real' but random or stochastic (Olesen and Petersen 2016).

In this paper, we extend the first direction of random deviations introduced by Banker (1988, 1996) and Banker et al. (2015) to provide confidence intervals for DEA efficiency scores. Specifically, in contrast with the second and third directions above, we argue that the observed data is unique, so does the efficient frontier enveloped from those data. However, estimation errors can happen in DEA, like in any regression/estimation model, resulting in randomness in the DEA inefficiency scores. To simply put, DEA inefficiency measurement is affected at some extent by errors and noise, and thus we need to disentangle them from the 'pure' inefficiency component. This is similar to the situation of SFA, as discussed above, so one can use the MLE approach to solve this problem and also extract important information on the 'pure' inefficiency scores, the 'pure' efficiency scores and their confidence intervals. This information is useful for managerial implications and our approach therefore overcomes the limitation of the first direction pointed out by Olesen and Petersen (2016) as discussed above.

Empirically, this study employed a two-stage DEA model where the frontier as well as the DEA inefficiency (or deviations) were estimated during the first-stage estimation via the conventional DEA model. These inefficiencies were then decomposed into 'pure' inefficiency and noise via the MLE approach. In this sense, our approach is a combination of DEA and SFA and therefore is named as the data envelopment analysis and stochastic (DEAS) model. This novel DEAS approach is similar to traditional SFA, except the DEAS deviations are computed between the observed data points and the (piece-wise) DEA frontier instead of the estimated OLS frontier, as in SFA. The validity and robustness of the proposed DEAS model in this study is further verified by a Monte Carlo simulation, in which the results show that the 'pure' inefficiency component of the novel DEAS model has similar characteristics to those of the DEA model, and the 'pure' random errors component of DEAS follows a normal distribution. Therefore, this novel DEAS approach is a good alternative to conventional DEA and bootstrap DEA for measuring the performance and efficiency of DMUs. In our view, it can be a compliment or robustness-checking tool for bootstrap DEA.

For an empirical application, this study applied the DEAS model to examine the operational and financial efficiency levels of 14 major Asia-Pacific airlines. The International Air Transport Association (IATA) has acknowledged that the Asia-Pacific aviation market in 2016 was the second largest aviation industry in the world, with a net profit of US\$6.3 billion, second only to North America's. IATA (2016) also pointed out that Asia-Pacific airlines performed differently, and

thus a thorough examination of the efficiency and performance of Asia-Pacific airlines via this novel DEAS model is justified.

The rest of the paper is constructed as follows. Section 2 reviews the methodologies of SFA and bootstrap DEA, and explains the proposed DEAS model. A Monte Carlo simulation for illustrative purposes is presented in Sect. 3. Section 4 applies the DEAS model to measure the operational performance of the sampled Asia-Pacific airlines and discusses the findings. Section 5 concludes the key findings and indicates directions for future research.

2 Confidence interval estimation in frontier analysis

As discussed previously, this section provides some information on SFA as well as bootstrap DEA, the most popular approach dealing with confidence interval estimation in DEA. We then discuss some limitations of bootstrap DEA in comparison to SFA and then introduce our DEAS model—note that relevant discussions on the so-called stochastic DEA (the first direction mentioned in Olesen and Petersen 2016) are also presented in this sub-section.

2.1 Using SFA to estimate the confidence intervals of efficiency scores

SFA was simultaneously introduced by Aigner et al. (1977) and Meeusen and van den Broeck (1977). The basic idea of SFA is that in the residuals ε_i between the observed and OLS-estimated data, one can distinguish the inefficiency u_i from the stochastic error v_i so that:

$$y_i = \alpha + \beta x_i + \varepsilon_i = \alpha + \beta x_i + v_i - u_i, \quad (1)$$

where y represents the output, x represents the set of inputs used to produce output y , β is a vector of coefficients to be estimated by using OLS, i denotes the observation or DMU, v is the ‘pure’ random error and u gets a negative sign, since it represents inefficiency. Note that in Eq. (1), the SFA frontier and the ‘pure’ inefficiency component u are still deterministic; only the error component v is stochastic.

One basic assumption for the two components u_i and v_i is that they are independent and identically distributed across observations [although there is an argument that u_i should not be identical (Olesen and Petersen 2016)], and thus u_i has a constant mean of μ_u and a variance of σ_u^2 , whereas v_i has a constant mean of 0 and a variance of σ_v^2 . The optimal values of the parameters α , β , σ_u and σ_v (and any other parameters) that can maximise the log-likelihood function in Eq. (2) also specify the SFA frontier, allowing us to estimate the efficiency and related confidence intervals. Note that it is under the assumption that u_i follows a half-normal distribution, whereas y_i follows a normal distribution (for other assumptions, please see Kumbhakar et al. (2015), among others).

$$L_i = -\ln\left(\frac{1}{2}\right) - \frac{1}{2} \ln(\sigma_u^2 + \sigma_v^2) + \ln\phi\left(\frac{\varepsilon_i}{\sqrt{\sigma_u^2 + \sigma_v^2}}\right) + \ln\Phi\left(\frac{\mu_{*i}}{\sigma_*}\right) \tag{2}$$

where:

$$\mu_{*i} = \frac{-\sigma_u^2 \varepsilon_i}{\sigma_u^2 + \sigma_v^2}, \tag{3}$$

$$\sigma_*^2 = \frac{\sigma_u^2 \sigma_v^2}{\sigma_u^2 + \sigma_v^2}. \tag{4}$$

Accordingly, the SFA efficiency scores can be estimated as in Eq. (5), following Jondrow et al. (1982):

$$SFA_i = E(u_i|\varepsilon_i) = \mu_{*i} + \sigma_* \left[\frac{\phi(\mu_{*i}/\sigma_*)}{\Phi(\mu_{*i}/\sigma_*)} \right]. \tag{5}$$

The confidence intervals of the efficiency scores in Eq. (5) are further estimated following Horrace and Schmidt (1996) and Hjalmarrsson et al. (1996), as in Eq. (6):

$$\exp\left(-\mu_{*i} - \sigma_* \Phi^{-1}\left\{\frac{\alpha}{2} \Phi\left(\frac{\mu_{*i}}{\sigma_*}\right)\right\}\right) \leq SFA_i \leq \exp\left(-\mu_{*i} - \sigma_* \Phi^{-1}\left\{\left(1 - \frac{\alpha}{2}\right) \Phi\left(\frac{\mu_{*i}}{\sigma_*}\right)\right\}\right) \tag{6}$$

2.2 Using bootstrap DEA to estimate the confidence intervals of efficiency scores

The bootstrap method is an intensive resampling method that can approximate the sampling distribution of the estimator and thus provides inferences in complex problems (Efron and Tibshirani 1994). Following this approach, Simar and Wilson (1998) and Lothgren and Tambour (1999) argued that by using bootstrap, one can resample the observed sample to be an approximation of the population. Simply put, the efficiency scores observed from DEA are only a sample of the population (of the efficiency scores) and therefore bootstrap DEA can provide better estimators of the efficiency scores as well as permitting valid inferences.

Conventional DEA examines a set of n firms (or DMUs), each using m inputs x_k ($k = 1, \dots, m$) to produce s outputs y_r ($r = 1, \dots, s$). In this sense, the efficiency score of DMU_j can be estimated as:

$$\begin{aligned}
 DEA_j &= \max_{u,v} \frac{\sum_r^s u_r y_{rj}}{\sum_k^m v_k x_{kj}} \\
 \text{subject to : } & \frac{\sum_r^s u_r y_{rj}}{\sum_k^m v_k x_{kj}} \leq 1, \quad j = 1, \dots, n; \\
 & u_r, v_k \geq 0, \quad r = 1, \dots, s, \quad k = 1, \dots, m,
 \end{aligned} \tag{7}$$

where u and v are the two sets of multipliers or weights to be estimated for the outputs y and the inputs x of DMU_j , respectively.

