Efficiency Analysis of European Freight Villages – Three Peers for Benchmarking

Congcong Yang, Alfred Taudes, Guozhi Dong

Abstract Measuring the efficiency of Freight Villages (FVs) has important implications for logistics companies and other related companies as well as governments. In this paper we apply Data Envelopment Analysis (DEA) to measure the efficiency of European FVs in a purely data-driven way, incorporating the nature of FVs as complex operations that use multiple inputs and produce several outputs. We employ several DEA models and perform a complete sensitivity analysis of the appropriateness of the chosen input and output variables, and an assessment of the robustness of the efficiency score. It turns out that about half of the 20 FVs analyzed are inefficient, with utilization of the intermodal area, warehouse capacity and level of goods handling being the most important areas of improvement. While we find no significant differences in efficiency between FVs of different sizes and in different countries, it turns out that the FVs Eurocentre Toulouse, Interporto Quadrante Europa and GVZ Nürnberg constitute more than 90% of the benchmark share.

Keywords Freight Village, benchmarking, efficiency measurement, Data Envelopment Analysis (DEA)

Congcong Yang
Institute for Production Management, Vienna University of Economics and Business
Welthandelsplatz 1, Building D2, 1020 Vienna, Austria

Alfred Taudes (Corresponding author)
Institute for Production Management, Vienna University of Economics and Business
Welthandelsplatz 1, Building D2, 1020 Vienna, Austria
Tel: (+43 1) 31336 5611 Fax: (+ 43 1) 31336 5610
Email: alfred.taudes@wu.ac.at

Guozhi Dong
Computational Science Center, University of Vienna
Oskar-Morgenstern-Platz 1,1090 Vienna, Austria

1 Introduction

The term "Freight Village" (FV) refers to a defined area organized for carrying out all activities related to transport, logistics and distribution for both national and international transit (Ballis 2006). Initially, FVs were established in response to the challenges posed by regional population and freight growth; however, with the ongoing increase in globalized trade, FVs are widely used in the process of trade and transportation in the world (Wu and Haasis 2013). Spurred by changes in freight and logistics processes, the FV concept has emerged around the world not only as a logistical interconnection point within a logistics network, but also as a "business generator" which contributes to the improvement of supply chain efficiency, regional economic growth and environmental protection (Meidute 2005). In the face of growing globalization of business activities and escalating demand for smoothing the flow of goods within a supply chain, FV management is becoming a daunting task and an important topic in supply chain management.

To achieve profitability and to survive in the market, all enterprises are required to perform activities in an efficient way (Andrejić et al. 2013). The FV is no exception. From a wider perspective, the FV serves as the backbone of the logistics system, affecting the performance of the entire transportation network and supply chain (Cezar-Gabriel 2010). Particularly, the intelligent multimodal transport chains that are implemented by FVs contribute to an efficient logistics network (Winkler and Seebacher 2011). Looking at it from another angle, measuring and improving FV efficiency has significant implications for a number of stakeholders. For example, it assists 3PL companies and other related companies (warehouse operators, transportation operators) in identifying and selecting the most efficient FV at which to base their operations. Also, it aids governments in making effective decisions in FV development programs. Finally, benchmarking is a good method for FV managers to ensure the competitiveness of their organization.

Due to the sizable investment and the significant operating and maintenance costs associated with FV infrastructure and the regional economic ramifications, measuring an FV's efficiency has aroused wide attention. In spite of this, a distinctive characteristic of FV appraisal research is that it lacks standard methodologies or decision criteria (Kapros et al. 2005). Wang and Chen (2009), Luo (2013) and Gronalt et al. (2008) assessed FV via performance index construction. Unfortunately, there are too many papers that replicate previous research while offering scant methodological and theoretical improvements. For instance, the majority of Chinese papers tend to construct logistics park performance frameworks using methods such as Analytic Hierarchy Process (AHP) and the Fuzzy Evaluation Method. Particularly the lack of explication on the variable selection process and its implications calls into question their usefulness as a framework to guide further study. Consequently, Liu et al. (2010) underscored the need to enhance the implementability of performance indicators.

