Joe Zhu

Foisie Business School, Worcester Polytechnic Institute, Worcester, MA 01609, USA

email: jzhu@wpi.edu

Tel.: +1 508-831-5467, Fax: +1 508-831-5720

Annals of Operations Research (in press)

August 2019
January 2020 (revised)
March 2020 (revised)
May 2020 (revised)

Acknowledgements: The author thanks anonymous reviewers for their helpful and constructive suggestions and comments on an earlier version of the paper. Earlier versions of the paper were presented at the Data Envelopment Analysis International Conference (June 28-30, 2019, Canterbury, UK) and the 2019 International Conference on Intelligent Transportation and Logistics with Big Data & the 7th International Forum on Decision Sciences, July 26-29, 2019, Windsor, Canada. The author is grateful for the comments and suggestions provided by the conference participants.

updates

DEA under big data: data enabled analytics and network data envelopment analysis

Joe Zhu[1]

© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

This paper proposes that data envelopment analysis (DEA) should be viewed as a method (or tool) for data-oriented analytics in performance evaluation and benchmarking. While computational algorithms have been developed to deal with large volumes of data (decision making units, inputs, and outputs) under the conventional DEA, valuable information hidden in big data that are represented by network structures should be extracted by DEA. These network structures, e.g., transportation and logistics systems, encompass a broader range of inter-linked metrics that cannot be modelled by conventional DEA. It is proposed that network DEA is related to the value dimension of big data. It is shown that network DEA is different from standard DEA, although it bears the name of DEA and some similarity with conventional DEA. Network DEA is big data enabled analytics (big DEA) when multiple (performance) metrics or attributes are linked through network structures. These network structures are too large or complex to be dealt with by conventional DEA. Unlike conventional DEA that are solved via linear programming, general network DEA corresponds to nonconvex optimization problems. This represents opportunities for developing techniques for solving non-linear network DEA models. Areas such as transportation and logistics system as well as supply chains have a great potential to use network DEA in big data modeling.

Keywords Data envelopment analysis (DEA) · Data enabled analytics · Big data · Performance · Productivity · Efficiency · Composite index · Transportation

1. Introduction

Data envelopment analysis (DEA) was coined by Charnes, Cooper, and Rhodes (1978). Since its first publication in 1978, DEA has been developed and applied in many different areas, resulting in over 5,000 publications in the Web of Science database. For comprehensive reviews on DEA literature, interested readers are referred to Cook and Seiford (2009), Liu, Lu, Lu, and Lin (2013a;2013b), and Liu, Lu, and Lu (2016).

As pointed by Cooper, Seiford, and Zhu (2004), the DEA literature has seen a great variety of applications in evaluating the performances of many different kinds of entities engaged in many different activities in many different contexts in many different countries. These DEA applications have used decision making units (DMUs) of various forms, such as hospitals, US Air Force wings, universities, cities, courts, business firms, countries, regions, etc. Because it requires very few assumptions, DEA has also opened up possibilities for use in cases which have been resistant to other approaches because of the complex (often unknown) nature of the relations between the multiple metrics labeled as inputs and outputs related to DMUs.

The focus of this paper is not on how great and versatile DEA has been, but rather on how DEA has been evolving. As big data research becomes an important area of Operations Analytics, DEA is evolving into Data Enabled Analytics. DEA can be viewed as a data-oriented data science tool for productivity analytics, benchmarking, performance evaluation, and composite index construction, among other new uses, in addition to the traditional uses such as, production efficiency and productivity measurement. Interestingly enough, Mahajan (1991) labeled DEA as "data envelopment analytics". The DEA community has witnessed the linkage between DEA and data analytics. A number of journal special issues have focused on DEA and its uses as a data-oriented and/or data science tool. INFOR has dedicated two volumes to DEA and its applications in operations (Lim and Zhu, 2017, 2018). Chen, Lim, and Cook (2019) have edited a special issue for Annals of Operations Research on DEA and data analytics. Charles, Aparicio, and Zhu (2020a) are editing a special issue for the Journal of the Operational Research Society on big data for better productivity. Charles, Aparicio, and Zhu (2020b) are editing a book on data science and productivity analytics.

The rest of the paper is organized as follows. Section 2 briefly introduces the conventional DEA and discusses some basic well-known properties of the DEA models. The emphasis is on the use of DEA as a benchmarking tool. Section 3 links the network structures to the big data concept and demonstrates that network DEA can derive insights and value from big data. Examples of network DEA in transportations and logistics are reviewed. It is also shown that network DEA is different from conventional DEA and may require the development and use of non-linear optimization techniques. Section 4 concludes.

2. Data Envelopment Analysis (DEA)

One often characterizes DEA as a tool for identifying best-practices when multiple performance metrics or measures are present for organizations. Although DEA has a strong link to production theory in economics, the tool is also used for benchmarking in operations management, where a set of measures is selected to benchmark the performance of manufacturing and service operations. In the circumstance of benchmarking, the efficient DMUs, as defined by DEA, may not necessarily form a "production frontier", but rather lead to a "best-practice frontier" (Cook, Tone, and Zhu, 2014). DEA can be a tool for constructing a composite index. For example, Shwartz, Burgess, and Zhu (2016) use an input only DEA model to develop a quality index for health care providers. Shen et al. (2012) develop a DEA based road safety model to measure the road safety risk. Chen et al. (2019) re-visit the global food security index by a hierarchical DEA.

Let us look at the very first standard DEA, often called the CCR model or CRS (constant returns to scale) model. I will talk about the use of returns to scale (RTS) in DEA in section 2.1. This standard DEA model can be presented in either its envelopment or multiplier form. For example, the multiplier CRS model is developed based upon the concept of engineering ratio by (Charnes, Cooper, and Rhodes, 1978):

$$\text{maximize } \frac{\sum_{r=1}^{s} u_r y_{ro}}{\sum_{i=1}^{m} v_i x_{io}}$$

s.t.

$$\frac{\sum_{r=1}^{s} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}} = \frac{u_1 y_{1j} + u_2 y_{2j} + \ldots + u_s y_{sj}}{v_1 x_{1j} + v_2 x_{2j} + \ldots + v_m x_{mj}} \leq 1, j = 1, \ldots, n \qquad (1)$$

where x_{ij} and y_{rj} are observations on the inputs and outputs, respectively, and v_i and u_r are unknown (non-negative) weights to be determined.

While in the DEA literature, the maximum value of the above model (DEA score) is often called "efficiency", in fact, the above model generates a composite index or measure, and "efficiency" does not necessarily mean "production efficiency" in many DEA applications. "Efficiency" is a standard terminology in DEA to represent the optimal value to the DEA model. Depending on the specific application, the DEA score can be a risk index, or a quality index, for example.

2.1. Returns to scale (RTS)

One of the best things happened to DEA is the linkage between DEA models to their economic meaning and foundations. This enables DEA to be used as a production function estimator. As a result, DEA models are often called by their frontier types, e.g., CRS or VRS (variable returns to scale). However, when DEA is not used to identify the production frontiers, RTS loses its economic meaning and merely indicates the shape of the best-practice frontier. For example, VRS simply means that the DEA model produces a tighter envelopment of the data than the CRS model does. It is well-known that VRS yields a better DEA score. However, such a conclusion may not be valid under the network DEA. Therefore, it is important to bear in mind that RTS only represents the shape of the best-practice frontier when DEA is not used to identify production functions.

2.2. Convexity (and ratio data)

It can be seen that the above (ratio) model (1) can include only inputs or only outputs. Therefore, the above model is not necessarily a model of "production" or "technology" in economics. Obviously, we can use ratio data (or mix of ratio and raw data) to define a new composite measure. Note that such a composite measure may not bear any economic meaning.

The requirement of convexity in DEA is related to production function or technology in economics. To see this, let us convert the ratio model (1) into the following linear envelopment DEA model

$$\theta^* = min\,\theta$$

subject to

$$\sum_{j=1}^{n} \lambda_j x_{ij} \leq \theta x_{io} \qquad i = 1,2,\ldots,m;$$

$$\sum_{j=1}^{n} \lambda_j y_{rj} \geq y_{ro} \qquad r = 1,2,\ldots,s$$

$$\lambda_j \geq 0 \qquad j = 1,2,\ldots,n. \tag{2}$$

In the above envelopment model, researchers discovered the "convexity" and established a link between DEA and the production function. Therefore, the economic meaning is justified *if* DEA is used as a tool for estimating production functions. As Olesen, Petersen, and Podinovski (2015) correctly point out, the use of the multiplier model along with ratio data is fine as long as we do not use DEA to estimate production functions.

2.3. A brief survey on DEA in the past 10 years

In the past 10 years or so, there are about 900 publications related to DEA. A significant amount of research is dedicated to (i) using DEA as a data-driven tool for descriptive analytics by gaining insight from historical data, and for prescriptive analytics by recommending decisions using DEA-based optimization and simulation, (ii) developing DEA models for studying network structures, and (iii) combining DEA with other data analysis tools.