Equation (7) seeks for an optimal weights u and v so that DEA_j is maximised. Note that the first constraint of Eq. (7) requires those weights to be applicable to all DMUs in the sample so that the efficiency scores of all DMUs cannot exceed unity. In bootstrap DEA, Simar and Wilson (1998) proposed that both DEA_j and the first constraint can be computed by using *pseudo*- (or resampling) data. In contrast, Lothgren and Tambour (1999) argued that although the calculation of DEA_j can use *pseudo*-data, the first constraint should use the original data instead. Simar and Wilson (2000) have criticized that Lothgren and Tambour (1999)'s approach is incorrect; however, there are evidence that the latter one still has some empirical values (e.g. Tziogkidis 2012). Nevertheless, it is obvious that bootstrap DEA can estimate the confidence intervals for the DEA efficiency scores by using data resampling; however, the trade-off is that the bootstrapped DEA scores and their confidence intervals are estimated from the *pseudo*-efficiency scores (derived from *pseudo*-data) but not the original ones, and thus there is limited implications for managers to improve the performance of the firms involved (Olesen and Petersen 2016). It is therefore necessary to construct the confidence intervals based on the observed efficiency scores of the sampled DMUs so that after the 'pure' efficiency scores are estimated, the results can be used by managers.

2.3 Using DEAS to estimate the confidence intervals of efficiency scores

Following Banker (1988, 1996) and Banker et al. (2015), the proposed DEAS model in this study differs from other DEA extensions in that the frontier is not based on the assumption of random data so that our estimated 'pure' efficiency scores are linked to the observed data. As discussed earlier, it is not very convincing to argue that the observed data for a DMU are not the 'true' data, or that the frontier constructed from the observed data is not the 'true' frontier, and there are many stochastic data and frontiers lying around that can affect the 'true' efficiency scores. Therefore, we base our study on the unique DEA frontier enveloped from the unique dataset of the observed DMUs. This is indeed consistent with SFA, where the data, frontier and inefficiency are unique or deterministic, and only the noise is stochastic.

The stochastic DEA introduced by Banker (1988, 1996) and Banker et al. (2015) attempted to test for the sensitivity of DEA measurements under different settings, arguing that the deviation between a certain DMU and the frontier is a combination of the DEA inefficiency and a symmetric random error. In this sense, by introducing (and adjusting) the weight of the inefficiency component, one can

estimate different efficiency scores for that DMU—this helps explain the sensitivity and stability of stochastic DEA. However, since stochastic DEA focuses more on the inefficiency term and its weight, no confidence intervals for the efficiency scores can be estimated. We therefore extend this direction to be able to provide the confidence intervals for DEA efficiency scores by exploring the statistics of both the inefficiency and the random errors, similar to the SFA approach.

In DEAS, we first estimate a unique DEA linear frontier by following the basic Charnes et al. (1978) or the Charnes–Cooper–Rhodes (CCR) DEA model using Eq. (7) under the assumption of constant returns to scale. Nguyen et al. (2018) have showed that the frontiers of CCR-DEA and OLS SFA are very similar whilst the difference between variable returns to scale DEA and OLS SFA is only a practical issue. What we argue next is that even though the frontier is unique as well as the deviations (i.e. the DEA inefficiency scores), those deviations could still be impacted by statistical noise or random errors—the same argument has been applied in SFA. In other words, if a DMU deviates from the frontier, it may not all be attributed to inefficiency, but deviations could also be caused by random errors, such as luck or bad weather. Following the SFA approach, we decompose those deviations into two components, the ‘pure’ inefficiency (which can be later transformed into ‘pure’ efficiency) and the ‘pure’ errors. It is therefore more convincing to understand the difference between DEA inefficiency and ‘pure’ inefficiency from a managerial perspective, since there is no stochastic data or frontier involved.

Following Stevenson (1980), Banker (1988), and Banker and Maindiratta (1992), it is convenient and common to assume that the ‘pure’ errors follow a normal distribution whereas the inefficiency scores follow a non-zero mean truncated-normal distribution. Since the efficiency scores are bounded by $[0, 1]$, one can even imply a half-normal distribution for these efficiency scores. However, we choose the truncated-normal distribution because it is more general (Kumbhakar et al. 2015)—it is only a technical matter to extend our model to other distribution assumptions. It is also well known that the deviations themselves also follow a normal distribution because they are consistent with the residuals estimated from regression (Banker, 1988). Therefore, our aim is to estimate the two components of the DEA inefficiency scores so that these conditions are satisfied. The DEAS model can be formalized as follows.

Step 1 Compute the DEA efficiency scores (*DEA*) by using the CCR DEA model as in Eq. (7) above. The inefficiency scores (*IEF*) can be consequently computed as follows—the construction of efficiency scores in DEA differs from that of SFA explaining why Eqs. (8) and (16) are also different:

$$IEF_j = 1 - DEA_j, \quad j = 1, \dots, n \quad (8)$$

Step 2 Use MLE to estimate the ‘pure’ inefficiency (*pIEF*) component and the ‘pure’ error (*pER*) component from the observed *IEF* by maximising the log-likelihood function in Eq. (9) below, given the conditions in Eqs. (10)–(14). Note that steps 2, 3 and 4 exactly follow the SFA process given in Eqs. (1)–(6) above.

$$L_j = -\ln\left(\frac{1}{2}\right) - \frac{1}{2} \ln\left(\sigma_{pIEF}^2 + \sigma_{pER}^2\right) + \ln\phi\left(\frac{IEF_j}{\sqrt{\sigma_{pIEF}^2 + \sigma_{pER}^2}}\right) + \ln\Phi\left(\frac{\mu_{*j}}{\sigma_*}\right), \quad (9)$$

where:

$$IEF_j \sim i.i.d.N(\mu_{IEF}, \sigma_{IEF}); \quad (10)$$

$$pIEF_j \sim i.i.d.N(\mu_{pIEF}, \sigma_{pIEF}); \quad (11)$$

$$pER_j \sim i.i.d.N(0, \sigma_{pER}); \quad (12)$$

$$\mu_{*j} = \frac{-\sigma_{pIEF}^2 IEF_j}{\sigma_{pIEF}^2 + \sigma_{pER}^2}; \quad (13)$$

$$\sigma_* = \frac{\sigma_{pIEF}^2 \sigma_{pER}^2}{\sigma_{pIEF}^2 + \sigma_{pER}^2}. \quad (14)$$

Step 3 The approach of Jondrow et al. (1982) can be used to derive the predictor of $pIEF$ as in Eq. (15). The values of the ‘pure’ efficiency scores ($DEAS_j$) can then be obtained via Eq. (16):

$$\widehat{pIEF}_j = \mu_{*j} + \sigma_* \left[\frac{\phi(\mu_{*j}/\sigma_*)}{\Phi(\mu_{*j}/\sigma_*)} \right]; \quad (15)$$

$$DEAS_j = \exp(-pIEF_j). \quad (16)$$

Step 4 We can further construct the confidence intervals for $DEAS_j$ at an $(1 - \alpha)100\%$ level of significance, following Horrace and Schmidt (1996) and Hjalmarsson et al. (1996):

$$L_j = \mu_{*j} + \sigma_* \Phi^{-1}\{(1 - \alpha/2)\Phi(\mu_{*j}/\sigma_*)\}; \quad (17)$$

$$U_j = \mu_{*j} + \sigma_* \Phi^{-1}\{\alpha/2\Phi(\mu_{*j}/\sigma_*)\}; \quad (18)$$

$$\exp(-U_j) \leq DEAS_j \leq \exp(-L_j). \quad (19)$$

3 Monte Carlo simulation

3.1 Data generation process

To illustrate the functions of the proposed approach, we used a simple Monte Carlo experiment, following Giraleas et al. (2012). In particular, we examine a set of DMUs, which can be in any industry or sector, operating under the same constant return to scale Cobb–Douglas production function:

$$Y_i = K_i^{0.5} L_i^{0.5} \exp(\varepsilon_i), \tag{20}$$

where Y_i is the output of DMU i , K_i is the capital input of DMU i , L_i is the labour input of DMU i and ε_i is the random deviation following a normal distribution associated with DMU i . Note that ε_i will be treated as inefficiency by DEA but is further decomposed by SFA and DEAS. It is also important to note that all DEA and SFA simulations apply similar models to Eq. (20), in which inefficiency is defined as ε_i . In this sense, SFA and DEAS identify and estimate ‘pure’ inefficiency (and thus its counterpart ‘pure’ efficiency) based on ε_i . In contrast, bootstrap DEA estimates the bias-corrected efficiency based on the original DEA efficiency scores DEA_i and thus has a weak relationship with ε_i . Consequently, bootstrap DEA will be less accurate than SFA and DEAS, as shown in Table 1.