Recent studies have attempted to evaluate the relative efficiency of FVs using Data Envelopment Analysis (DEA). The logic is that an organization's efficiency is a complex phenomenon requiring more than a single criterion to characterize it (Hung and Lu 2008). Therefore, it is reasonable to see efficiency as a relative concept, and a natural measure is the ratio of output to inputs, where larger values of this ratio are associated with better efficiency (Coelli et al. 2005). For example, de Carvalho and Lima (2010) measured and compared the efficiency of six logistics platforms in Europe with DEA to guide the development of new logistics platforms. Haralambides and Gujar (2012) proposed an eco-DEA model and applied it to sixteen dry ports in India. Liu et al. (2013) treated the number of employees as a dual variable and utilized the dual variable DEA model to measure the efficiency of logistics parks in Inner Mongolia.

While introducing DEA as a possible technique for efficiency measurement in FV, this stream of research is still in an early stage. On the one hand, the application process from variable selection to the interpretation of the results has not been exhibited completely and systematically. On the other hand, the number of variables and FVs included in the models has been limited. For example, with only four variables and six FVs taken into account, de Carvalho and Lima (2010, p.10) recommended that future research should use other mathematical models as well and include more variables and FVs in the analysis. Markovits-Somogyi et al. (2011) claimed that although the features of DEA make it appropriate for the efficiency assessment of FVs, it has not been fully utilized for that purpose. Taking these limitations into account, extending previous studies by showing how DEA can be fully applied as a benchmarking tool for FV operations is of great importance to enrich the evaluation research on FVs.

Apart from these research gaps, another motivation for this work derives from the comparison of efficiency measurement studies of FVs from practice and academia. To assess the development level of European FVs, the FV associations EUROPLATFORM EEIG and Deutsche GVZ-Gesellschaft mbH (DGG) carried out a large-scale benchmark study in 2010, in which 78 FVs were assessed and ranked based on a SWOT analysis (Koch et al. 2010). Given different benchmark methods, it is of interest to make a comparison between this study and our approach to check whether the findings in Koch et.al. (2010) are robust and to understand the similarities and differences between SWOT analysis and DEA.

The remainder of the paper is organized as follows: Section 2 introduces the DEA models used in the present study. In Section 3 we deal with the specification of input and output variables. Section 4 describes the results of the empirical analysis, including relative efficiency scores derived from the DEA models introduced by Charnes, Cooper and Rhodes (1978; DEA-CCR; hereafter CCR) and Banker, Charnes and Cooper (1984; DEA-BCC; hereafter BCC). Subsequently, we check the robustness of efficiency scores with regard to changes in the DMUs and variables included in the model. Then we investigate which of the efficient FVs can serve as benchmark partners for many inefficient ones and whether there are significant differences in efficiency between FVs of different size and location. Finally, we compare our results to those of the SWOT based study of Koch at.al. (2010). Section 5 outlines the most relevant conclusions, along with a scope for future research.

2 Research Methodology

DEA is the main methodology employed in this paper. DEA is a non-parametric approach for evaluating the efficiency of a set of peer entities called Decision-Making Units (DMUs), which convert multiple inputs into multiple outputs (Cooper et al. 2011). We argue that DEA is particularly useful for the efficiency measurement of FVs for a number of reasons. Firstly, DEA allows one to gauge the efficiency of FVs without opening the "black box". This is very important, as the operational processes of FVs are complicated so that it is hard to determine a production function. Secondly, unlike AHP and the Fuzzy Comprehensive Evaluation method, which require experts or managers to judge the importance of each indicator, DEA automatically determines the endogenous weights that represent a relative value system for each FV. DEA captures the efficiency of FVs comprehensively by taking multiple inputs and outputs into account. In particular, DEA helps to identify "best practice" from a large number of FVs rather than study only one FV, which thus solves the problem of generalization and applicability when several FVs are involved simultaneously. Last but not least, DEA also works in the case of a small sample size (Sufian 2007), which is important as gathering data from FVs is a daunting task.

