DEA under Big Data: Big Data Enabled Analysis of Data and Network Data Envelopment Analysis

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

June 2019

Draft

Keynote presented at Data Envelopment Analysis International Conference (DEAIC2019)
Canterbury, UK, June 28-30, 2019

DEA under Big Data: Big Data Enabled Analysis of Data and Network Data Envelopment Analysis

Joe Zhu

Abstract: This paper proposes that data envelopment analysis (DEA) should be viewed as a method (or tool) for data-oriented analytics. DEA is a data-driven tool for performance evaluation and benchmarking. While computational algorithms have been developed to deal with large volume 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 encompass a broader range of metrics that cannot be modelled by the conventional DEA. It is shown that network DEA is different from the standard DEA, although it bears the name of DEA and some similarity with the conventional DEA. Network DEA is big Data Enabled Analysis (big DEA) of data when multiple (performance) metrics or attributes are linked through network structures. These network structures are too large or complex to be dealt with by the conventional DEA. Unlike the 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.

Keywords: Data Envelopment Analysis (DEA); big data; analysis; performance; productivity; efficiency; composite index

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 multiple outputs related to DMUs.

The focus of this paper is not about how great and versatile DEA has been, but rather how DEA has been evolving. As big data research becomes an important area of Operations Analytics, DEA is evolving into Data Envelopment 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 use as a data-oriented and/or data science tool. INFOR has dedicated two volumes to DEA and its applications in operations (Lim & 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 (2019a) are editing a special issue for the Journal of the Operational Research Society on big data for better productivity. Charles, Aparicio, and Zhu (2019b) 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 information and value from the big data. It is also shown that network DEA is different from the conventional DEA and requires the development 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, & 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-visits 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, & 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 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 that have happened to DEA is linking the DEA models to their economic meaning and foundations. This enabled 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 does. It is well-known that VRS will yield 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, we convert the ratio model (1) into the following linear envelopment DEA model

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

subject to

$$\sum_{j=1}^{n} \lambda_j x_{ij} \leq \theta x_{io} \qquad i = 1,2,\ldots,m;$$
$$\sum_{j=1}^{n} \lambda_j y_{rj} \geq y_{ro} \qquad r = 1,2,\ldots,s$$
$$\lambda_j \geq 0 \qquad j = 1,2,\ldots,n. \qquad (2)$$

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 pointed 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 have been about 900 publications related to DEA. A significant amount of research has been 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 also been used as a predictive analytics by assisting 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 data base. 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, & Zhu, 2010; Tone & 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 use network DEA and dynamic DEA under big data context. A double frontier network DEA approach is used by Badiezadeh, Saen, and Samavati (201x) to study the sustainability of supply chains. Dynamic DEA is used in evaluating the performance of power grid enterprises by Sun et al. (201x) and supply chains by Kahi et al. (201x).

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) incorporate 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). Kuo and Kusiak (2019) note 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 (20xx) 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 (201x). 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).

3. Big Data Enabled Analysis of Data: Network DEA

I think the discussion on what big data represent is still on-going. However, it is clear that big data represent *volume* and *value*. It is very straightforward that one can think of the inclusion of large quantity of DMUs when DEA is applied. However, a large number of DMUs in itself may not 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 as embedded in the big data. Of course, one can include more performance metrics. However, such an action may weaken the discriminatory power of DEA.

Khezrimotlagh et al. (2019) develop algorithms to handle large volume of data (decision making units, inputs, and outputs) under the conventional DEA. In this section, I will talk about how DEA can be used to deal with the "value" aspect of the big data.

Using bank operations as an example, a variety of data or performance metrics is available to be analyzed. A mixture of data, e.g., service and sales data, in the traditional DEA may not clearly characterize the benchmarking purpose. For example, Figure 1 depicts a simple bank operation where the bank loans the funds generated by the deposits (savings and checking accounts) to make profits.