DEA has been used as a predictive analytics tool that assists medical professionals to accurately predict best donor/recipient pairings in organ sharing programs. Misiunas et al. (2016) combine artificial neural networks (ANN) and DEA to develop a tool that results in accurate predictions and faster training time. Note that in their study, over 400 variables and 100,000 observations exist in the United Network for Organ Sharing database. DEA plays a critical role in training the ANN and its accuracy in predication.

In the model development area, some noticeable contribution lies in the area of network DEA (e.g., Chen, 2009; Cook, Liang, and Zhu, 2010; Tone and Tsutsui, 2014; and Kao, 2014), hierarchical DEA Models (e.g., Kao, 2015), and non-homogeneous DEA models (e.g., Li et al., 2016). In the next section, I will demonstrate that these network DEA models provide opportunities for DEA to be applied under the concept of big data. Note that the following three studies have already used network DEA and dynamic DEA under big data context. A double frontier network DEA approach is used by Badiezadeh, Saen, and Samavati (2018) to study the sustainability of supply chains. A dynamic DEA is used in evaluating the performance of power grid enterprises by Sun et al. (2017) and supply chains by Kahi et al. (2017).

Another noticeable area is the combined use of DEA with other data analytics tools. For example, Lahdelma and Salminen (2006) introduce a method combining DEA with stochastic multicriteria acceptability analysis (SMAA) so that DEA can handle uncertain or imprecise data to provide stochastic efficiency measures. Afsharian (2019) incorporates DEA into a location analysis where facilities are managed by a central authority who wishes to improve the efficiency of the whole system rather than maximizing the individual ones. In a real-life case study of first-tier automotive supplier, Ihrig et al. (2019) combine DEA and a resource allocation technique in setting productivity targets.

Other new DEA-related research has also been developed in the area of productivity and benchmarking (see, e.g., Aparicio et al., 2017 and Cook et al., 2019). It is worthwhile to point out that Kuo and Kusiak (2019) show that production research enabled by data has shifted from analytical models to data-driven, and manufacturing and DEA have been the most popular application areas of data-driven methodologies.

In addition to the big data algorithms provided in Khezrimotlagh et al. (2019), Zhu et al. (2018) provide a hierarchical decomposition algorithm, and Chu, Wu, and Song (2018) develop procedures for environmental efficiency evaluation when the number of DMUs is massive.

A topic search was conducted using the "advanced search" function on the Web of Science (WoS) database. A combination of keywords "big data" and "data envelopment analysis" yielded 29 studies. In addition, the combination of keywords "big data" and "DEA" yielded 30 studies. After compiling these results and excluding duplicates, the final number of complete studies totaled 38. The citation indexes in which these studies are covered include: Science Citation Index Expanded (SCI-EXPANDED), Social Sciences Citation Index (SSCI), Arts & Humanities Citation Index (A&HCI), Conference Proceedings Citation Index- Science (CPCI-S), Conference Proceedings Citation Index- Social Science & Humanities (CPCI-SSH), and Emerging Sources Citation Index (ESCI). The timespan of the search is from January 1970 to June 2019. A cleaning process was designated to remove papers that were present in the initial literature collection by means of the WoS topic search but are irrelevant to applying DEA in the context of big data, and this process has reduced our number of papers from 38 to 23.

A significant number of studies are carried out for environmental issues under the context of big data. Wu, Chen, and Xia (2018)) propose a DEA-based dynamic environmental performance evaluation model using real time big data. DEA is used to evaluate environmental efficiency of China's regional industry by Chen and Jia (2017). Liu, Chu, Yin, and Sun (2017) use a cross efficiency DEA to study the eco-efficiency of coal-fired power plants in a big data environment. An et al. (2017) and Ji et al. (2017) set the carbon dioxide emission permits for each DMU. See Song et al. (2017, 2018) for a survey on environmental performance evaluation with big data.

The use of DEA for big data has also been adopted in supply chain performance evaluations. See, e.g., Badiezadeh, Saen, and Samavati (2018), Song and Wang (2017), and Herranz et al. (2017). Other applications include China's forestry resources efficiency (Li, Hao, and Chi (2017), production performance in iron and steel enterprises (Gong et al. (2017), regional energy efficiency and resource allocation (Zhang et al (2017); Zhu et al. (2017)), transportation management (Chen et al. (2019)), and disaster recovery systemic innovations (Yang et al. (2015).

The above literature analysis indicates that DEA has evolved into a tool for big data analysis and a significant body of DEA research has focused on network DEA. In the section below, I will demonstrate how network DEA is related to the "value" dimension of big data.

3. Data Enabled Analytics and Network DEA

I think the discussion on what big data represents is still on-going. Often 3Vs (volume, variety and velocity) are called the three defining properties of big data. Under the context of DEA, the number of DMUs becomes the "volume". Special algorithms are needed in order to process large amounts of DMUs in a short period of time. This is then related to the "velocity" aspect of big data. Fortunately, researchers in the DEA field have already begun to develop effective algorithms. Most recently, Khezrimotlagh et al. (2019) develop algorithms to handle large volumes of data (decision making units, inputs, and outputs) under conventional DEA.

The "variety" dimension is reflected by the non-homogenous DMUs and different types of performance measures (or called inputs and outputs in DEA). While it is usually assumed in conventional DEA that DMUs under consideration must be homogenous, Cook et al (2013) show that DEA can be adopted for modeling non-homogenous DMUs. The issue of different types of performance measures boils down to whether a DEA model can deal with ordinal or scale data or mixtures of real data and scale data (see, e.g., Zhu (2003). The "variety" aspects can also lead to large amounts of performance measures being used by DEA. Charles, Aparicio and Zhu (2019) develop simple techniques to reduce the number of DEA inputs and outputs. In the previous section, I have discussed that the types of the performance measures and frontiers are not limited to the situation when DEA is used to estimate production functions.

It is now well known that "value" is another important dimension of big data and it sits at the top of the big data pyramid. The "value" aspect refers to the ability of transferring data into useful information. While conventional DEA already generates useful insights in helping business to improve their performance, for example, not all the operations and performance measures can be addressed in a single conventional DEA model. Fortunately, in recent years, we have seen a significant development on network DEA where various (operational) processes are linked by a variety of performance measures.

It is very straightforward that one can think of the inclusion of large quantities of DMUs when DEA is applied. However, a large number of DMUs in itself may not entirely reflect the big data concept. For example, a bank can only have a limited number of branches. A DEA analysis of all the bank branches would not characterize the information embedded in the big data. Of course, one can include more performance metrics. However, such an action may weaken the discriminatory power of DEA.

In this section, I will focus on how DEA, in particular, network DEA, can be used to deal with the "value" dimension of big data. Note that DEA or network DEA is a "ratio" based analysis". Certainly, "ratio" analysis in general can be applied to big data. In contrast to the conventional DEA, this paper emphasizes that network DEA can extract "value" from big data that are presented in network structures, such as supply chain and logistics and transportation systems. Covering all these "Vs" in a one paper is of course infeasible. Therefore, in the current paper, we only emphasize some aspects of the "Vs".

Using airline operations as an example, a variety of data or performance metrics are available to be analyzed. A mixture of data, e.g., airline capacity and marketing data, in the traditional DEA may not clearly characterize the benchmarking purpose. For example, Figure 1 illustrates a simple process depicting the airline operations where the "capacity" component determines the fleet sizes to generate revenues in the "operation" stage. The fleet sizes can reflect such measures as load factor defined as the percentage of available seats filled, available seat-miles for passenger transport segment and available tone-miles for the freight transport segment. (see, e.g., Zhu (2011) and Kottas and Madas (2018).)

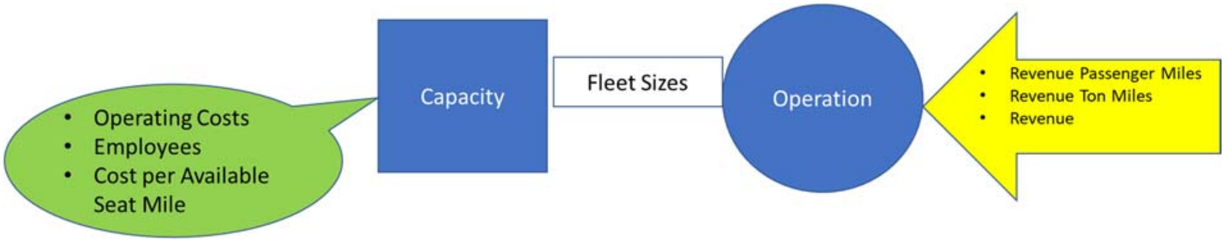

Figure 1. Airline Performance

A conventional DEA model can evaluate the performance of the capacity or operation stage. However, if we combine the performance metrics, whether the measures related to fleet sizes should be used as inputs or outputs is not clear. In fact, these performance metrics may represent "coordination" inbetween the two stages. For example, in a supplier and buyer supply chain, the "optimal" values of measures that link the two members are often determined via coordination between the supplier and the buyer.