The DEA method offers input-side and output-side models. When analyzing the efficiency of a particular DMU, the former pursue minimal inputs when the output levels remain the same as the reference DMU, while the latter seek maximal outputs without changing the input quantities of the reference DMU (Yu and Chen 2011). Within the context of the FVs, both orientations are useful. Managers who are concerned with "how to fully and efficiently use resources" might prefer input-oriented models. By contrast, output-side models are more associated with planning and strategy formulation (Cullinane et al., 2006, Golany and Roll 1989). In this study, output-oriented models are chosen, because (i) outputs in our model are more controllable than inputs. FVs are normally associated with long-lived infrastructures and facilities so that adjusting e.g. the size of a facility in the short term is impossible once it has been built; (ii) an output-oriented model can provide information for managers on the capacity utilization of an

FV, indicating whether output has been maximized given the input, which, in turn provides reference for further expansion planning.

2.1 Data envelopment analysis

2.1.1 The DEA concept

DEA is a mathematical programming approach for evaluating the relative efficiency of DMUs (Malekmohammadi et al. 2011). The concept of DEA is based on the idea that the efficiency of an observation (DMU) is determined by its ability to transform inputs into desired outputs (Tongzon 2001). The basic efficiency measure used in DEA is the ratio of total outputs to total inputs. More specifically, DEA constructs a relative efficiency measure based on a single "virtual" input and a single "virtual" output to compute the efficiency of DMU, when multiple inputs and multiple outputs are present. DMUs which are most efficient in producing the virtual output from the virtual input constitute the efficient frontier and have an efficiency score of one, whereas inefficient DMUs are scored somewhere between zero and one. It is worthwhile to note that the efficiencies estimated using DEA are relative, that is, relative to the best performing DMU(s). Because of this, efficient DMUs in DEA do not necessarily optimize the use of inputs to produce outputs.

DEA is a powerful quantitative, analytical tool. DEA accounts for multiple objectives simultaneously without attaching ex-ante weights to each indicator. DEA compares each DMU to the efficient set of observations, and assumes neither a specific functional form for the production function nor an inefficiency distribution (Adler et al. 2013). DEA makes it possible to identify top performing units in a particular sector and indicates possible ways to improve DMU's efficiency for those units that are far away from "best-practice frontier" (Liang et al. 2008). Up to now, DEA has been extensively used to compare the efficiencies of non-profit and profit organizations in areas such as banking, health care, agriculture and farming, transportation, research and education and corporate real estate (see e.g. Liu et al. 2013; Kritikos et al. 2010; Amirteimoori 2011; Merkert and Mangia 2014; Cullinane 2006).

Although various forms of DEA models are available in the literature, the most widely used models are the CCR and the BCC model (Ho and Zhu 2004). The CCR model was initially proposed by Charnes et al. (1978) under the assumption of constant returns to scale (CRS), that is, all DMUs are operating at an optimal scale. The use of a CCR model yields a global technical efficiency measure without taking any scale effects into consideration. However, the CRS assumption is not always true. Banker et al. (1984) therefore revised the CCR model by allowing variable returns to scale (VRS).

2.1.2 The CCR model

It is assumed that n DMUs are to be evaluated. Each $DMU_j\,(\mathrm{j}=1,2,...,\mathrm{n})$ consumes a vector of inputs $x_j=(x_{1j},x_{2j},...x_{mj})^{\mathrm{T}}$ to produce a vector of outputs $y_j=(y_{1j},y_{2j},...y_{sj})^{\mathrm{T}}$, where the superscript T represents transposition. The DMU to be evaluated is designated as *DMUo* and its input-output vector is denoted as (x_0,y_0). The CCR construction can be interpreted as the reduction of the multiple output/multiple input to that of a single "virtual" output and a "virtual" input. The fractional form of the DEA-CCR model with output-orientation is expressed as:

$$
\begin{aligned}
& Min\ \frac{\sum_i v_i x_{io}}{\sum_r u_r y_{ro}} \\
& s.t\quad \frac{\sum_i v_i x_{ij}}{\sum_r u_r y_{rj}} \ge 1 \quad j=1,2,...,n, \\
& \qquad u_r, v_i > \varepsilon > 0 \quad \textit{for all i and r}
\end{aligned}
\tag{1}
$$

where u_r and v_i are weights assigned to output r and input i, respectively; ε is a non-Archimedean infinitesimal (i.e., a very small positive number such as $\varepsilon=10^{-9}$).