The data for K and L are randomly generated following a uniform distribution of $U[0, 1]$, whereas ε_i is randomly generated following an uniform distribution of $U(0, 0.1)$. The data of Y_i are then computed via Eq. (20) above, in which no DMU is fully (i.e. 100%) technically efficient because of the inclusion of ε_i . Since most DEA studies have a limited number of DMUs, we focused more on the simulated data of 100 DMUs; however, another simulation dataset of 200 DMUs was also generated for testing the robustness of our DEAS model.

As discussed previously, one basic characteristic that distinguishes SFA from DEA is how they treat ε_i . Since in DEAS, we are trying to do the same as in SFA, in contrast to DEA, we are also interested in how the pure inefficiency component

Table 1 Accuracy results of different efficiency measures

Simulated Size	Measure	MAD	MSE	AC
100 DMUs	SFA_i	0.0529	0.0051	1.0236
	DEA_i	0.0458	0.0030	0.9987
	$SW98_i$	0.0503	0.0034	1.0261
	$LT99_i$	0.0799	0.0087	1.0598
	$DEAS_i$	0.0441	0.0027	0.9984
200 DMUs	SFA_i	0.0535	0.0050	1.0222
	DEA_i	0.0380	0.0021	0.9899
	$SW98_i$	0.0352	0.0018	0.9964
	$LT99_i$	0.0567	0.0040	1.0645
	$DEAS_i$	0.0341	0.0017	0.9893

(DEAS) behaves compared with that of SFA. Therefore, we compare and contrasted the deviations from the frontiers by the simulated DMUs, which are ε_i (for the true frontier), v_i (for the SFA frontier), IEF_i (for the DEA frontier) and pER_i (for the DEAS frontier). In particular, by looking at the distributions of those components (i.e. the probability density function (PDF) and the cumulative distribution function (CDF)), we can check if they are normally distributed with a zero mean, a basic characteristic of the residuals.

Several indicators are also calculated from these simulated datasets, including the true efficiency scores $EF_i = \exp(-\varepsilon_i)$, the SFA efficiency scores SFA_i (Eq. 5), the DEA efficiency scores DEA_i (Eq. 7) and the DEAS efficiency scores $DEAS_i$ (Eq. 16), for each of the simulated DMUs. To compare and contrast with bootstrap DEA, two indicators using Simar and Wilson (1998) approach (*SW98*) and Lothgren and Tambour (1999) approach (*LT99*) are also estimated. The results from SFA, DEA, bootstrap DEA and DEAS will be compared with those of the true frontier [as in Eq. (20)] to gain a better insight into the accuracy of each method. In particular, we used the following accuracy indicators:

$$MAD = \frac{1}{n} \sum_{i=1}^n |EF_i - sEF_i|; \tag{21}$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (EF_i - sEF_i)^2; \tag{22}$$

$$AC = \frac{1}{n} \sum_{i=1}^n \frac{EF_i}{sEF_i}, \tag{23}$$

where EF_i is the ‘true’ efficiency scores; sEF_i are SFA_i , DEA_i , *SW98*, *LT99* and $DEAS_i$, respectively; *MAD* measures the mean absolute deviation of each indicator sEF_i from EF_i ; *MSE* measures the mean square error of each indicator sEF_i from EF_i ; *AC* measures the accuracy level of each indicator sEF_i compared with EF_i and n is the number of DMUs. According to Eqs. (21, 23), the indicator sEF_i will be more accurate and will more closely estimate the ‘true’ EF_i when *MAD* and *MSE* are small and when *AC* is closer to unity. *MAD* and *MSE* were used by Giraleas et al. (2012) and *AC* was used by Cross and Färe (2015), among others.

3.2 Results

In general, the analysis found that DEAS outperforms SFA in terms of deviation decomposition. In particular, DEAS provides less skewed efficiency scores and more normally distributed random errors.

In Fig. 1, we first illustrated the PDF and CDF of the simulated deviations from the estimated frontiers of SFA, DEA and DEAS, represented by v_i , IEF_i and pER_i , respectively. The ‘true’ frontier is defined in Eq. (20) and thus the ‘true’ deviations are given as ε_i . Although the distributions of IEF_i and pER_i are very similar to those

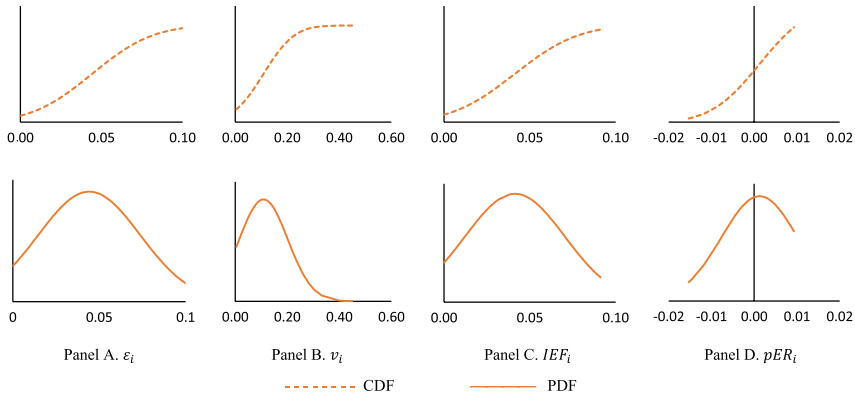


Fig. 1 Distribution functions of the random errors from examined methods. CDF: cumulative distribution function, PDF: probability density function, represents the (simulated) deviations from the ‘true’ frontier, indicates the deviations estimated by SFA, IEF_i indicates the deviations estimated by DEA and indicates the deviations estimated by DEAS

of ε_i (Panels A, C and D in Fig. 1), one can see that v_i (Panel B in Fig. 1) is an exception: it is left-skewed with a significantly lower variance (and thus a shorter tail). One may argue that since our data generation process only created positive deviations $\varepsilon_i \subset U(0, 0.1)$, whereas SFA will estimate $v_i \subset N(0, \sigma_v^2)$, some skewness in v_i is acceptable. We are not neglecting this issue; however, even if we account for the normal distribution of the random errors, DEAS still can be considered to be better than SFA, as pER_i follows $N(0, \sigma_v^2)$ more closely than v_i (see Panel D in Fig. 1). We therefore concluded that the ‘pure’ random error component pER_i estimated from DEAS is well-behaved and satisfies the requirements for decomposition.

Regarding the confidence intervals of the stochastic efficiency scores, we compared results from SFA, bootstrap DEA (SW98 and LT99) and DEAS for the 95% level of confidence. Since those estimated efficiency scores are different, especially for bootstrap DEA due to the characteristic of randomness in resampling, we expected to see some differences in their lower bounds and upper bounds as well. As discussed by many studies, SFA_i and its confidence intervals are higher than those based on DEA (Hjalmarsson et al. 1996; Bauer et al. 1998) and thus it is not a surprise to see that those measures are closer to unity (Panel A in Fig. 2). Panels B, C and D in Fig. 2 visually show that the three DEA-based measures (namely $SW98_i$, $LT99_i$ and $DEAS_i$) have a similar pattern, suggesting that DEAS can be seen as an alternative to bootstrap DEA, which is more computation intensive, bringing DEA closer to SFA in terms of statistical characteristics.