While Figure 1 depicts a very simple two-stage operation, air carriers can be classified as combination carriers, fright-only carriers, and integrators (see, e.g., Kottas and Madas (2018)). The network shown in Figure 1 needs to be modified to reflect the three air transportation classifications and results in a more complicated network structure to better characterize the airline performance. In fact, Gan et al. (2019) present a hierarchical network structure related to an international shipping company in Taiwan.

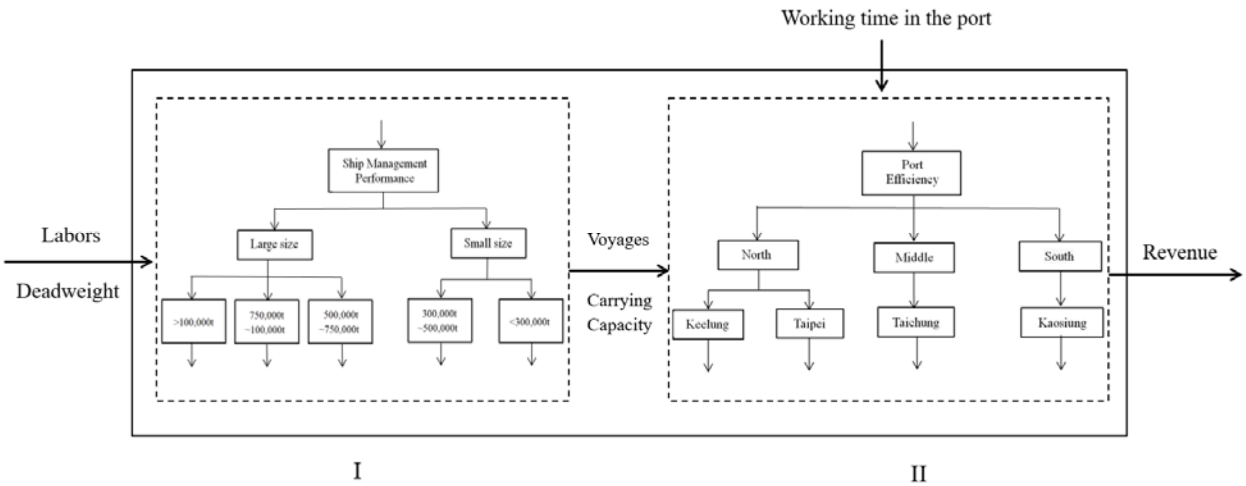

Figure 2. Taiwan’s international shipping company

Figure 2 shows the hierarchical network structure of this Taiwanese international shipping company. It is a two-stage network consisting of ship management stage and port management stage. Each stage has a hierarchical structure due to the sizes of the ships and capacity of the ports. Other network structures can be found in Cook et al. (1998), Cook and Green (2005), and Kao (1998, 2009, 2015), for example.

If one examines the big data technique literature, one would find the usual techniques such as classification techniques (e.g., support vector machine), (deep) neural networks, clustering, and hierarchical learning. While standard DEA can be regarded as a “classification” or “clustering” technique, network DEA can provide additional insights or “value” if the evaluation or benchmarking issue itself needs to be characterized by multiple aspects or dimensions. For example, in a recent article by Summerfield et al. (2019), network DEA is used to study whether drivers should cooperate on simulated road networks. This is an important and valuable piece of information to the transportation department.

Since transportation and logistics systems naturally consist of network systems, let us take a look at the DEA applications in this particular field. I should point out that while network DEA is developed for studying the internal structures of DMUs, its underlying applications are from supply chains, multi-stage production systems, and transportation systems (see, e.g., Liang et al. (2006) and Tone and Tsutsui, (2009)).

In the past 20 years or so, there are more than 600 published papers using DEA in transportation and logistics system. The majority of them are using the basic DEA models. Given the recent development in network DEA, a few studies have used the simple network DEA models. Therefore, let us focus on the period of 2014-2019.

The last column of Table 1 shows the total number of papers published in each year. The second and third columns show the number of papers that use conventional and network DEA, respectively. It can be seen that the majority the papers use conventional DEA and about 15% of the publications use the network DEA technique in each year (except for a 31% in 2016). This could indicate that the recent network DEA development takes time to be adopted.

Table 2 provides detailed information on top 10 journals in which most of the DEA papers published. The majority of the application areas are air transportation, followed by sea transportation, and road, as indicated by Figure 3.

Table A.1 in Appendix A lists the measures used in various network DEA models for studying air transportation. It can be seen that the usual inputs are the number of employees, operating expenses, salaries and wages, materials cost, fleet sizes, and fuel costs. Typical intermediate measures include available seat kilometers (or miles), and revenue passenger kilometers. The outputs from the second stage usually include revenues. Depending on the goals of particular studies, some studies treat fleet size as an exogenous input which does not come from a previous stage, for example. Tables A.2 and A.3 list the measures in sea transportations and supply chain, respectively.

Table 1. DEA Techniques used in publications

Year	conventional	network	Total
2014	26	5	31
2015	22	4	26
2016	27	12	39
2017	30	6	36
2018	30	6	36
2019	11	2	13

Table 2. Journals

Journal	201 9	201 8	201 7	201 6	201 5	201 4	TOTA L
Journal of Air Transport Management	1	5	5	7	5	6	29
Transportation Research Part A: Policy and Practice	2	6	6	4	4	5	27
Transport Policy	3	4	5	6	2	1	21
Transportation Research Part D: Transport and Environment		3	4	8	2	4	21
Maritime Economics & Logistics	1	5	4		1	4	15
Maritime Policy & Management	1	2	2	4	3		12
Transportation Research Part E: Logistics and Transportation Review	1		1	3	4		9
Journal of Advanced Transportation		2		2		2	6
Journal of Transport Geography		1	2	1		2	6