The objective function of equation (1) seeks to minimize the efficiency score of a DMU by choosing a set of weights for all inputs and outputs. The two constraints restrict the efficiencies of all of the DMUs to have a lower bound of 1 and all weights should be positive. DMU_j is considered efficient if the objective function of the equation (1) results in an efficiency score of 1, otherwise it is considered inefficient.

The fractional program (1) is subsequently converted to a linear programming format (2). The model (2) is called multiplier model or primal model and its dual problem can be expressed in (3) as envelopment model. This formulation is an LP problem and therefore can be solved more efficiently than the primal model. This is important when the number of DMUs is larger than the total number of inputs and outputs, which is normally the case when applying DEA (Peng Wong and Yue Wong 2007).

$$
\begin{aligned}
& \min\quad q=\sum_{i=1}^{m} v_i x_{io} \\
& s.t\quad \sum_{i=1}^{m} v_i x_{ij} - \sum_{r=1}^{s} \mu_r y_{rj} \ge 0 \\
& \qquad \sum_{r=1}^{s} \mu_r y_{ro} = 1 \\
& \qquad \mu_r, v_i \ge \varepsilon > 0
\end{aligned}
\tag{2}
$$

$$
\begin{aligned}
&\max \ \varphi + \varepsilon\left(\sum_{i=1}^{m} s_i^{-} + \sum_{r=1}^{s} s_r^{+}\right) \\
&\text{s.t.} \quad \sum_{j=1}^{n} x_{ij}\lambda_j + s_i^{-} = x_{io} \qquad i = 1,2,\dots,m; \\
&\qquad \sum_{j=1}^{n} y_{rj}\lambda_j - s_r^{+} = \varphi y_{ro} \qquad r = 1,2,\dots,s; \\
&\qquad \lambda_j \geq 0 \qquad\qquad j = 1,2,\dots,n.
\end{aligned} \tag{3}
$$

Here, φ is the technical efficiency score which we want to calculate, ε is the previously defined non-Archimedean element, λ_j is the convexity coefficient; s_i^{-} and s_r^{+} represent input and output slack variables, respectively. In equation (3), $DMUo$ is DEA efficient if, and only if, the following two conditions are satisfied: (i) $\varphi^* = 1$ and (ii) $s_i^{-*} = s_r^{+*} = 0, \forall i, r$, where $*$ designates an optimum. The output-oriented measure of technical efficiency of DMU_j is then given by:

$$
TE = \frac{1}{\varphi} \tag{4}
$$

2.1.3 The BCC model

As mentioned previously, the BCC model considers variable returns to scale (Zhou et al. 2008). DMUs can exhibit increasing returns to scale (IRS) or decreasing returns to scale (DRS) for different output ranges. To handle the VRS property in DEA, Banker et al. (1984) incorporated the additional convexity constraint $\sum_{j=1}^{n} \lambda_j = 1$ into equation (3). This constraint ensures that DMUs are only compared to DMUs of similar size.

By using the BCC model, the technical efficiency captured by CCR model TE can be decomposed into pure technical efficiency PTE and scale efficiency SE. Scale efficiency measures the divergence between the efficiency rating of a DMU under the CRS and the VRS assumption (Kumar Singh and Kumar Bajpai 2013, p.417). It can be calculated by the following formula:

$$
SE = \frac{TE}{PTE} \tag{5}
$$

Under this metric, SE=1 indicates scale efficiency, and SE<1 scale inefficiency. If scale inefficiency exists, it is of interest to determine whether IRS or DRS is the cause (Banker and Thrall 1992). A useful test of the RTS properties of DMUs can be obtained by observing the corresponding values of λ^*: if $\sum \lambda_j^* = 1$, the DMU operates at CRS; if $\sum \lambda_j^* > 1$, the DMU operates at DRS; and IRS prevails if $\sum \lambda_j^* < 1$. FVs operating at IRS (DRS) can improve their operation by increasing (reducing) their resources (Zhu and Shen 1995).