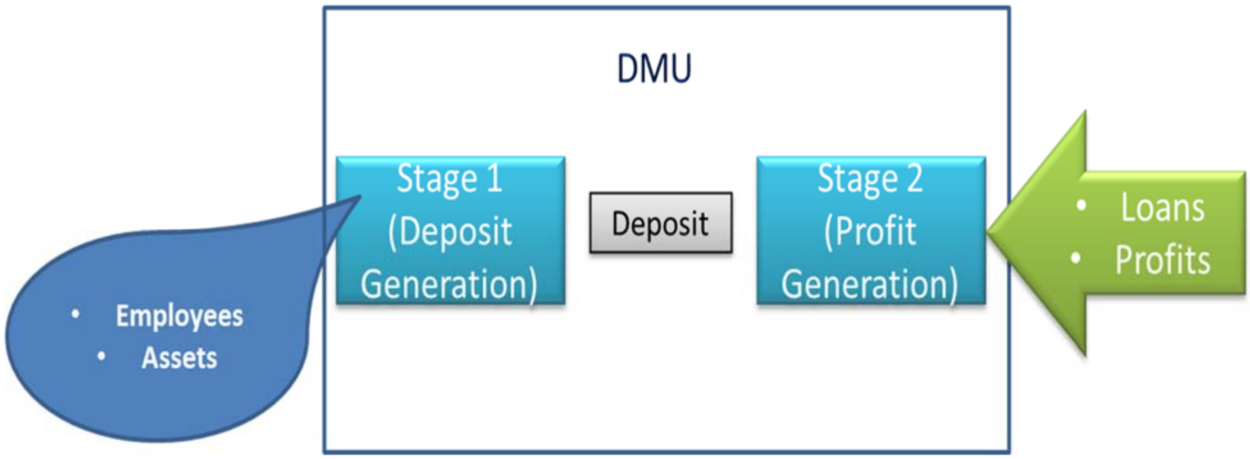

Figure 1. A simple bank operation

A conventional DEA model can evaluate the deposit or loan operation. However, if we combine the performance metrics, whether the “savings” and “checking accounts” should be used as inputs or outputs is not clear. In fact, these performance metrics may represent coordination in-between 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 by coordination between the supplier and the buyer.

While Figure 1 depicts a very simple two-stage operation, Gan et al. (2019) present a more complicated network structure related to an international shipping company in Taiwan.

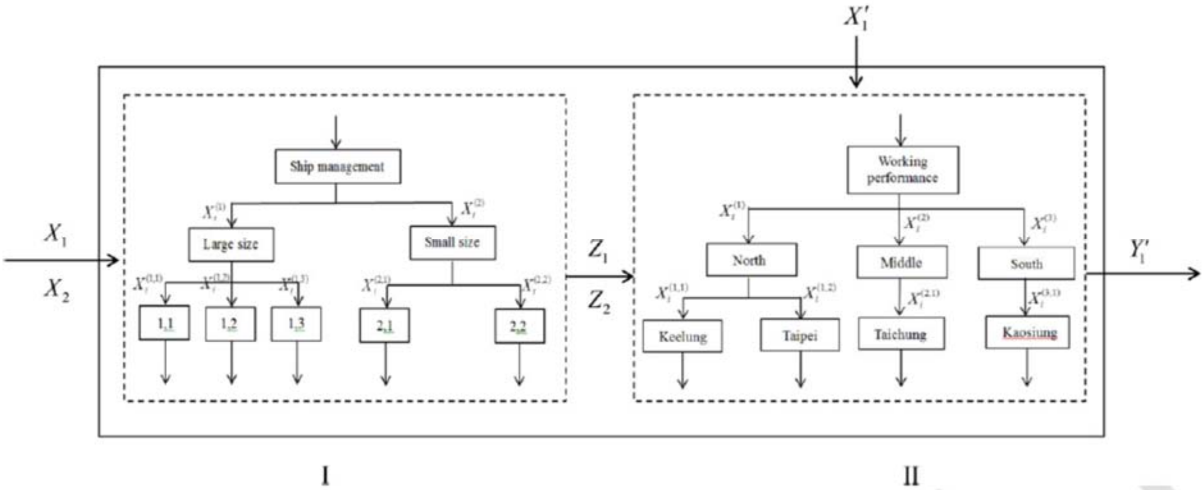

Figure 2. Taiwan's international shipping company

Figure 2 shows the hierarchical network structure of this Taiwanese international shipping company. Interested readers are referred to Gan et al. (2019) for detailed discussion on this complicated network structure. Other network structures can be found in Cook et al. (1998), Cook and Green (2005), and Kao (1998, 2009, 2015), among others.

For illustration purposes, we consider a general two-stage network structure as shown in Figure 3. Each DMU$_j$ (j=1,2,…,n) has m inputs x_{ij} , (i=1,2,…,m) to the first stage and P outputs y^1_{pj} (p=1,2,…,P) that leave the system. In addition to these P outputs, stage 1 has D intermediate outputs z_{dj} (d=1,2,…,D) that become inputs to the second stage. The second stage has as well, its own inputs x^2_{hj} (h=1,2,…,H) that enter from outside the system. The outputs from the second stage are y_{rj} (r=1, 2, …,s).

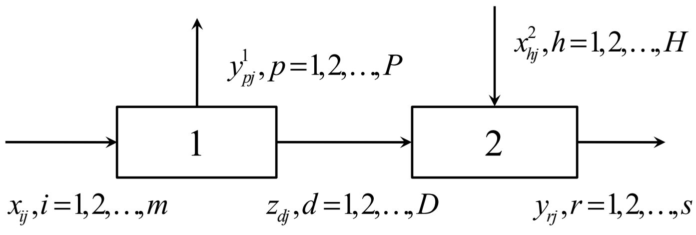

Figure 3. General two-stage network structure

The (efficiency) ratios of stages 1 and 2 for a specific DMU_o under evaluation can be expressed as:

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

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).