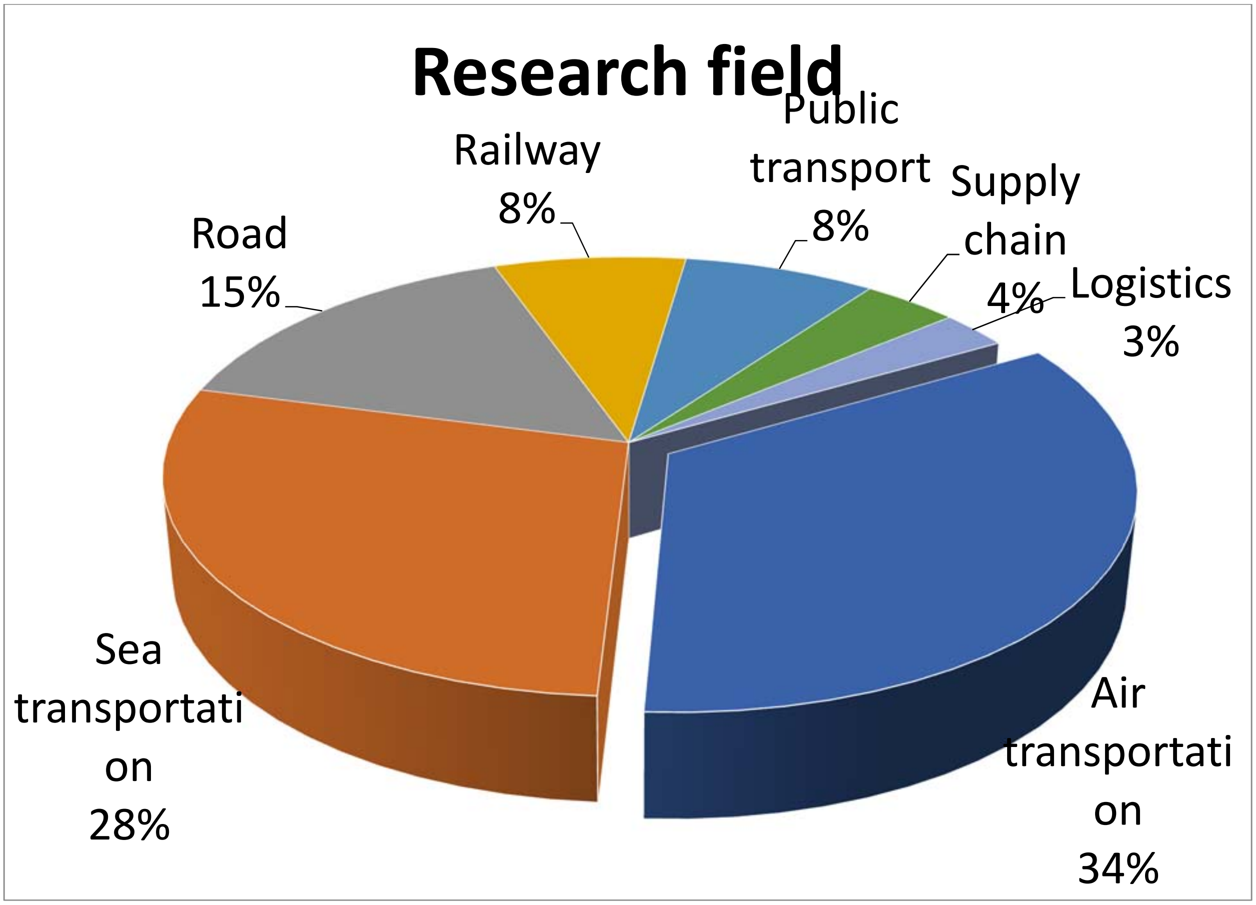

Figure 3. Research Field Breakdown

Finally, to demonstrate the network DEA model, let us consider a general two-stage network structure as shown in Figure 4. Each DMU$_j$ (j=1,2,…,n) has m inputs $x_{ij} \geq 0$, (i=1,2,…,m) to the first stage and P outputs $y^1_{pj} \geq 0$ (p=1,2,…,P) that leave the system. In addition to these P outputs, stage 1 has D intermediate outputs $z_{dj} \geq 0$(d=1,2,…,D) that become inputs to the second stage. The second stage has as well, its own inputs $x^2_{hj} \geq 0$(h=1,2,…,H) that enter from outside the system. The outputs from the second stage are $y_{rj} \geq 0$ (r=1, 2, …,s). Note in DEA it is assumed that all observations on these performance measures are non-negative.

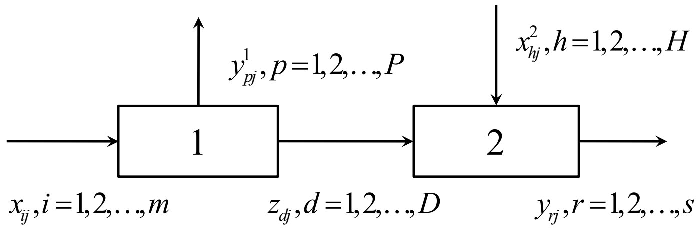

Figure 4. General two-stage network structure

Based upon the conventional DEA, the (performance) ratios or (indexes) of stages 1 and 2 for a specific DMU_o under evaluation can be expressed as:

$$e_o^1 = \frac{\sum_{d=1}^{D} \eta_d z_{do} + \sum_{p=1}^{P} \lambda_p y_{p0}^1}{\sum_{i=1}^{m} v_i x_{io}} \quad \text{and} \quad e_0^2 = \frac{\sum_{r=1}^{s} u_r y_{r0}}{\sum_{d=1}^{D} \eta_d z_{d0} + \sum_{h=1}^{H} Q_h x_{h0}^2}$$

where v_i, η_d, λ_p, u_r, and Q_h are weights which are assumed to be positive in the current study, by incorporating the small non-Archimedean ε into the DEA models. Note that the weights on the intermediate measures are assumed to be the same for stages 1 and 2, as in Kao and Hwang (2008) and Liang, Cook, and Zhu (2008). This is an important assumption that establishes a linkage between the two stages.

Under (weighted) additive efficiency aggregation, we can define the overall performance index as $e_o^1 + e_o^2 = \alpha \frac{\sum_{d=1}^{D} \eta_d z_{d0} + \sum_{p=1}^{P} \lambda_p y_{p0}^1}{\sum_{i=1}^{m} v_i x_{i0}} + (1-\alpha) \frac{\sum_{r=1}^{s} u_r y_{r0}}{\sum_{d=1}^{D} \eta_d z_{d0} + \sum_{h=1}^{H} Q_h x_{h0}^2}$, which is a nonconvex function where α is a predetermined weight satisfying $0 \le \alpha \le 1$.

The corresponding network DEA model can be expressed as:

$$\begin{aligned}
\max \quad & \alpha \frac{\sum_{d=1}^{D} \eta_d z_{d0} + \sum_{p=1}^{P} \lambda_p y_{p0}^1}{\sum_{i=1}^{m} v_i x_{i0}} + (1-\alpha) \frac{\sum_{r=1}^{s} u_r y_{r0}}{\sum_{d=1}^{D} \eta_d z_{d0} + \sum_{h=1}^{H} Q_h x_{h0}^2} \\
s.t. \quad & \frac{\sum_{d=1}^{D} \eta_d z_{dj} + \sum_{p=1}^{P} \lambda_p y_{pj}^1}{\sum_{i=1}^{m} v_i x_{ij}} \le 1, \quad \forall j \\
& \frac{\sum_{r=1}^{s} u_r y_{rj}}{\sum_{d=1}^{D} \eta_d z_{dj} + \sum_{h=1}^{H} Q_h x_{hj}^2} \le 1, \quad \forall j \\
& \eta_d, u_r, v_i, \lambda_p, Q_h \ge \varepsilon, \quad \forall d, r, i, p, h
\end{aligned} \qquad (3)$$

Note that although model (3) is constructed by using the DEA ratio similar to that in model (1), the resulting model (3) is a nonlinear and nonconvex optimization problem that cannot be converted into a linear model. Consequently, we do not have an equivalent of dual linear (envelopment) model. One has to build the "envelopment" network DEA model that is not related to the multiplier form (3).

While one can set a specific α so that model (3) can be converted into a linear program, such a technique introduces weight restrictions into the model (3) (see, e.g., Cook et al., 2010). Chen and Zhu (2017) and Chen, Cook, and Zhu (2009) develop second order cone programming (SOCP) and conic relaxation model to solve non-linear network DEA models. It generates, in a more convenient manner, feasible approximations and tighter upper bounds on the global optimal solution. Compared with a line-parameter search method that has been applied to solve non-linear network DEA models, the conic relaxation model keeps track of the distances between the optimal overall efficiency and its approximations. As a result, it is able to determine whether a qualified approximation has been achieved or not, with the help of a branch and bound algorithm.

Given the nonlinearity of the network DEA models, network DEA is already significantly different from conventional DEA from the computational perspective. This offers opportunities for the DEA community to develop and/or apply optimization techniques in solving these network DEA models. Both the SOCP of Chen and Zhu (2020) and Chen, Cook and Zhu (2020) and semi-definiteness programming of Halická and Trnovská (2018) are two possible useful tools for solving non-linear network DEA models and big data modeling under DEA. In my personal view, big data can be reflected in the (complex) network structures. As such, this offers both challenges and opportunities in applying network DEA to big data analysis.

Because there is no dual model to (3), a different line of network DEA research on envelopment form has been developed. Such envelopment-based models are based upon the production possibility set. See, e.g., Färe and Grosskopf (2000) and Tone and Tsutsui (2009).

Färe and Grosskopf (2000) suggest that the production possibility set (PPS) of network system is the aggregation of PPS of individual divisions. Thus, based upon Tone and Tsutsui (2009) and Kao (2018), the PPS of the general two-stage network shown in Figure 3 can be defined as follows:

$$T = \left\{ \left(x, x^2, y, y^1, z\right) \middle| \begin{array}{l} \sum_{j=1}^{n} x_{ij}\lambda_j^1 \le x_i, \sum_{j=1}^{n} x_{hj}^2\lambda_j^2 \le x_h, \forall h, \sum_{j=1}^{n} y_{rj}\lambda_j^2 \ge y_r, \forall r, \\ \sum_{j=1}^{n} y_{pj}^1\lambda_j^1 \ge y_p, \forall p, \sum_{j=1}^{n} z_{dj}\lambda_j^1 \ge z_d, \forall d, \sum_{j=1}^{n} z_{dj}\lambda_j^2 \le z_d, \forall d, \\ \sum_{j=1}^{n} \lambda_j^1 = 1, \sum_{j=1}^{n} \lambda_j^2 = 1 \end{array} \right\} \quad (4)$$

where x_{ij}, x_{hj}^2, y_{pj}^1, and y_{rj} are exogenous variables which are visible to outsiders. Then, based on the PPS (4), we have the following after slacks are introduced:

$$\begin{array}{l} \sum_{j=1}^{n} x_{ij}\lambda_j^1 + s_i^- = x_{i0}, \forall i \\ \sum_{j=1}^{n} x_{hj}^2\lambda_j^2 + s_h^- = x_{h0}^2, \forall h \\ \sum_{j=1}^{n} y_{rj}\lambda_j^2 - s_r^+ = y_{r0}, \forall r \\ \sum_{j=1}^{n} y_{pj}^1\lambda_j^1 - s_p^+ = y_{p0}^1, \forall p \\ \sum_{j=1}^{n} \lambda_j^1 = 1, \sum_{j=1}^{n} \lambda_j^2 = 1, \\ s_i^-, s_h^-, s_r^+, s_p^+, \lambda_j^1, \lambda_j^2 \ge 0 \end{array} \quad (5)$$