Figure 2 also shows that SFA sometimes estimates a proportion of DMUs to be located far from their ‘true’ efficiency (i.e. large differences between EF_i and SFA_i), and that estimates from DEA and DEAS are located closer together (i.e. smaller differences between EF_i and both DEA_i and $DEAS_i$). We then used mean absolute deviation (MAD), mean square error (MSE) and accuracy level (AC) to further test for the accuracy of different efficiency measures. The simulation

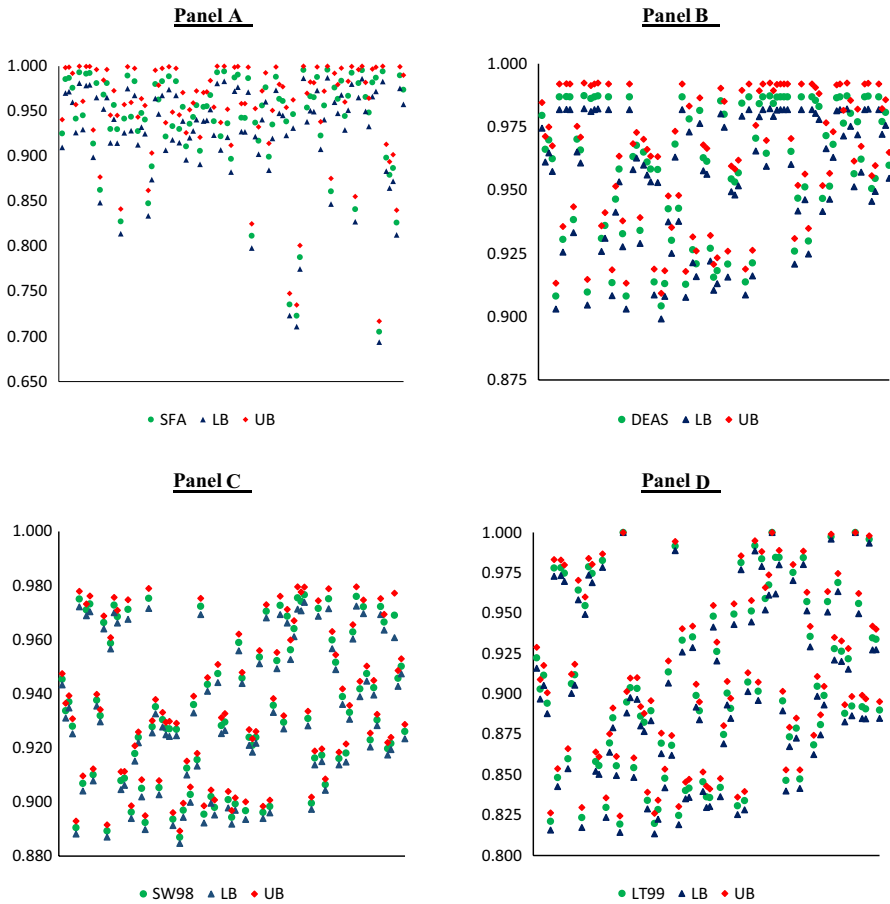


Fig. 2. 95% confidence intervals for the efficiency scores of the examined methods. Values on the vertical axis represent the efficiency scores; the horizontal axis represents the DMUs from 1 to 100. SW98 indicates efficiency scores estimated via Simar and Wilson (1998). LT99 indicates efficiency scores estimated via Lothgren and Tambour (1999). LB, lower bound; UB, upper bound

sample was increased to 200 DMUs to test for the robustness of DEAS – most DEA applications used DEA because of the small sample size issue and ended up with less than 200 DMUs. Table 1 presents the simulation results, where $DEAS_i$ has the lowest MAD and MSE throughout all simulations. In addition, we also observed that SFA tends to underestimate the efficiency ($AC > 1$), whereas DEA and its derivatives (i.e. $SW98_i$, $LT99_i$ and $DEAS_i$) tend to overestimate the ‘true’ efficiency score ($AC < 1$), which is caused by the characteristics of DEA and SFA, as discussed in Smith (1997) and Bauer et al. (1998). Table 1 also suggests that increasing the simulated sample size from 100 to 200 DMUs helps to improve the

accuracy of all DEA-based models, though it is not significant for SFA (e.g. *MAD* increases, but *MSE* and *AC* decrease).

4 Measuring the operational performance of Asia-Pacific airlines via the DEAS model

The previous section suggested that the proposed DEAS approach is a good alternative efficiency measure that combines the strengths and addresses the weaknesses of the SFA and DEA models. This section applies the new DEAS approach to examine the operational and financial efficiency of 14 Asia-Pacific airlines.

4.1 Data and model specification

The global economy is significantly boosted up by the aviation sector, with 62.7 million jobs created and US\$2.7 trillion generated [corresponding to 3.5% of the global gross domestic product (GDP)] in 2016 (Air Transport Action Group 2016). The direct and indirect impacts of the aviation industry have been recognised in business and trade development, infrastructure development and tourism, among others (Acar and Karabulak 2015). These developments, in turn, fuel the demand for air transportation, especially in the Asia-Pacific region (Heshmati and Kim 2016). A recent report from the IATA forecast revealed that the net profits of the Asia-Pacific's aviation industry reached US\$6.3 billion in 2016, second to North America. However, the report pointed out that airlines operating in the Asia-Pacific region performed differently, and thus an examination of the efficiency and performance of Asia-Pacific's airlines is justified (IATA 2016).

In our DEAS model, three key airline indicators (available seat-kilometres (*ASK*), available tonne-kilometres (*ATK*) and operating expenses (*EXPENSES*)) are used as inputs; another three key indicators (revenue passenger-kilometres (*RPK*), revenue tonne-kilometres (*RTK*) and operating revenues (*REVENUES*)) are used as outputs. These variables are commonly used in prior DEA studies in air transportation (Barbot et al. 2008; Lee and Worthington 2014; Min and Joo 2016) which allow us to analyse the overall performance of the sampled Asia-Pacific airlines in both operational and financial aspects. Note that operational performance and financial performance are equally important to airlines in seeking for profit maximization.

We further extended the DEAS model by using a two-stage approach: after the 'pure' efficiency scores (*DEAS_i*) were estimated as the true efficiency scores of the sampled airlines (excluding errors and biases) during the first-stage efficiency analysis, they were then regressed on the identified explanatory variables to see whether the air transport-related external factors affected the efficiency of the sampled Asia-Pacific airlines during the study period. Equation (24) presents the model specification of the second-stage regression analysis with the explanatory variables selected following prior air transportation studies (e.g. Heshmati and Kim (2016) and are discussed below.

$$\begin{aligned}
DEAS_{it} = & \beta_0 + \beta_1 FLAG_CARRIER_i \\
& + \beta_2 ALLIANCE_{it} \\
& + \beta_3 LISTED_{it} \\
& + \beta_4 \ln ASSETS_{it} \\
& + \beta_5 GFC_t + \beta_6 MH_ACCIDENTS_t \\
& + \beta_7 \ln GDP \text{ per capita}_{it} \\
& + \beta_8 \ln TOURIST_ARRIVALS_{it} \\
& + \beta_9 \ln TRADE_VOLUMES_t \\
& + \beta_{10} \ln FUEL_PRICES_{it} + \epsilon_{it}
\end{aligned} \tag{24}$$

where:

- $FLAG_CARRIER_i$ is a dummy variable that equals 1 if airline i is a national flag carrier and 0 otherwise. This variable was previously used in Fethi et al. (2002), Barros and Peypoch (2009) and Chen et al. (2017), among others;
- $ALLIANCE_{it}$ is a dummy variable that equals 1 if airline i belongs to an airline alliance (Oneworld, SkyTeam or Star Alliance) at year t and 0 otherwise (Park and Cho 1997; Kleymann and Seristö, 2001; Min and Joo 2016);
- $LISTED_{it}$ is a dummy variable that equals 1 if airline i was listed in the stock market at year t and 0 otherwise (Mar and Young 2001; Lu et al. 2012);
- $\ln ASSETS_{it}$ is the logarithmic value of the total assets of airline i in year t . This variable is used to capture the size of an airline (Feng and Wang 2000; Backx et al. 2002);
- GFC_t is a dummy variable that equals 1 for the year 2008 and 0 otherwise. This variable is used to capture the impact of the Global Financial Crisis (GFC) of 2008 on airline operations (Merkert and Morrell 2012; Tsui 2017);
- $MH_ACCIDENTS_t$ is a dummy variable that equals 1 for the period of 2014–2015, 0 otherwise. This variable is used to capture the impact of the two Malaysia Airlines flight accidents (MH370 and MH17) in 2014 on international air transportation demand (Hunter and Lambert 2016; Yang et al. 2018);
- $\ln GDP \text{ per capita}_{it}$ is the logarithmic value of the GDP per capita at year t of the country where airline i is based. This variable is used to capture the size of an economy as well as its air transportation demand (Khadaroo and Seetanah 2008; Gaggero and Bartolini 2012);
- $\ln TOURIST_ARRIVALS_{it}$ is the logarithmic value of the number of international tourist arrivals at year t to the country where airline i is based on. This variable is used to capture tourist arrivals using air transport (Bieger and Wittmer 2006; Khadaroo and Seetanah 2008; Rehman Khan et al. 2017);
- $\ln TRADE_VOLUMES_{it}$ is the logarithmic value of the total trade volumes (exports + imports) at year t of the country where airline i is based. This variable is used to capture the air cargo demand of the country (Yamaguchi 2008; Tsui and Fung 2016);
- $\ln FUEL_PRICES_t$ is the logarithmic value of fuel prices at year t (Assaf 2011; Lim and Hong 2014);

- ϵ_{it} is random errors.