Under (weighted) additive efficiency aggregation, we have:

$$e_o^1 + e_o^2 = \alpha \frac{\sum_{d=1}^{D} \eta_d z_{d0} + \sum_{p=1}^{P} \lambda_p y_{p0}^1}{\sum_{i=1}^{m} v_i x_{i0}} + (1-\alpha) \frac{\sum_{r=1}^{s} u_r y_{r0}}{\sum_{d=1}^{D} \eta_d z_{d0} + \sum_{h=1}^{H} Q_h x_{h0}^2}$$

, 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:

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

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, the network DEA is already significantly different from the 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. 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 we do not have a 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:

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

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:

$$\sum_{j=1}^{n} x_{ij}\lambda_j^1 + s_i^- = x_{i0}, \forall i$$
$$\sum_{j=1}^{n} x_{hj}^2\lambda_j^2 + s_h^- = x_{h0}^2, \forall h$$
$$\sum_{j=1}^{n} y_{rj}\lambda_j^2 - s_r^+ = y_{r0}, \forall r \tag{5}$$
$$\sum_{j=1}^{n} y_{pj}^1\lambda_j^1 - s_p^+ = y_{p0}^1, \forall p$$
$$\sum_{j=1}^{n} \lambda_j^1 = 1, \sum_{j=1}^{n} \lambda_j^2 = 1,$$
$$s_i^-, s_h^-, s_r^+, s_p^+, \lambda_j^1, \lambda_j^2 \geq 0$$

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

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

Chen and Zhu (2019) develop the following envelopment form of the network DEA model:

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

Unlike the envelopment models in Tone and Tsutsui (2009), model (7) is a non-linear model that can be solved via the SOCP technique (see Chen & Zhu, 2019).

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 the 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

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 DEA 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 the 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.

From the very first DEA paper (Charnes et al., 1978), it is clear that DEA is a data-oriented technique. While the conventional DEA is linear program based, the network DEA can remain as a non-linear and non-convex model. As a data-oriented technique, DEA and network DEA will 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 the conventional DEA, environmental and sustainability issues are by nature multifaceted that need to be categorized by social, environmental, and financial performances.

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,
6. Charnes A., Cooper, W.W., & Rhodes, E. (1978). Measuring the inefficiency of decision making units. European Journal of Operational Research 2 (6) 429-444.
7. 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.
8. 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.
9. 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.
10. 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.

11. 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.
12. 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
13. 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.
14. Chen, K., & Zhu, J. (2019). Additive slacks-based measure: computational strategy and extension to network DEA. Omega. https://doi.org/10.1016/j.omega.2018.12.011.
15. 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.
16. Cook, W. D., Chai, D., Doyle, J., & Green, R. (1998). Hierarchies and groups in DEA. Journal of Productivity Analysis, 10(2), 177-198.
17. Cook, W. D., & Green, R. H. (2005). Evaluating power plant efficiency: a hierarchical model. Computers & Operations Research, 32(4), 813-823.
18. 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
19. 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
20. 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.
21. Cook, W.D., Tone, K., and Zhu, J. (2014), Data envelopment analysis: Prior to choosing a model, OMEGA, Vol. 44, 1-4.
22. Färe, R., and Grosskopf, S., (2000), Network DEA. Socio-Economic Planning Sciences 34, 35-49
23. 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.

24. 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.
25. 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.
26. 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.
27. 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.
28. 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
29. Kao, C. (1998). Measuring the efficiency of forest districts with multiple working circles. Journal of the Operational Research Society, 49(6), 583-590.
30. Kao, C. (2009). Efficiency decomposition in network data envelopment analysis: A relational model. European Journal of Operational Research, 192(3), 949-962.
31. Kao, C. (2014). Network data envelopment analysis: A review, European Journal of Operational Research 239, 1–16.
32. Kao, C. (2015). Efficiency measurement for hierarchical network systems. Omega, 51, 121-127.
33. 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.
34. 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.
35. Lahdelma R., Salminen, K., (2006), Stochastic multicriteria acceptability analysis using the data envelopment model. European Journal of Operational Research 170, 241–252.

36. 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.
37. 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.
38. Lim, S. and Zhu, J. (2017), DEA and its Applications in Operations - Part I, INFOR, Volume 55, Issue 3, 159-273.
39. Lim, S. and Zhu, J. (2018), DEA and its Applications in Operations - Part II, INFOR, Volume 56, Issue 3, 265-359.
40. Lim, S., and Zhu, J., (2019), Primal-dual correspondence and frontier projections in two-stage network DEA models, OMEGA, Vol. 83, 236-248.
41. 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.
42. Liu, J.S., Lu, L.Y., Lu, W. and Lin, B.J. (2013a), A survey of DEA applications, Omega, 41(5), pp. 893-902.
43. 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.
44. 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.
45. 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.
46. Matthews, K. (2013). Risk management and managerial efficiency in Chinese banks: a network DEA framework. Omega, 41(2), 207-215.
47. 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.
48. 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.

49. 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.
50. 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.
51. 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.
52. 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.
53. 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.
54. 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.
55. 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.
56. Tone, K. and M. Tsutsui, (2009), Network DEA: A slacks-based measure approach, European Journal of Operational Research, Volume 197, Issue 1, 243-252.
57. Tone, K, and Tsutsui, M., (2014), Dynamic DEA with network structure: A slacks-based measure approach. OMEGA 42, 124–131.
58. 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.

59. 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.
60. 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.
61. 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.
62. 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.