Given that z_{dj} are intermediate measures that link the two stages, we assume here that (Kao, 2018):

$$\sum_{j=1}^{n} \lambda_j^1 z_{dj} = \sum_{j=1}^{n} \lambda_j^2 z_{dj}, \quad \forall d \quad (6)$$

Chen and Zhu (2020) develop the following envelopment form of the network DEA model:

$$\begin{array}{ll} \min & \frac{1}{S+P+M+H}\left(\sum_{r=1}^{S} \frac{y_{r0}}{y_{r0}+s_r^+} + \sum_{p=1}^{P} \frac{y_{p0}^1}{y_{p0}^1+s_p^+} + \sum_{i=1}^{M} \frac{x_{i0}-s_i^-}{x_{i0}} + \sum_{h=1}^{H} \frac{x_{h0}^2-s_h^-}{x_{h0}^2}\right) \\ s.t & \text{Constraint Sets (5) and (6)} \end{array} \quad (7)$$

Unlike the envelopment models in Tone and Tsutsui (2009), model (7) is a non-linear model that can be solved via SOCP technique (see Chen and Zhu, 2020).

I should point out that in the existing DEA literature, proofs have never been provided that a model like (7), for example, actually yields the overall and divisional scores. In fact, Chen et al. (2013) point out that the overall and divisional scores generated by the multiplier and envelopment network DEA models do not correspond to each other. While the envelopment model generates frontier projection points for inefficient units, the multiplier model is needed for overall and divisional scores. Interested readers are referred to Chen et al. (2013) for a list of network DEA pitfalls.

Note that the assumption of VRS (or the VRS shape of the frontier) is reflected on the convexity constraints of $\sum_{j=1}^{n}\lambda_j^1 = 1$ and $\sum_{j=1}^{n}\lambda_j^2 = 1$. In other words, if we impose such a convexity constraint in the envelopment form, we assume VRS. In the multiplier form, VRS is reflected by a free variable which represents the y-intercept, depending on whether the optimal value of the free variable is positive, negative, or zero. Note also that when the network DEA models are not linear, there is no duality relationship between the convexity constraint and the free variable. As a result, whether the convexity condition assumes VRS shape of the frontier needs to be further studied.

In fact, as pointed out by Lim and Zhu (2019), the overall network DEA score under VRS is not smaller than that under CRS for all DMUs as is the case in the conventional DEA. However, some individual component scores under VRS are found to be smaller than the corresponding score under CRS, unlike the conventional DEA.

In the conventional envelopment DEA, it is obvious that VRS scores are always greater than CRS scores due to the additional convexity constraint in the VRS model. The same holds true with the overall network DEA scores. The problematic situation, where the VRS scores are smaller than the CRS scores, happens only with divisional scores. This discovery indicates that network DEA cannot be viewed as a (simple) extension to the conventional DEA, although the network DEA model is based upon the ratios in the multiplier form or the PPS in the envelopment form. While the overall index in network DEA is built upon the assumption of VRS or CRS shape of the frontier, its divisional efficiency may not obey the CRS or VRS assumption. This is due to the lack of duality between the network DEA multiplier and envelopment models and the treatment of the intermediate measures that link the network components.

Finally, note that in conventional DEA, there is always at least one DMU that is efficient or on the best-practice frontier. However, it is possible that none of the DMUs is overall efficient in network DEA.

4. Conclusions

The goal of this paper is to explore the idea on how network DEA can be used in big data research. The focus is the value aspect of the big data reflected in network structures. Given the existing examples in transportation and logistics systems and other areas, the need for using network DEA to gain valuable information in data analytics is obvious. I demonstrate that network DEA can be different from conventional DEA in many aspects. In particular, techniques in solving non-linear programming problems will be very useful in network DEA computations. While there exist simple network DEA structures, consequently, one is able to convert the related network DEA models into linear programs. However, the dual to the linear multiplier network DEA does not resemble the envelopment DEA network models. This is obviously a topic for future research in network DEA when we study the multiplier and envelopment-based models. In general, we expect that the non-linear optimization techniques need to be developed for solving network DEA models under general network structures.

Research built upon conventional DEA is also extremely important for big data research. Misiunas et al. (2016) is one example where basic conventional DEA can be used to assist decision making under big data. While Khezrimotlagh et al. (2019) offer algorithms to deal with large value of DEA data, Charles, Aparicio and Zhu (2019) develop simple techniques to reduce the number of DEA inputs and outputs.

Under big data, any methodology has its limitation with respect to real time update. However, under DEA, if a new DMU (or a group of new DMUs) appears, one does not have to run the entire big data set. In DEA or network DEA, we only need to compare the new DMU(s) to the established or identified frontier. This is a much smaller data set and can be calculated quickly. Note that one challenge for DEA under big data is the quick identification of DMUs that are on the frontier. While the above methods can effectively address such a challenge, we need to look at the possibility of combining DEA or network DEA with typical data mining and machine learning techniques, for example random forest, support vector machine, and artificial neural networks, to expediate the process for identification of frontier DMUs. This is an important future research.

From the very first DEA paper (Charnes et al., 1978), it is clear that DEA is a data-oriented technique. While conventional DEA is linear program based, network DEA can remain as a non-linear and non-convex model. As a data-oriented technique, data will enable network DEA play important roles in big data related researches. In addition to the top DEA application areas, such as, banking, health care, transportation, education, and agriculture, recent years have seen a significant amount of applications in environmental issues and sustainability research. While many of these applications are based upon conventional DEA, environmental and sustainability issues are by nature multifaceted that need to be categorized by social, environmental, and financial performances. Therefore, there is an opportunity to revisit these areas by network DEA.

References

1. Afsharian, M., (2019), A frontier-based facility location problem with a centralized view of measuring the performance of the network, Journal of the Operational Research Society, in press.
2. An, Q. X., We, Y., Xiong, B. B., Yang, M., & Chen, X. H. (2017). Allocation of carbon dioxide emission permits with the minimum cost for Chinese provinces in big data environment. Journal of Cleaner Production, 142, 886-893.
3. Aparicio, J., Pastor, J.T., Vidal, F., and Zofío, J.L. (2017), Evaluating productive performance: A new approach based on the product-mix problem consistent with Data Envelopment Analysis, OMEGA, 67, 134-144.
4. Badiezadeh, T., Saen, R. F., & Samavati, T. (2018). Assessing sustainability of supply chains by double frontier network DEA: A big data approach. Computers & Operations Research, 98, 284-290.
5. Charles, V., Aparicio, J., and Zhu, J. (2019), The curse of dimensionality of decision-making units: A simple approach to increase the discriminatory power of data envelopment analysis, European Journal of Operational Research, Vol. 279, Issue 3, 929-940.
6. Charles, V., Aparicio, J., and Zhu, J. (2020a), Data Science for Better Productivity, Journal of the Operational Research Society,
7. Charles, V., Aparicio, J., and Zhu, J. (2020b), Preface, in Data Science and Productivity Analytics, Springer Nature, New York.
8. Charnes A., Cooper, W.W., & Rhodes, E. (1978). Measuring the inefficiency of decision making units. European Journal of Operational Research 2 (6) 429-444.
9. Chen, C-M. (2009). A network-DEA model with new efficiency measures to incorporate the dynamic effect in production networks. European Journal of Operational Research, 194, 687–699.
10. Chen, C. L., Achtari, G., Majkut, K., & Sheu, J. B. (2017). Balancing equity and cost in rural transportation management with multi-objective utility analysis and data envelopment analysis: A case of Quinte West. Transportation Research Part a-Policy and Practice, 95, 148-165.