This study examines 14 major airlines operating in the Asia-Pacific region during the period of 2008–2015 (see Table 2). Data were collected from the statistical reports of the Association of Asia Pacific Airlines (AAPA 2016). Because of data limitations, and for the panel data regression in the second stage of analysis, we devised a multi-period sample covering the years 2008, 2010, 2012, 2014, and 2015. Table 2 presents the variables of interest used for the DEAS model in the first-stage efficiency analysis and the second-stage regression analysis in this study. The values of all variables are in million US dollars, except for *TOURIST_ARRIVALS*, which is in persons; *FUEL_PRICES*, which is in US dollars; and dummy variables, which are binary. The variation in the variables of interest across the sampled airlines suggests that independent results from single-dimensional analysis such as *EXPENSES* or *REVENUES* may contradict each other and therefore a multi-dimensional efficiency analysis of the Asia-Pacific airlines with different indicators is needed.

Table 2 Descriptive statistics and data source

Variables	Obs	Mean	Std. dev	Min	Max	Data source
<i>First-stage airline inputs and outputs variables for the DEAS model</i>						
<i>ASK</i>	70	59,629,141	32,676,174	5,054,942	122,543,607	AAPA
<i>ATK</i>	70	9,808,676	5,535,441	752,407	20,036,772	AAPA
<i>EXPENSES</i>	70	6,175,911	4,562,154	285,837	17,237,370	AAPA
<i>RPK</i>	70	44,565,384	24,487,433	3,619,115	94,267,400	AAPA
<i>RTK</i>	70	6,723,464	3,887,410	446,241	15,116,355	AAPA
<i>REVENUES</i>	70	6,377,378	4,746,190	228,547	16,612,764	AAPA
<i>Second-stage variables for regression analysis</i>						
<i>FLAG_CARRIER_j</i>	70	0.786	0.413	0	1	Authors' own calculation
<i>ALLIANCE_{jt}</i>	70	0.857	0.352	0	1	Authors' own calculation
<i>LISTED_{jt}</i>	70	0.843	0.367	0	1	Authors' own calculation
<i>lnASSETS_{jt}</i>	70	15.562	0.998	13.171	17.024	AAPA
<i>GFC_t</i>	70	0.800	0.403	0	1	World Bank
<i>MH_ACCIENTS_t</i>	70	0.400	0.493	0	1	Authors' own calculation
<i>lnGDP per capita_{jt}</i>	70	9.601	1.181	7.102	10.914	World Bank
<i>lnTOURIT_ARRIVALS_{jt}</i>	70	15.785	1.112	12.211	17.214	World Bank
<i>lnTRADE_VOLUMES_{jt}</i>	70	26.871	1.200	23.177	28.288	World Bank
<i>lnFUEL_PRICES_t</i>	70	0.876	0.261	0.420	1.117	EIA

AAPA denotes Association of Asia Pacific Airlines, EIA denotes U.S. Energy Information Administration

4.2 Efficiency of Asia-Pacific airlines

4.2.1 First-stage efficiency analysis (DEAS model)

In this stage, the DEAS model is independently applied to yearly data (e.g. 2008 or 2010) to obtain the ‘pure’ efficiency scores, $DEAS_i$. In general, the sampled Asia-Pacific airlines performed relatively well during the period of 2008–2015, with high $DEAS_i$ values between 0.943 and 0.997 (see Table 3).

Table 3 also indicates a decreasing trend in the average efficiency of the sampled airlines, where the average $DEAS_i$ dropped from the level of 0.994 in 2008 to 0.973 in 2015. This finding is in line with Li et al. (2016), who found that the efficiency of the major Asian carriers (e.g. Air China, Korean Air, Singapore Airlines, Asiana and Malaysia Airlines) declined during the period of 2008–2012. The results in the last two columns of Table 3 further confirm the impact of the GFC of 2008 and the two Malaysia Airlines flight accidents in 2014, as presented later in this section.

However, a further examination of the airline efficiency levels in Fig. 3 provides a different perspective of Asia-Pacific airlines’ performance over the study period. At first, if one looks at the results from the DEA model (i.e. the dotted line for DEA_i in Fig. 3), it shows that the sampled Asia-Pacific airlines performed differently from each other: the high-performing group with the highest possible efficiency score of 1.000 or 100% efficiency includes EVA Air, Qantas and Vietnam Airline; the low-performance group includes Garuda Indonesia, Royal Brunei, All Nippon Airways and Thai Airways. If we look at the ‘pure’ efficiency derived from the DEAS model (i.e. the solid line for $DEAS_i$ in Fig. 3), however, we still see that airlines performed

Table 3 Pure efficiency scores of Asia-Pacific airlines estimated via DEAS

Airlines	2008	2010	2012	2014	2015	ΔGFC	ΔMH
Royal Brunei (BI)	0.997	0.976	0.991	0.973	0.943	-0.054	-0.018
EVA Air (BR)	0.993	0.983	0.991	0.973	0.980	-0.014	-0.018
China Airlines (CI)	0.993	0.983	0.992	0.973	0.980	-0.014	-0.019
Garuda Indonesia (GA)	0.993	0.975	0.991	0.974	0.969	-0.024	-0.017
Japan Airlines (JL)	0.997	0.983	0.991	0.973	0.980	-0.017	-0.018
Korean Air (KE)	0.993	0.985	0.992	0.973	0.980	-0.014	-0.019
Malaysia Airlines (MH)	0.993	0.983	0.991	0.973	0.980	-0.014	-0.018
All Nippon Airways (NH)	0.993	0.985	0.983	0.979	0.970	-0.023	-0.004
Asiana Airlines (OZ)	0.993	0.983	0.991	0.977	0.975	-0.019	-0.014
Philippine Airlines (PR)	0.993	0.983	0.992	0.973	0.955	-0.038	-0.019
Qantas (QF)	0.993	0.983	0.991	0.973	0.980	-0.014	-0.018
Singapore Airlines (SQ)	0.993	0.983	0.991	0.973	0.980	-0.014	-0.018
Thai Airways (TG)	0.997	0.985	0.991	0.977	0.975	-0.022	-0.014
Vietnam Airlines (VN)	0.993	0.983	0.991	0.973	0.980	-0.014	-0.018
Average	0.994	0.982	0.991	0.974	0.973	-0.021	-0.017

ΔGFC measures the differences between $DEAS_i$ for 2015 and 2008, ΔMH measures the differences between $DEAS_i$ for 2014 and 2012

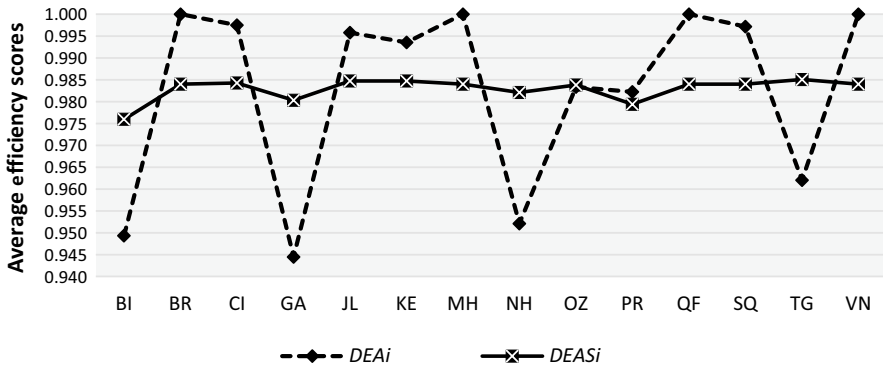


Fig. 3 Average efficiency scores derived from DEA and DEAS models

differently but were less volatile (the average $DEAS_i$ efficiency scores are between 0.976 and 0.985). Differences between the means of DEA_i and $DEAS_i$ are statistically significant at 1% (t -test); however, the two approaches ranked the examined airlines similarly (Mann–Whitney test). As shown in this study, DEAS has less discriminatory power than DEA. It is not surprising, however, since DEAS is in the same line with SFA, and SFA is well known to have higher efficiency scores and lower standard deviation compared to DEA (Hjalmarsson et al. 1996; Bauer et al. 1998; Michaelides et al. 2009; Roberts 2014).