2.1.4 Slack analysis

In DEA, slacks indicate sources of inefficiency. In the case of output maximization, input slack implies over-utilized inputs, and in the case of input minimization, output slack is defined as the output that is under-produced (Jacobs et al. 2006). Through the analysis of slack managers can allocate resources more efficiently and improve their organization's efficiency (Grewal et al. 1999). Input and output slacks can be calculated via

$$s_{io}^{-} = x_{io} - \lambda_j x_{ij} \quad (6a)$$

$$s_{ro}^{+} = \lambda_j y_{rj} - \varphi y_{ro} \quad (6b)$$

An output-oriented framework with two outputs ($q_1$and q_2) and a single input (x) is used to illustrate how the output can be expanded in the case of slack. As depicted in Fig 1, A, B & C are the three efficient DMUs that define the efficient frontier, whereas P and Q are two inefficient DMUs. In line with Farrell (1957), the inefficiency of units P and Q is calculated as OP/OP' and OQ/OQ' respectively. However, it is questionable as to whether the point P' is an efficient point since one could expand the amount of output q_1 produced by the amount P'A without using any more inputs. Thus, in this case there is output slack of P'A in output q_1 (Coelli et al. 2005).

(Fig. 1 insert here)

2.2 Sensitivity analysis

DEA is sensitive to data and measurement error as the efficient frontier is formed by the best-performing units (Kummar Singh and Kummar Bajpai 2013) and there is no direct way to assess whether a DMU's deviation from the frontier is statistically significant or not (Smith and Mayston 1987). Therefore, it is necessary to test the reliability and robustness of the position of the efficiency frontier (Smith and Mayston 1987, p.186). In this paper, sensitivity analysis is conducted via the removal of variables and by jack-knifing.

2.2.1 Removal of variables

When applying DEA, a DMU can be judged as efficient if it achieves exceptionally better results in terms of just one input (output) but performs below average in all other inputs (outputs) (Ramanathan 2003). Stated differently, one critical factor might result in an efficient score of 1 for a DMU, while without this factor the efficiency score of the DMU would be much less. An easy way to test if this kind of situation is present is to remove one variable at a time from the variables set and to compare the efficiencies of the resulting structurally perturbed models with the efficiencies calculated on the basis of the original model. In order to maintain the same degree of freedom, the removed variable is returned before the next round of analysis (Singh and Bajpai 2013, Yadav et al. 2010). An efficient DMU that is ranked inefficient due to the omission of just one input or one output should be viewed with caution (Ramanathan 2003). In order to further capture the impact of variables on efficiency, FVs can be classified into different groups, as depicted in Table 1 (Yadav et al. 2010; Kumar Singh and Kumar Bajpai 2013; Jha and Shrestha 2006).

(Table 1insert here)

This sensitivity-based classification is useful for selecting FVs for efficiency improvement. FVs that fall into the distinctly inefficient category obviously need to improve efficiency. Since marginally efficient FVs are sensitive to changes in data and might become inefficient quickly when a few variables change, they take priority over marginally inefficient FVs which are inefficient FVs in all the cases and exhibit low sensitivity to changes in variables. That is, the marginally inefficient FVs need to improve efficiency broadly, while marginally efficient FVs need to pay attention to particular aspects.

2.2.2 Jack-knifing analysis

Jack-knifing is an iterative technique that produces a distribution of estimates by dropping one efficient DMU at a time and observing the change of efficiency scores (Ondrich and Ruggiero 2002, Charles et al. 2012). If significant changes are observed, possible outliers may be present. Otherwise, one can argue that the efficiencies calculated are robust with regard to the set of DMUs chosen.