11. Chen, Y., Cook, W.D., Kao, C., & Zhu, J. (2013), Network DEA pitfalls: Divisional efficiency and frontier projection under general network structures, *European Journal of Operational Research*, 226(3), pp. 507-515.
12. Chen, Y., Cook, W.D., & Lim, S. (2019), Preface: DEA and its applications in operations and data analytics, Annals of Operations Research, Vol. 278, Issue 1-2, pp. 1-4.
13. Chen, K., Cook, W.D., and Zhu, J. (2020), A conic relaxation model for searching global optimum of network data envelopment analysis, *European Journal of Operational Research*, Vol. 280, Issue 1, 242-253.
14. Chen, L., & Jia, G. Z. (2017). Environmental efficiency analysis of China's regional industry: a data envelopment analysis (DEA) based approach. Journal of Cleaner Production, 142, 846-853.
15. Chu, J. F., Wu, J., & Song, M. L. (2018). An SBM-DEA model with parallel computing design for environmental efficiency evaluation in the big data context: a transportation system application. Annals of Operations Research, 270(1-2), 105-124.
16. Chen, P-C, Yu, M-M, Shih, J-C, Chang, C-C, and Hsu, S-H, (2019), A reassessment of the Global Food Security Index by using a hierarchical data envelopment analysis approach, European Journal of Operational Research 272, 687–698
17. Chen, K., and Zhu, J., (2017), Second order cone programming approach to two-stage network data envelopment analysis. European Journal of Operational Research 262, 231–238.
18. Chen, K., and Zhu, J. (2020). Additive slacks-based measure: computational strategy and extension to network DEA. Omega. OMEGA, Vol. 91, 102022
19. Cook, W.D., Zhu, J., Bi, G-B, and Yang, F., (2010), Network DEA: Additive efficiency decomposition, European Journal of Operational Research, Vol. 207, Issue 2, 1122-1129.
20. Cook, W. D., Chai, D., Doyle, J., & Green, R. (1998). Hierarchies and groups in DEA. Journal of Productivity Analysis, 10(2), 177-198.
21. Cook, W. D., & Green, R. H. (2005). Evaluating power plant efficiency: a hierarchical model. Computers & Operations Research, 32(4), 813-823.
22. Cook, W.D., Harrison, J., Imanirad, R., Rouse, P., and Zhu, J. (2013). Data envelopment analysis with non-homogeneous DMUs, Operations Research, Vol. 61, No. 3, 666-676.

23. Cook, W.D., Liang, L., and Zhu, J. (2010), Measuring performance of two-stage network structures by DEA: A review and future perspective. OMEGA 38, 423–430
24. Cook, W.D., Ramón, N., Ruiz, J.L., Sirvent, I. and Zhu, J., (2019), DEA-based benchmarking for performance evaluation in pay-for-performance incentive plans. OMEGA 84,45–54
25. Cook, W.D., and Seiford, L.M., (2009), Data envelopment analysis (DEA) – Thirty years on, European Journal of Operational Research, Vol. 192, Issue 1, 1-17.
26. Cook, W.D., Tone, K., and Zhu, J. (2014), Data envelopment analysis: Prior to choosing a model, OMEGA, Vol. 44, 1-4.
27. Färe, R., and Grosskopf, S., (2000), Network DEA. Socio-Economic Planning Sciences 34, 35-49
28. Gan, G., Lee, H-S, Lee, L., Wang, X., and Wang, Q, (2019), Network hierarchical DEA with an application to international shipping industry in Taiwan, Journal of the Operational Research Society, in press.
29. Gong, B. G., Guo, D. D., Zhang, X. Q., & Cheng, J. S. (2017). An approach for evaluating cleaner production performance in iron and steel enterprises involving competitive relationships. Journal of Cleaner Production, 142, 739-748.
30. Halická, M. and Trnovská, M. (2018). The Russell measure model: Computational aspects, duality, and profit efficiency, European Journal of Operational Research 268, 386-397.
31. Herranz, R. E., Estevez, P. G., Oliva, M., & De, R. (2017). LEVERAGING FINANCIAL MANAGEMENT PERFORMANCE OF THE SPANISH AEROSPACE MANUFACTURING VALUE CHAIN. Journal of Business Economics and Management, 18(5), 1005-1022.
32. Ihrig, S., Ishizaka, A., Brech, C., and Fliedner, T., (2019), A new hybrid method for the fair assignment of productivity targets to indirect corporate processes, Journal of the Operational Research Society, in press.
33. Ji, X., Sun, J. S., Wang, Y. Y., & Yuan, Q. Q. (2017). Allocation of emission permits in large data sets: a robust multi-criteria approach. Journal of Cleaner Production, 142, 894-906.

34. Kahi, V. S., Yousefi, S., Shabanpour, H., & Saen, R. F. (2017). How to evaluate sustainability of supply chains? A dynamic network DEA approach. Industrial Management & Data Systems, 117(9), 1866-1889
35. Kao, C. (1998). Measuring the efficiency of forest districts with multiple working circles. Journal of the Operational Research Society, 49(6), 583-590.
36. Kao, C. (2009). Efficiency decomposition in network data envelopment analysis: A relational model. European Journal of Operational Research, 192(3), 949-962.
37. Kao, C. (2014). Network data envelopment analysis: A review, European Journal of Operational Research 239, 1–16.
38. Kao, C. (2015). Efficiency measurement for hierarchical network systems. Omega, 51, 121-127.
39. Khezrimotlagh, D., Zhu, J., Cook, W., & Toloo, M. (2019). Data Envelopment Analysis and Big Data, European Journal of Operational Research, Vol. 274, Issue 3, 1047-1054.
40. Kottas, A. T., and Madas, M. A. (2018), Comparative efficiency analysis of major international airlines using Data Envelopment Analysis: Exploring effects of alliance membership and other operational efficiency determinants, Journal of Air Transport Management, Vol. 70, 1-17.
41. Kuo, Y-H and Kusiak, A. (2019), From data to big data in production research: the past and future trends, International Journal of Production Research, in press.
42. Lahdelma R., Salminen, K., (2006), Stochastic multicriteria acceptability analysis using the data envelopment model. European Journal of Operational Research 170, 241–252.
43. Li, L., Hao, T. T., & Chi, T. (2017). Evaluation on China's forestry resources efficiency based on big data. Journal of Cleaner Production, 142, 513-523.
44. Li, W.H, Liang, L., Cook, W.D., and Zhu, J. (2016), DEA models for non-homogeneous DMUs with different input configurations. European Journal of Operational Research 254, 946–956.
45. Liang, L., Yang, F. Cook, W.D. and Zhu, J, DEA models for supply chain efficiency evaluation, Annals of Operations Research, Vol. 145, No. 1 (2006), 35-49.
46. Lim, S. and Zhu, J. (2017), DEA and its Applications in Operations - Part I, INFOR, Volume 55, Issue 3, 159-273.

47. Lim, S. and Zhu, J. (2018), DEA and its Applications in Operations - Part II, INFOR, Volume 56, Issue 3, 265-359.
48. Lim, S., and Zhu, J., (2019), Primal-dual correspondence and frontier projections in two-stage network DEA models, OMEGA, Vol. 83, 236-248.
49. Liu, X. H., Chu, J. F., Yin, P. Z., & Sun, J. S. (2017). DEA cross-efficiency evaluation considering undesirable output and ranking priority: a case study of eco-efficiency analysis of coal-fired power plants. Journal of Cleaner Production, 142, 877-885.
50. Liu, J.S., Lu, L.Y., Lu, W. and Lin, B.J. (2013a), A survey of DEA applications, Omega, 41(5), pp. 893-902.
51. Liu, J.S., Lu, L.Y. and Lu, W. (2016), Research Fronts and Prevailing Applications in Data Envelopment Analysis', in Data Envelopment Analysis. Springer, pp. 543-574.
52. Liu, J.S., Lu, L.Y., Lu, W. and Lin, B.J. (2013b), Data envelopment analysis 1978–2010: A citation-based literature survey, Omega, 41(1), pp. 3-15.
53. Mahajan, J., (1991), A data envelopment analytic model for assessing the relative efficiency of the selling function, European Journal of Operational Research, Volume 53, Issue 2, 189-205.
54. Matthews, K. (2013). Risk management and managerial efficiency in Chinese banks: a network DEA framework. Omega, 41(2), 207-215.
55. Misiunas, N., A., Oztekin, Y. Chen, and K. Chandra, (2016), DEANN: A Healthcare Analytic Methodology of Data Envelopment Analysis and Artificial Neural Networks for the Prediction of Organ Recipient Functional Status, OMEGA, 58 (2016), 46-54.
56. Olesen, O. B., Petersen, N. C., and Podinovski, V. V. (2015). Efficiency analysis with ratio measures. European Journal of Operational Research, 245 (2), 446–462.
57. Seles, B., Jabbour, A., Jabbour, C. J. C., Fiorini, P. D., Mohd-Yusoff, Y., & Thome, A. M. T. (2018). Business opportunities and challenges as the two sides of the climate change: Corporate responses and potential implications for big data management towards a low carbon society. Journal of Cleaner Production, 189, 763-774.
58. Shen, Y., Hermans, E., Brijs, T., Wets, G., and Vanhoof, K., (2012), Road safety risk evaluation and target setting using data envelopment analysis and its extensions, Accident Analysis & Prevention, Volume 48, 430-441.