In particular, the findings of the $DEAS_i$ scores of the efficient airlines (e.g. Singapore Airlines, China Airlines, Korean Air or Japan Airlines) are consistent with those in previous studies (Arjomandi and Seufert 2014; Chang et al. 2014), which suggests that these airlines are top performers in the Asia-Pacific region as well as globally (notice that the dEF_i scores rank them as second-tier performers, behind EVA Air, MH, Qantas and Vietnam Airline). In addition, Thai Airways is now considered as a good performer in the group, which is more reasonable and in line with the argument that Thai Airways should perform well thanks to its high managerial efficiency (Lee and Worthington 2014; Min and Joo 2016). The relatively low-performing group according to their $DEAS_i$ scores still consists of Garuda Indonesia and Royal Brunei but also Philippine Airlines. It is noted that DEA also provides information on peer DMUs (i.e. efficient DMUs that can be the role models for inefficient ones) while SFA does not. Since DEAS stands between DEA and SFA, for the managerial implications and decision-making, airline managers still can utilise the peers’ information from the first stage DEA [i.e. Equation (7)] to compare their performance against other sampled airlines, although this is not reported here due to the scope of our research.

In terms of the confidence intervals for the $DEAS_i$ scores, at the 95% level of significance, we can confidently argue that it was possible that some of the sampled Asia-Pacific airlines in this study did not perform well and were less efficient than other airlines in the group during the analysis period. For example, the efficiency of Royal Brunei went down to 0.810 in 2015 and the efficiency of Garuda Indonesia in 2014 also dropped to 0.752 (see Fig. 4). By comparing the ‘pure’ efficiency scores

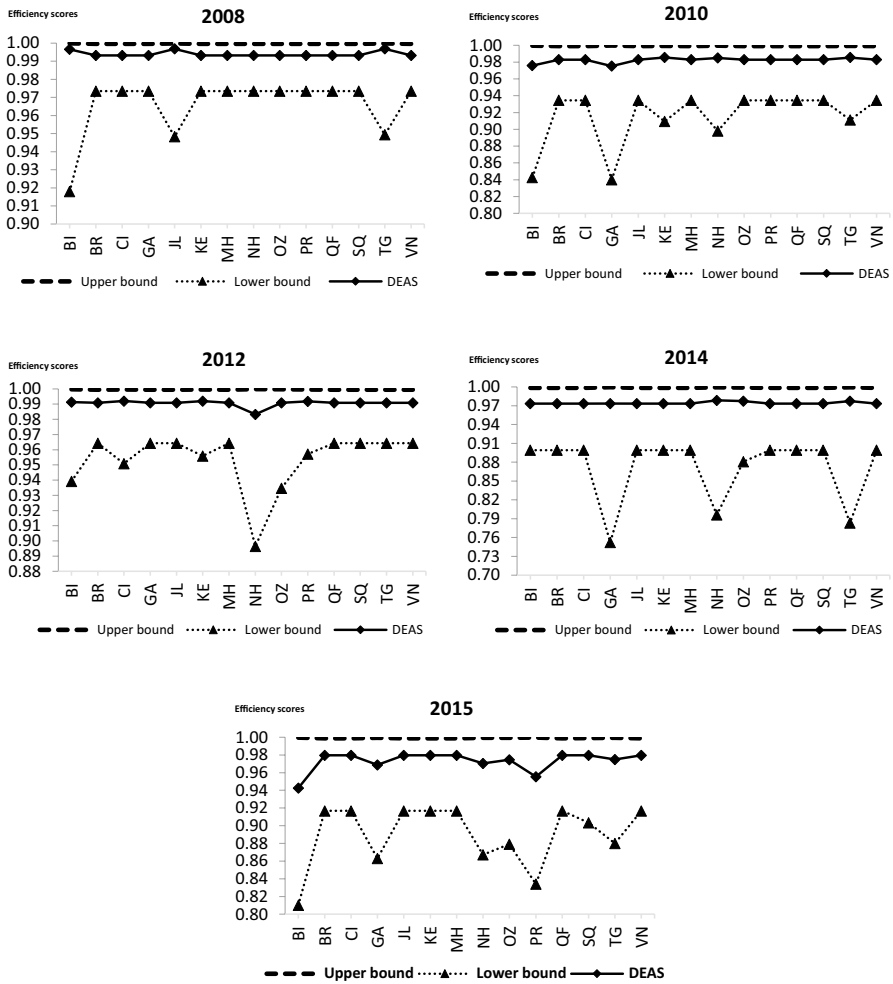


Fig. 4 Confidence intervals of the DEAS efficiency scores of Asia-Pacific airlines. All the plots are presented in 95% confidence intervals

of Royal Brunei and Garuda Indonesia to those of other sampled airlines in the same year, we can see that if we exclude the random errors and statistical noise from the DEA efficiency scores, other factors potentially had significant impact on the sampled Asia-Pacific airlines’ operations and hindered their performance. Therefore, a second-step regression analysis is performed for examining the significant determinants of the efficiency and performance of the sampled Asia-Pacific airlines.

4.2.2 Second-stage of regression analysis

The second-stage Tobit regression analysis was estimated following Eq. (24) in Sect. 4.1. The coefficients of the explanatory variables are shown in Table 4, where

Table 4 Results of the second-stage Tobit regression of airlines' 'pure' efficiency

Variables	DEAS		Bootstrap DEA	
	Coefficient	Std. error	Coefficient	Std. error
Intercept	0.9681***	0.0301	0.8734***	0.2197
<i>FLAG_CARRIER_i</i>	0.0020	0.0019	0.0097	0.0142
<i>ALLIANCE_{it}</i>	0.0043	0.0036	0.0090	0.0262
<i>LISTED_{it}</i>	0.0039	0.0031	0.0071	0.0229
<i>lnASSETS_{it}</i>	- 0.0016	0.0015	0.0063	0.0110
<i>GFC_t</i>	- 0.0064***	0.0017	- 0.0381**	0.0146
<i>MH_ACCIDENTS_t</i>	- 0.0114***	0.0015	- 0.0043	0.0110
<i>lnGDPpercapita_{it}</i>	0.0004	0.0009	- 0.0021	0.0063
<i>lnTOURIST_ARRIVALS_{it}</i>	0.0008	0.0017	- 0.0018	0.0124
<i>lnTRADE_VOLMES_t</i>	0.0005	0.0017	0.0019	0.0127
<i>lnFUEL_VOLMES_t</i>	0.0081***	0.0028	0.0293	0.0199

***, **Significant at 1% and 5% levels, respectively

results from DEAS and bootstrap DEA are generally consistent but DEAS provides more insights (i.e. more significant variables). Those results therefore strengthen the value of DEAS in comparison to bootstrap DEA.

With regard to *DEAS_i*, the significant variables affecting the sampled Asia-Pacific airlines' performance during the study periods are the three macroeconomic variables (i.e. *GFC*, *MH_ACCIDENTS* and *lnFUEL_PRICES*). The remaining variables are statistically insignificant although the directions of the relationships (i.e. the signs of the coefficients) between the predictors and the pure efficiency scores are consistent with those in the literature (*lnGDPpercapita*, *lnTRADE_VOLUMESS* and *lnTOURIST_ARRIVALS* are positively associated with *DEAS_i*). Note that Tobit regression using *DEA_i* as a dependent variable resulted in no significant predictor—this result was not reported here but is available upon request.

Regarding the significant explanatory factors that influenced the performance of the sampled Asia-Pacific airlines during the period of 2008–2015, both the GFC of 2008/09 and the two Malaysia Airlines flight accidents in 2014 had statistically significant and negative impacts on the airlines' pure efficiency: these empirical findings confirm our expectation of the expected negative impacts of exogenous shocks, as shown in the last two columns of Table 3. These empirical results are also consistent with the findings of Heshmati and Kim (2016), who found that the Asian financial crisis of 1997/98, the 9/11 terrorist attacks, and the GFC 2008 had negative shocks to (cost) efficiency of 39 airlines in 33 different countries during the period of 1998–2012. For the managers, it indicates that Asia-Pacific airlines need to take steps to improve their efficiency by overcoming and mitigating the hindrances of any exogenous shocks, and by leveraging other inputs (such as careful strategic planning in airline connectivity and routes).