2.3 Benchmark share measure

The benchmark share measure characterizes the suitability of a DMU to act as a benchmark for a particular input or output variable. It consists of two steps: (i) applying factor-specific models (input/output-specific model) for each inefficient DMU to determine the maximal possible decrease in a certain input (or increase in a certain output) without adjusting the remaining inputs and outputs (Eq. 7-8); (ii) calculating each efficient DMU's benchmark share (Eq. 9-11) Zhu (2000). The bigger the benchmark share measure, the more important an efficient DMU is in benchmarking. A benchmark share of zero indicates that an efficient DMU does not act as a reference for any inefficient unit.

For a particularly inefficient DMU, the factor-specific (kth input-specific and qth output-specific) measure is defined via the following two linear programming problems and the existing variable RTS model's best practice frontier.

The kth input-specific DEA model is given as

$$
\begin{aligned}
&\theta_d^{k*} = \min \theta_d^k, \qquad d \in N,\\
&s.t. \sum_{j \in E} \lambda_j^d x_{kj} = \theta_d^k x_{kd}, \qquad k \in \{1, \ldots . m\},\\
&\sum_{j \in E} \lambda_j^d x_{ij} \le x_{id}, \quad i \ne k,\\
&\sum_{j \in E} \lambda_j^d y_{rj} \ge y_{rd}, \quad r = 1, \ldots s,\\
&\sum_{j \in E} \lambda_j^d = 1,\\
&\lambda_j^d \ge 0, \quad j \in E,
\end{aligned}
$$

(7)

The qth output-specific DEA model is given as

$$
\begin{aligned}
&\phi_d^{q*} = \max \phi_d^q, \qquad d \in N,\\
&s.t. \sum_{j \in E} \lambda_j^d y_{qj} = \phi_d^q y_{qd}, \qquad q \in \{1, \ldots . s\},\\
&\sum_{j \in E} \lambda_j^d y_{rj} \ge y_{rd}, \quad r \ne q,\\
&\sum_{j \in E} \lambda_j^d x_{ij} \le x_{id}, \quad i = 1, \ldots m,\\
&\sum_{j \in E} \lambda_j^d = 1,\\
&\lambda_j^d \ge 0, \quad j \in E,
\end{aligned}
\tag{8}
$$

where $\lambda_j^{d^*}$ and $\theta_d^{k^*}$ are optimal values in (7), $\lambda_j^{d^*}$ and $\phi_d^{q^*}$ are optimal values of (8).

E and N represent the index sets for the efficient and inefficient DMUs identified by the variable returns to scale model. The factor-specific measures in Eqs. 7 and 8 determine the maximum potential decrease of an input and increase of an output without altering other inputs and outputs at current levels. These factor-specific measures are multi-factor efficiency measures, because all related factors are considered in a single model.

The kth input-specific benchmark-share measure for each efficient FV is given as

$$
\Delta_j^k = \frac{\sum_{d \in N} \lambda_j^{d*} (1 - \theta_d^{k*}) x_{kd}}{\sum_{d \in N} (1 - \theta_d^{k*}) x_{kd}} \tag{9}
$$

The qth output-specific benchmark-share measure for each efficient FV is given as

$$
\Pi_j^q = \frac{\sum_{d \in N} \lambda_j^{d*} (\phi_d^{q*} - 1) y_{qd}}{\sum_{d \in N} (\phi_d^{q*} - 1) y_{qd}} \tag{10}
$$

The benchmark share Δ_j^k (or $\prod_j^q$) measures the contribution of efficient units to the potential input (output) improvement in inefficient units and depends on the value of $\lambda_j^{d^*}$ and $\theta_d^{k^*}$ (or $\lambda_j^{d^*}$ and $\phi_d^{q^*}$).