59. Shwartz, M., Burgess, J.F. Jr, and Zhu, J, (2016), A DEA based composite measure of quality and its associated data uncertainty interval for health care Provider profiling and pay-for-performance, European Journal of Operational Research, Vol. 253, Issue 2, 489-502.
60. Song, M. L., Cen, L., Zheng, Z. X., Fisher, R., Liang, X., Wang, Y. T., & Huisingh, D. (2017). How would big data support societal development and environmental sustainability? Insights and practices. Journal of Cleaner Production, 142, 489-500.
61. Song, M. L., Fisher, R., Wang, J. L., & Cui, L. B. (2018). Environmental performance evaluation with big data: theories and methods. Annals of Operations Research, 270(1-2), 459-472.
62. Song, M. L., & Wang, S. H. (2017). Participation in global value chain and green technology progress: evidence from big data of Chinese enterprises. Environmental Science and Pollution Research, 24(2), 1648-1661.
63. Summerfield, N., Deoka, A., Xu, M., and Zhu, W.W. 2019. Should driver cooperate? Performance evaluation of cooperative navigation on simulated road networks using network DEA, Journal of the Operational Research Society, in press.
64. Sun, Y. X., Yu, X. B., Tan, Z. F., Xu, X. F., & Yan, Q. Y. (2017). Efficiency Evaluation of Operation Analysis Systems Based on Dynamic Data Envelope Analysis Models from a Big Data Perspective. Applied Sciences-Basel, 7(6), 14.
65. Tone, K. and M. Tsutsui, (2009), Network DEA: A slacks-based measure approach, European Journal of Operational Research, Volume 197, Issue 1, 243-252.
66. Tone, K, and Tsutsui, M., (2014), Dynamic DEA with network structure: A slacks-based measure approach. OMEGA 42, 124–131.
67. Wu, Y. M., Chen, Z. X., & Xia, P. P. (2018). An extended DEA-based measurement for eco-efficiency from the viewpoint of limited preparation. Journal of Cleaner Production, 195, 721-733.
68. Yang, C. L., Yuan, B. J. C., Huang, C. Y., & Chang, C. N. (2015). Evaluating the Performance of Disaster Recovery Systemic Innovations by Using the Data Envelopment Analysis. Asia Pacific Journal of Innovation and Entrepreneurship, 9(2), 51-75.

69. Zhang, W., Pan, X. F., Yan, Y. B., & Pan, X. Y. (2017). Convergence analysis of regional energy efficiency in china based on large-dimensional panel data model. Journal of Cleaner Production, 142, 801-808.
70. Zhu, Q. Y., Wu, J., Li, X. C., & Xiong, B. B. (2017). China's regional natural resource allocation and utilization: a DEA-based approach in a big data environment. Journal of Cleaner Production, 142, 809-818.
71. Zhu, Q., Wu, J., & Song, M. (2018). Efficiency evaluation based on data envelopment analysis in the big data context. Computers & Operations Research, 98, 291-300.
72. Zhu, J. (2011), Airlines performance via two-stage network DEA approach, Journal of CENTRUM Cathedra, Vol. 4, Issue 2, 260-269.
73. Zhu, J. (2013). Efficiency evaluation with strong ordinal input and output measures, European Journal of Operational Research, Vol. 146, Issue 3, 477-485.

Appendix A Network DEA Applications in Air Transportation, Sea Transportation, and Supply Chains

Table A.1 Network DEA Measures in Air Transportation Studies

Authors	Original Inputs	Final outputs	Intermediate measures/links	Exogenous inputs	Exogenous outputs
Cui et al. (2018)	• number of employees • tons of aviation kerosene	• Total Revenue、Greenhouse gases emission	• Capital Stock		
Cui and Li (2018)	• Operating Expenses	• Total Revenue	• Available Seat Kilometers (Link 1-2) • Revenue Passenger Kilometers (Link 2-3)	• Fleet Size (stage 2) • Sales Costs (stage 3)	• Greenhouse Gases Emission (stage 2)
Kottas and Madas (2018)	• Number of Employees • Total Operating Costs • Number of Operated Aircraft	• Total Operating Revenue • Revenue assenger-Kilometers (RPKs • Revenue Tonne-Kilometers (RTKs)	• Available Seat-Kilometers (ASKs) • Available Tonne-Kilometers (ATKs)		
Storto (2018)	• soft operating expenditures • labor cost	• aviation revenues • non-aviation revenues	• terminal size、apron size、total area of runways、employees (Link 1-2) • movements、passengers、cargo (Link 2-3)		
Cui and Li (2017)	• Number of employees • Aviation kerosene	• Revenue tonne kilometers • Revenue passenger kilometers • Total revenue	• Capital Stock (Carry-over)		
Cui et al.(2017)	• Operating Expenses	• Total Revenue	• Available Seat Kilometers (Link 1-2) • Revenue Passenger Kilometers (Link 2-3)	• Fleet Size (stage 2) • Sales Costs (stage 3)	• Greenhouse Gases Emission (stage 2)
Liu (2017)	• Runway area	• passengers and cargo	• aircraft movements		

	• Staff costs • Other operating costs	• non-aeronautical revenues			
Yu et al. (2017)	• Size of leased fleet • Labor expenses • Fuel expenses • Other operational expenses	• Revenue passenger kilometers (RPK) • Freight revenue ton kilometers (FRTK)	• ASK、FATK (Link) • Size of self-owned fleet、Waypoints(Carry -over)		
Chang et al. (2016)	• Net asset • Material cost • Labor cost	• Cargo • Enplanemen t	• PFC/AIP determined (Link) • aircraft operations (Carry-over)	• Promotion (stage 2)	• Delay (stage 1)
Chou et al. (2016)	• Labor cost • Fuel cost (million US$) • Fleet size	• Passenger kilometers	• Available seatkilometers (Link) • Net revenue、Number of accidents (Carry-over)		
Cui and Li (2016)	• Salaries, Wages and Benefits , Fuel Expenses and Total Assets	• Carbon Dioxide (CO2)	• Estimated Carbon Dioxide (ECO2)	• Abatement Expense (AE)	• Revenue Passenger Kilometers (RPK), Revenue Tonne Kilometers (RTK)
Cui et al. (2016)	• Number of employees and Aviation Kerosene	• Total Business Income	• Available seat kilometres (ASK) and available tonne kilometres (ATK) (Link 1-2) • Revenue Passenger Kilometres (RPK) and Revenue Tonne Kilometres (RTK)(Link 2-3)	• Fleet Size (stage 2) • Sales Costs (stage 3)	• Greenhous e Gases Emission (stage 2)
Olfat et al. (2016)	• Policy making based on sustainable developmen t concept • Budget	• Satisfaction • Pollutions levels	• The number of takeoff and landing aircraft、Social responsibility、Service quality (Link 1-2) • Corporate reputation (Carry-over)		• Non-aviation income (stage 1)
Omrani and Soltanzade h (2016)	• the number of employees	• passenger-kilometer performed • passenger	• available seat-kilometer, available ton-kilometer, and		

		ton-kilometer perfumed (Y2)	number of scheduled flights • the number of fleet's seat (Carry-over)		
Shao and Sun (2016)	• Flights related cost	• Passenger throughput • Freight throughput	• Available seats、Available tonnage		
Li et al. (2015)	• Number of employees and Aviation Kerosene	• Total Business Income (TBI)	• Available seat kilometres (ASK) and available tonne kilometres (ATK) (Link 1-2) • Revenue Passenger Kilometres (RPK) and Revenue Tonne Kilometres (RTK)(Link 2-3)	• Fleet Size (stage 2) • Sales Costs (stage 3)	
Mallikarjun (2015)	• operating expenses	• operating revenue	• available seat miles (Link 1-2) • revenue passenger miles (Link 2-3)	• fleet Size、destinations (stage 2)	
Chang and Yu (2014)	• labor • capital • fuel consumptio n	• passenger miles	• the number of destinations • Seat miles	• Environmenta l factor: adjusted GDP (stage 1&2)	
Lozano and Gutiérrez (2014)	• fuel cost of airline • non-current assets of airline • wages and salaries of airline • other operating costs of airline	• revenue passenger kilometres of airline • revenue tonne kilometres of airline	• available seat kilometres of airline • available tonne kilometres of airline	• selling costs of airline	
Tavassoli et al. (2014)	• Number of passenger planes • Labor • Number of cargo • planes	• P-plane-km • C-plane-km	• P-km • Ton-km	/	

References for Table A.1:
Chang, Y. T., Park, H. K., Zou, B., & Kafle, N. (2016). Passenger facility charge vs. airport improvement program funds: A dynamic network DEA analysis for US airport financing. Transportation Research Part E: Logistics and Transportation Review, 88, 76-93.
Chang, Y. C., & Yu, M. M. (2014). Measuring production and consumption efficiencies using the slack-based measure network data envelopment analysis approach: the case of low-cost carriers. Journal of Advanced Transportation, 48(1), 15-31.