Interestingly, fuel prices were reported to have a significant and positive impact on airlines' efficiency. This empirical finding is mainly because fuel prices dropped

continuously during the study period (i.e. from US\$2.964 per gallon in 2008 to a sharp decrease in 2010 and continued to US\$1.522 per gallon in 2015). One might argue that lower fuel prices allowed the sampled Asia-Pacific airlines to save more on operating costs, as fuel prices are the largest operating cost for airlines (Ryerson et al. 2014; Zou et al. 2014) and thus they improved their financial performance. However, the revenues of the sampled airlines also decreased, cancelling out the effect of lower fuel prices. Specifically, as shown in Fig. 5, $\ln FUEL_PRICES$ and $\ln EXPENSES$ for the sampled Asia-Pacific airlines were moving together with a similar pattern, and although the two variables decreased continuously during the 2012–2015 period, $\ln REVENUES$ of the sampled Asia-Pacific airlines also faced a sharp decline during the same period. Consequently, it is reasonable that the drop in fuel prices did not contribute much to the revenue as well as the financial and overall performance of the sampled Asia-Pacific airlines.

5 Conclusions

This study has proposed a novel DEAS approach to decompose the deviations of DMUs from the frontier derived via DEA with a SFA-like method. The key contribution of the novel DEAS approach in this study is its ability to estimate the confidence intervals for the calculated efficiency scores of DMUs. In addition, Monte Carlo simulations showed that the ‘pure’ efficiency scores ($DEAS_i$) decomposed from DEA (in)efficiency are accurate measures of the ‘true’ efficiency scores (EF_i), and thus the DEAS approach is believed to be a good alternative or complement for conventional DEA and bootstrap DEA for analysing the efficiency levels of DMUs.

For the empirical application, this study applied the proposed model to measure the performance of key Asia-Pacific airlines during the period of 2008–2015. We found that after accounting for the ‘pure’ random errors, the sampled Asia-Pacific airlines performed well during the study period but their ‘pure’ efficiency was

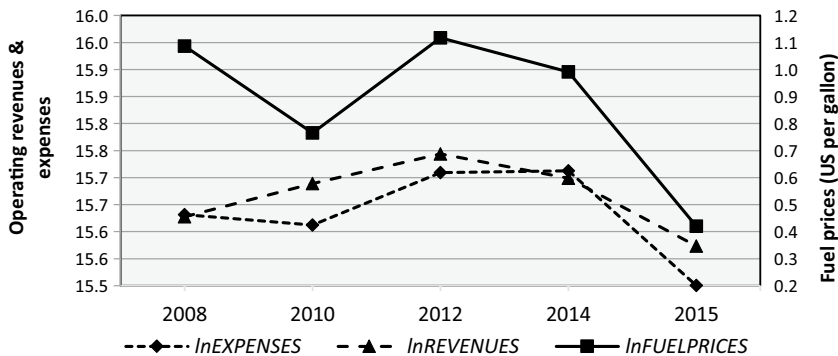


Fig. 5 Relationship between fuel prices and the operating revenues and expenses of Asia-Pacific airlines (2008–2015)

declining. Although the key reasons for efficiency deterioration were mainly attributed to uncontrollable factors (i.e. the GFC 2008/09 and the two Malaysia Airlines flight accidents in 2014), there is still room for improvement.

Although the novel DEAS model used in this study is shown to be a robust and fruitful approach to measuring the performance and efficiency of DMUs (e.g. Asia-Pacific airlines) in the Monte Carlo simulation, this study examines neither the likely impacts of airline-related characteristics (e.g. national flag carrier status or alliance membership) nor the macroeconomic environment (e.g. GDP per capita or number of international tourist arrivals) on their performance and efficiency because of data limitations. For future research, it would be meaningful to extend the dataset to include more airlines (legacy and low-cost carriers) and longer (continuous) time periods to explore their likely impact on airline efficiency and performance. Furthermore, this novel DEAS model can be further extended; for example, for considering the variable return to scale assumption (Banker et al. 1984) or incorporating the explanatory variables directly into the DEAS model (Battese and Coelli 1995; Simar and Wilson 2007) for performance and efficiency measurement.

References

- AAPA (2016) Statistical report 2016. Association of Asia Pacific Airlines, Kuala Lumpur
- Acar AZ, Karabulak S (2015) Competition between full service network carriers and low cost carriers in Turkish airline market. *Proc Soc Behav Sci* 207:642–651
- Adler N, Martini G, Volta N (2013) Measuring the environmental efficiency of the global aviation fleet. *Transp Res Part B Methodol* 53:82–100
- Aigner DJ, Lovell CAK, Schmidt P (1977) Formulation and estimation of stochastic frontier production function models. *J Econom* 6:21–37
- Air Transport Action Group (2016) Aviation: benefits beyond borders. Air Transport Action Group (ATAG), Geneva
- Air Transport Action Group. (2019). Facts and figures. Retrieved from <https://www.atag.org/facts-figures.html>
- Arjomandi A, Seufert JH (2014) An evaluation of the world's major airlines' technical and environmental performance. *Econ Model* 41:133–144
- Assaf AG (2011) The operational performance of UK airlines: 2002–2007. *J Econ Stud* 38(1):5–16
- Assaf AG, Gillen D, Tsionas EG (2014) Understanding relative efficiency among airports: a general dynamic model for distinguishing technical and allocative efficiency. *Transp Res Part B Methodol* 70:18–34
- Azadeh A, Rahimi Y, Zarrin M, Ghaderi A, Shabanpour N (2017) A decision-making methodology for vendor selection problem with uncertain inputs. *Transp Lett* 9(3):123–140
- Azadeh A, Salehi V, Kianpour M (2018) Performance evaluation of rail transportation systems by considering resilience engineering factors: Tehran railway electrification system. *Transp Lett* 10(1):12–25
- Backx M, Carney M, Gedajlovic E (2002) Public, private and mixed ownership and the performance of international airlines. *J Air Transp Manag* 8(4):213–220
- Banker RD (1996) Hypothesis tests using data envelopment analysis. *J Prod Anal* 7(2):139–159
- Banker RD, Maindiratta A (1992) Maximum likelihood estimation of monotone and concave production frontiers. *J Prod Anal* 3(4):401–415
- Banker RD, Charnes A, Cooper WW (1984) Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Manag Sci* 30(9):1078–1092
- Banker RD, Kotarac K, Neralić L (2015) Sensitivity and stability in stochastic data envelopment analysis. *J Oper Res Soc* 66(1):134–147
- Banker RD (1988) Stochastic data envelopment analysis. Carnegie-Mellon University, Working Paper