The normalized weights are expressed as

$$\left[\frac{(1-\theta_d^{k*})x_{kd}}{\sum_{d\in N}(1-\theta_d^{k*})x_{kd}}\right], \left[\frac{(\phi_d^{q*}-1)y_{qd}}{\sum_{d\in N}(\phi_d^{q*}-1)y_{qd}}\right] \tag{11}$$

Here, $(1-\theta_d^{k*})x_{kd}$ and $(\phi_q^{d*}-1)y_{qd}$ describe the potential decrease in the kth input and increase in the qth output, respectively, and the value of $\sum_{j\in E}\Delta_j^k=1$ and $\sum_{j\in E}\prod_j^q=1$.

3 Data and variable construction

When applying DEA to study the efficiency of FVs, data availability is an important bottleneck. Due to business confidentiality, only limited information is publicly available on web sites and in annual or statistical reports. We therefore decided to survey FVs in Europe. We specifically addressed those FVs that took part in the study by Koch et al. (2010), which ensured comparability between the FVs and the studies as well as familiarity with the subject. The unit of analysis chosen for this purpose was the FV itself rather than a specific internal facility such as a warehouse or intermodal terminal. Given that an FV is a conglomerate of various organizations and actors that operate independently to a certain extent, the relative efficiency of FVs is measured from a systematic perspective without consideration of certain operators' efficiency. Correspondingly, the efficiency value we captured is a mean value of a certain FV, and it is possible that the efficiency of single operators might be above or below the mean value.

3.1 Identification of input and output variables

Choosing the right input and output variables is very important as DEA results are highly influenced by this choice (De Witte and Marques 2010). However, DEA itself does not provide guidance for the specification of the input and output variables (Nataraja and Johnson 2011). Rather, literature survey and data availability assist in identifying suitable indicators (Bhanot and Singh 2014). Theoretically, the identification of variables should be based upon the operational processes of an FV to ensure precise and complete results. An FV, however, is a highly complex system with a large number of entities, a wide variety of services and complicated relationships among processes. We therefore started out by considering general input and output factors and then took account of the main functions of a typical FV.

Essentially, the inputs are the various resources consumed by DMUs for their operations, while the outputs represent a set of quantitative measures of results expected from operation (George and Rangaraj 2008). In general, resource input can include any combination of labor, equipment, capital and/or information; outputs can be categorized as aggregate revenue, profits, quality, utilization and customer satisfaction (Ross and Droge 2004). As a rule, the initial list of potential variables to be considered for DEA should be large so as to ensure that all potential variables are taken into account. Accordingly, we start out with a list of all variables suggested in previous studies (e.g. Chakraborty et al. 2011; Haralambides and Gujar 2012; Markovits-Somogyi et al. 2011):

- *Total area:* In many cases, this variable is used to measure the input of land. The total area of an FV in hectares, however, is somewhat subjective as some FVs report a gross area including an expanding area that is not yet utilized. Since the undeveloped area does not have an influence on the current output levels, the already-developed area in hectares is more appropriate.
- *Amount of investment:* As an aggregated concept, this indicator includes investment in land, equipment and infrastructure. Due to the huge investment and diversity of shareholders, it is difficult to collect crisp data on this item. For the sake of simplicity, a measurement unit of one million euros is chosen. This measure is affected by differences in purchasing power between different countries. Therefore, adaptions are necessary before this variable is used to measure efficiency in a multinational setting.
- *Intermodal and warehouse area:* Warehouses and the intermodal terminal are the most important infrastructures inside a logistics center (Europlatforms 2004). The warehouse is the infrastructure where the transport operators perform most of their business. The intermodal terminal is the heart of the FV and multimodal transshipment enables the consolidation of transport (Ballis and Golias 2002). Accordingly, it is essential to consider the area of these two facilities measured in hectares as input factors.
- *Number of employees:* Usually, the number of employees is a proxy variable of labor input; however, as argued by Liu et al. (2013) it is reasonable to treat it both as input and output factor in the FV context. Employees of the FV management company assist in the FV's development and operation so that this variable can rightly be viewed as an input. At the same time, job creation is the most important political incentive to establish a new FV and one of the most important goals of FV development (Vrochidis 2013). Furthermore, given today’s high degree of automation in logistics operations the aforementioned input variables are sufficient to capture the capacity of a FV. Hence, it makes

sense to treat the number of jobs created by the operators settled at the site as outcome.