Chou, H. W., Lee, C. Y., Chen, H. K., & Tsai, M. Y. (2016). Evaluating airlines with slack-based measures and meta-frontiers. Journal of Advanced Transportation, 50(6), 1061-1089.
Cui, Q., & Li, Y. (2016). Airline energy efficiency measures considering carbon abatement: A new strategic framework. Transportation Research Part D: Transport and Environment, 49, 246-258.
Cui, Q., & Li, Y. (2017). Airline efficiency measures using a Dynamic Epsilon-Based Measure model. Transportation Research Part A: Policy and Practice, 100, 121-134.
Cui, Q., & Li, Y. (2018). CNG2020 strategy and airline efficiency: A Network Epsilon-Based Measure with managerial disposability. International Journal of Sustainable Transportation, 12(5), 313-323.
Cui, Q., Li, Y., & Lin, J. L. (2018). Pollution abatement costs change decomposition for airlines: An analysis from a dynamic perspective. Transportation Research Part A: Policy and Practice, 111, 96-107.
Cui, Q., Li, Y., & Wei, Y. M. (2017). Exploring the impacts of EU ETS on the pollution abatement costs of European airlines: An application of Network Environmental Production Function. Transport Policy, 60, 131-142.
Cui, Q., Wei, Y. M., Yu, C. L., & Li, Y. (2016). Measuring the energy efficiency for airlines under the pressure of being included into the EU ETS. Journal of Advanced Transportation, 50(8), 1630-1649.
Kao, C. (2018). A classification of slacks-based efficiency measures in network data envelopment analysis with an analysis of the properties possessed. European Journal of Operational Research, Vol. 270, Issue 3, 1109-1121.
Kottas, A. T., & Madas, M. A. (2018). Comparative efficiency analysis of major international airlines using Data Envelopment Analysis: Exploring effects of alliance membership and other operational efficiency determinants. Journal of Air Transport Management, 70, 1-17.
Liu, D. (2017). Evaluating the multi-period efficiency of East Asia airport companies. Journal of Air Transport Management, 59, 71-82.
Li, Y., Wang, Y. Z., & Cui, Q. (2015). Evaluating airline efficiency: an application of virtual frontier network SBM. Transportation Research Part E: Logistics and Transportation Review, 81, 1-17.
Lozano, S., & Gutiérrez, E. (2014). A slacks-based network DEA efficiency analysis of European airlines. Transportation Planning and Technology, 37(7), 623-637.
Mallikarjun, S. (2015). Efficiency of US airlines: a strategic operating model. Journal of Air Transport Management, 43, 46-56.
Olfat, L., Amiri, M., Soufi, J. B., & Pishdar, M. (2016). A dynamic network efficiency measurement of airports performance considering sustainable development concept: A fuzzy dynamic network-DEA approach. Journal of Air Transport Management, 57, 272-290.
Omrani, H., & Keshavarz, M. (2016). A performance evaluation model for supply chain of shipping company in Iran: an application of the relational network DEA. Maritime Policy & Management, 43(1), 121-135.
Shao, Y., & Sun, C. (2016). Performance evaluation of China's air routes based on network data envelopment analysis approach. Journal of Air Transport Management, 55, 67-75.
Storto, C. L. (2018). The analysis of the cost-revenue production cycle efficiency of the Italian airports: A NSBM DEA approach. Journal of Air Transport Management, 72, 77-85.
Tavassoli, M., Faramarzi, G. R., & Saen, R. F. (2014). Efficiency and effectiveness in airline performance using a SBM-NDEA model in the presence of shared input. Journal of Air Transport Management, 34, 146-153.
Yu, M. M., Chen, L. H., & Chiang, H. (2017). The effects of alliances and size on airlines' dynamic operational performance. Transportation Research Part A: Policy and Practice, 106, 197-214.

Table A.2 Network DEA Measures in Sea Transportation Studies

Authors	Original Inputs	Final outputs	Intermediate measures	Exogenous inputs/links
Chao et al. (2018)	● Chartered-in fleet capacity	● Revenue	● Lifting ● Owned fleet capacity (Carry-over)	● Expenses ● Employees
Chen and Lam (2018)	● Terminal area ● Berth length ● Number of quay crane	● GDP ● CO2 emissions	● Annual container throughput	● Land area、Energy consumption ● Labor ● Annual container throughput
Chao (2017)	● Owned fleet capacity ● Chartered fleet capacity ● Operating expenses	● Revenue	● Number of port calls (main ports)、 ● Number of port calls (side ports)	
Omrani and Keshavarz (2016)	● Ship purchase cost ● Crew cost ● Costs of spare parts, provisions, insurance ● Costs of repairs (voyage + dry dock) ● Commercial container operation cost + other costs	● Net income (Profits)	● Lease + purchasing、Ship manning cost、Supply of spares & provisions plus 3% overhead、Total available days per year (on-hire days) (Link 1-2) ● Time charter to service provider (container)、Time charter to service provider (passenger) (Link 2-3) ● No. of containers carried per year、No. of passenger + cars carried per year (Link 3-4)	● Commercial container operation cost + other costs ● Commercial passenger operation cost + other costs (stage 2)
Yu and Chen (2016)	● vessel capacity ● handling cost ● other cost ● fuel cost	● revenue ● carbon emissions	● TEU-nautical miles ● the number of destination ports	
Díaz-Hernández et al. (2014)	● Total cost ● Labor cost ● Intermediate input cost ● Capital cost	● Containerized general cargo ● Non-containerized general cargo ● Liquid bulk ● Solid bulk ● Passengers ● Area under concession	● Linear meters of docks ● Total surface area ● Price per linear meter of dock ● Price of total surface area	

References for Table A.2
Chao, S. L. (2017). Integrating multi-stage data envelopment analysis and a fuzzy analytical hierarchical process to evaluate the efficiency of major global liner shipping companies. Maritime Policy & Management, 44(4), 496-511.
Chao, S. L., Yu, M. M., & Hsieh, W. F. (2018). Evaluating the efficiency of major container shipping companies: A framework of dynamic network DEA with shared inputs. Transportation Research Part A: Policy and Practice, 117, 44-57.
Chen, C., & Lam, J. S. L. (2018). Sustainability and interactivity between cities and ports: a two-stage data envelopment analysis (DEA) approach. Maritime Policy & Management, 1-18.
Díaz-Hernández, J. J., Martínez-Budría, E., & Salazar-González, J. J. (2014). Measuring cost efficiency in the presence of quasi-fixed inputs using dynamic data envelopment analysis: the case of port infrastructure. Maritime Economics & Logistics, 16(2), 111-126.
Omrani, H., & Soltanzadeh, E. (2016). Dynamic DEA models with network structure: An application for Iranian airlines. Journal of Air Transport Management, 57, 52-61.
Yu, M. M., & Chen, L. H. (2016). Centralized resource allocation with emission resistance in a two-stage production system: Evidence from a Taiwan's container shipping company. Transportation Research Part A: Policy and Practice, 94, 650-671.

Table A.3 Network DEA Measures in Supply Chain Studies

Authors	Original Inputs	Final outputs	Intermediate measres/links
Amirteimoori et al. (2016)	● Annual cost ● Annual personnel turnover ● Environmental cost	● Number of trained personnel in the fields of job, safety, and health ● Number of green products ● Revenue	● Number of products from supplier to manufacturer ● Partnership cost in green production plans
Izadikhah and Saen (2016)	● Cost of work safety and labor health ● Annual cost ● Environmental cost	● Number of obtained ISO certificates ● Number of trained personnel in the fields of job, safety, and health ● Rate of increasing of number of green products ● Rate of increasing of Revenue	● Rate of increasing of partnership cost in green production plans ● Number of products from supplier to manufacturer
Mahdiloo et al. (2016)	● Engineering specifications	● Environment performance	● Product attributes
Azadi et al. (2015)	● Number of seats ● Operating network ● Cars-labor cost ● Fuel cost ● CO2 emission	● Revenue ● Passenger-km ● Fuel saving	● preventive maintenance ● vehicle-km ● environmental cost

References for Table A.3

Amirteimoori, A., Kordrostami, S., & Azizi, H. (2016). Additive models for network data envelopment analysis in the presence of shared resources. Transportation Research Part D: Transport and Environment, 48, 411-424.

Azadi, M., Shabani, A., Khodakarami, M., & Saen, R. F. (2015). Reprint of "Planning in feasible region by two-stage target-setting DEA methods: An application in green supply chain management of public transportation service providers". Transportation Research Part E: Logistics and Transportation Review, 74, 22-36.

Izadikhah, M., & Saen, R. F. (2016). Evaluating sustainability of supply chains by two-stage range directional measure in the presence of negative data. Transportation Research Part D: Transport and Environment, 49, 110-126.

Mahdiloo, M., Jafarzadeh, A. H., Saen, R. F., Tatham, P., & Fisher, R. (2016). A multiple criteria approach to two-stage data envelopment analysis. Transportation Research Part D: Transport and Environment, 46, 317-327.