- Barbot C, Costa Á, Sochirca E (2008) Airlines performance in the new market context: a comparative productivity and efficiency analysis. *J Air Transp Manag* 14(5):270–274
- Barros CP, Peypoch N (2009) An evaluation of European airlines' operational performance. *Int J Prod Econ* 122(2):525–533
- Battese GE, Coelli TJ (1995) A model for technical inefficiency effects in a stochastic frontier production function for panel data. *Empir Econ* 20(2):325–332
- Bauer PW, Berger AN, Ferrier GD, Humphrey DB (1998) Consistency conditions for regulatory analysis of financial institutions: a comparison of frontier efficiency methods. *J Econ Bus* 50(2):85–114
- Bieger T, Wittmer A (2006) Air transport and tourism—perspectives and challenges for destinations, airlines and governments. *J Air Transp Manag* 12(1):40–46
- Bogetoft P, Otto L (2011) *Benchmarking with DEA, SFA, and R*. Springer
- Chang Y-T, Park H-S, Jeong J-B, Lee J-W (2014) Evaluating economic and environmental efficiency of global airlines: a SBM-DEA approach. *Transp Res Part D: Transp Environ* 27:46–50
- Charnes A, Cooper WW, Rhodes E (1978) Measuring the efficiency of decision making units. *Eur J Oper Res* 2(6):429–444
- Chen S-J, Chen M-H, Wei H-L (2017) Financial performance of Chinese airlines: does state ownership matter? *J Hosp Tour Manag* 33:1–10
- Coelli TJ, Rao DSP, O'Donnell CJ, Battese GE (2005) *An introduction to efficiency and productivity analysis*, 2nd edn. Springer, Berlin
- Cross RM, Färe R (2015) Value data and the fisher index. *Theor Econ Lett* 5(2):262–267
- Efron B, Tibshirani RJ (1994) *An introduction to the bootstrap*. CRC Press, Boca Raton
- Feng C-M, Wang R-T (2000) Performance evaluation for airlines including the consideration of financial ratios. *J Air Transp Manag* 6(3):133–142
- Fethi MD, Jackson PM, Weyman-Jones T (2002) *Measuring the efficiency of European airlines: An application of DEA and tobit analysis*. Discussion Paper, University of Leicester
- Gaggero AA, Bartolini D (2012) The determinants of airline alliances. *JTEP* 46(3):399–414
- Giraleas D, Emrouznejad A, Thanassoulis E (2012) Productivity change using growth accounting and frontier-based approaches: evidence from a Monte Carlo analysis. *Eur J Oper Res* 222(3):673–683
- Gong B-H, Sickles RC (1992) Finite sample evidence on the performance of stochastic frontiers and data envelopment analysis using panel data. *J Econ* 51(1–2):259–284
- Grosskopf S (1996) Statistical inference and nonparametric efficiency: a selective survey. *J Prod Anal* 7(2–3):161–176
- Heshmati A, Kim J (2016) *Efficiency and competitiveness of international airlines*. Springer, Berlin
- Hjalmarsson L, Kumbhakar S, Heshmati A (1996) DEA, DFA and SFA: a comparison. *J Prod Anal* 7(2–3):303–327
- Horrace W, Schmidt P (1996) Confidence statements for efficiency estimates from stochastic frontier models. *J Prod Anal* 7(2–3):257–282
- Huang Z, Li SX (2001) Stochastic DEA models with different types of input-output disturbances. *J Prod Anal* 15(2):95–113
- Hunter JA, Lambert JR (2016) Do we feel safer today? The impact of smiling customer service on airline safety perception post 9–11. *J Transp Secur* 9(1):35–56
- IATA (2016) *Industry economic performance*. International Air Transport Association (IATA), Geneva
- Jondrow J, Lovell CAK, Materov IS, Schmidt P (1982) On the estimation of technical inefficiency in the stochastic frontier production function model. *J Econ* 19:233–238
- Kang CC, Feng CM, Liao BR, Khan HA (2019) Accounting for air pollution emissions and transport policy in the measurement of the efficiency and effectiveness of bus transits. *Transportation Letters*, pp. 1–13
- Kao C, Liu S-T (2009) Stochastic data envelopment analysis in measuring the efficiency of Taiwan commercial banks. *Eur J Oper Res* 196(1):312–322
- Khadaroo J, Seetanah B (2008) The role of transport infrastructure in international tourism development: a gravity model approach. *Tour Manag* 29(5):831–840
- Kleymann B, Seristö H (2001) Levels of airline alliance membership: balancing risks and benefits. *J Air Transp Manag* 7(5):303–310
- Kumbhakar SC, Wang H-J, Horncastle AP (2015) *A practitioner's guide to stochastic frontier analysis using stata*. Cambridge University Press, New York
- Lan LW, Lin ETJ (2005) Measuring railway performance with adjustment of environmental effects, data noise and slacks. *Transportmetrica* 1(2):161–189

- Lee BL, Worthington AC (2014) Technical efficiency of mainstream airlines and low-cost carriers: new evidence using bootstrap data envelopment analysis truncated regression. *J Air Transp Manag* 38:15–20
- Li Y, Wang Y-Z, Cui Q (2016) Has airline efficiency affected by the inclusion of aviation into European Union Emission Trading Scheme? Evidences from 22 airlines during 2008–2012. *Energy* 96:8–22
- Lim SH, Hong Y (2014) Fuel hedging and airline operating costs. *J Air Transp Manag* 36:33–40
- Lothgren M, Tambour M (1999) Bootstrapping the data envelopment analysis Malmquist productivity index. *Appl Econ* 31(4):417–425
- Lu W-M, Wang W-K, Hung S-W, Lu E-T (2012) The effects of corporate governance on airline performance: Production and marketing efficiency perspectives. *Transp Res Part E Logist Transp Rev* 48(2):529–544
- Mar P, Young MN (2001) Corporate governance in transition economies: a case study of two Chinese airlines. *J World Bus* 36(3):280–302
- Meeusen W, van den Broeck J (1977) Efficiency estimation from Cobb-Douglas production functions with composed error. *Int Econ Rev* 18(2):435–444
- Merkert R, Morrell PS (2012) Mergers and acquisitions in aviation: management and economic perspectives on the size of airlines. *Transp Res Part E Logist Transp Rev* 48(4):853–862
- Michaelides PG, Belegri-Roboli A, Karlaftis M, Marinou T (2009) International air transportation carriers: evidence from SFA and DEA technical efficiency results (1991–2000). *Eur J Transp Infrastruct Res* 9(4):347–362
- Min H, Joo S-J (2016) A comparative performance analysis of airline strategic alliances using data envelopment analysis. *J Air Transp Manag* 52:99–110
- Ngo T, Tsui KWH (2020) A data-driven approach for estimating airport efficiency under endogeneity: an application to New Zealand airports. *Res Transp Bus Manag* 34:100412
- Ngo T, Le T, Tran SH, Nguyen A, Nguyen C (2019) Sources of the performance of manufacturing firms: evidence from Vietnam. *Post-Communist Econ* 31(6):790–804
- Nguyen T, Tripe D, Ngo T (2018) Operational efficiency of bank loans and deposits: a case study of vietnamese banking system. *Int J Financ Stud* 6(1):14
- Olesen OB, Petersen NC (2016) Stochastic data envelopment analysis: a review. *Eur J Oper Res* 251(1):2–21
- Park NK, Cho D-S (1997) The effect of strategic alliance on performance: a study of international airline industry. *J Air Transp Manag* 3(3):155–164
- Rehman Khan SA, Qianli D, SongBo W, Zaman K, Zhang Y (2017) Travel and tourism competitiveness index: The impact of air transportation, railways transportation, travel and transport services on international inbound and outbound tourism. *J Air Transp Manag* 58:125–134
- Roberts CM (2014) Efficiency in the U.S. Airline Industry. (PhD Thesis), The University of Leeds
- Ruggiero J (2004) Data envelopment analysis with stochastic data. *J Oper Res Soc* 55(9):1008–1012
- Ryerson MS, Hansen M, Bonn J (2014) Time to burn: flight delay, terminal efficiency, and fuel consumption in the National Airspace System. *Transp Res Part A Policy Pract* 69:286–298
- Simar L, Wilson PW (1998) Sensitivity analysis of efficiency scores: how to bootstrap in nonparametric frontier models. *Manag Sci* 44(1):49–61
- Simar L, Wilson PW (2000) Statistical inference in nonparametric frontier models: the state of the art. *J Prod Anal* 13(1):49–78
- Simar L, Wilson PW (2007) Estimation and inference in two-stage, semi-parametric models of production processes. *J Econ* 136:31–64
- Smith P (1997) Model misspecification in Data Envelopment Analysis. *Ann Oper Res* 73:233–252
- Stevenson RE (1980) Likelihood functions for generalized stochastic frontier estimation. *J Econ* 13(1):57–66
- Tovar B, Wall A (2017) Dynamic cost efficiency in port infrastructure using a directional distance function: accounting for the adjustment of quasi-fixed inputs over time. *Transp Sci* 51(1):296–304
- Tsui WHK (2017) Does a low-cost carrier lead the domestic tourism demand and growth of New Zealand? *Tour Manag* 60:390–403
- Tsui WHK, Fung MKY (2016) Analysing passenger network changes: the case of Hong Kong. *J Air Transp Manag* 50:1–11
- Tziogkidis P (2012) Monte Carlo experiments on bootstrap DEA. Cardiff Economics Working Papers
- Viton PA (1997) Technical efficiency in multi-mode bus transit: a production frontier analysis. *Transp Res Part B Methodol* 31(1):23–39

- Yamaguchi K (2008) International trade and air cargo: analysis of US export and air transport policy. *Transp Res Part E: Logist Transp Rev* 44(4):653–663
- Yang L, Tjiptono F, Poon WC (2018) Will you fly with this airline in the future? An empirical study of airline avoidance after accidents. *J Travel Tour Mark* 35(9):1145–1159
- Zou B, Elke M, Hansen M, Kafle N (2014) Evaluating air carrier fuel efficiency in the US airline industry. *Transp Res Part A Policy Pract* 59:306–330

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.