- *Number of companies settled:* We treat this indicator as an output factor, reflecting the development and utilization of FVs. As with the number of employees, it could also be viewed at as an input factor, since it is the companies settled at a site that perform most of the operations of an FV with their employees. Actually, the variables used in DEA analysis do not necessarily have to represent inputs and outputs in the standard notion of production (Cook et al. 2014, p.2). If the goal of DEA analysis is to find the best practice, the efficient DMUs, as defined by DEA, lead to a "best-practice frontier"; therefore, it is more appropriate to interpret them as outcomes, indicators or metrics. Similarly, Nyhan and Martin (1999, p.354) proposed that the term output in DEA can be broadly interpreted to include not only output measures but also quality measures and outcome measures. In our case, ranking FVs is not the primary intention; instead, we are interested in identifying the best-practice frontier and illustrating the appropriateness of DEA in FVs context. From a benchmarking perspective, the inputs and outputs represent the features that managers consider for comparison.
- *Annual load handling:* This indicator is also an output: the more goods are handled, the better the relative efficiency of an FV. Since the majority of FVs can provide intermodal transportation, this variable is the sum of the load handled in an FV, including road, rail and water measured in million tons.
- *Annual revenue:* In Europe, FVs are regarded as commercial organizations, implying that obtaining economic benefits is one of the goals of FVs. Annual revenue is a monetary indicator for measuring the operational profitability and the sustainable development of an FV. It is measured in millions of euros. This variable is affected by differences in prices in different countries. Therefore, adaptions are necessary before this variable is used to measure efficiency in a multinational setting.

3.2 Sampling and data collection

A survey was carried out to collect the variables described above. Questions of both the total area and the already developed area were asked. The questionnaire was developed in the light of Koch et al. (2010) because (i) the data related to the questions can be expected to be available among the members of FV associations which participated in the survey of Koch et al. (2010), and (ii) the results can be compared.

The survey was created on Survey Monkey and sent to 150 FVs. If the mail address of the manager was known, the mail was addressed directly to this person; otherwise the enterprise mail address was used. The first round was followed by two reminder emails and follow-up calls. As the survey involves nine different countries in Europe, the survey language could be expected to affect the response rate. Thus, the questionnaire and invitation letter were translated from English into the respective languages (such as Italian and German) and follow-up calls were made by a native speaker to further explain the survey's purpose and to increase credibility. In addition, company brochures and annual reports were requested as supplementary material. Once feedback became available, we asked the respondent to help by recommending other respondents. In total, the survey was carried out over three months (March 2014-June 2014).

Despite the use of a cover letter assuring data confidentiality, the response rate was quite low (12 responses) and, as expected, respondents skipped some questions. In particular, many FVs did not disclose their annual revenue, so that this variable had to be skipped from the variable set. Also, too few observations were available for the number of employees of the FV management companies. Instead of answering the questionnaire, the companies could also provide information by sending brochures and reports and some information could be obtained from secondary sources such as FV websites, brochures and research reports. By combining these different sources, enough data could be collected for 20 FVs, which were selected for further analysis. The data available for 2013 are summarized in Table 2. To lessen the impact of large differences in data magnitudes (scaling difficulties), the data was normalized before the efficiency value calculation (Sarkis and Talluri 2004).

(Table 2 insert here)

3.3 Variable selection

To further confirm whether the selection of input and output variables is able to fully explain the effect on efficiency, it needs to be verified that an increase of an input will not decrease the output of another item (Liu 2008). Pearson's correlation analysis was employed to determine if this isotonicity property exists between the selected input and output variables (Lin and Hong 2006; Yadav et al. 2011). The resulting Pearson's correlation coefficients are shown in Table 3 together with their significance levels. It is observed that the correlations of "Amount of investment" with all of the output variables are not significant. One reason for this could be that this variable is affected by different prices across Europe. Due to the fact that investments are made over time, no simple weighting scheme for accounting for these differences exists and we therefore decided to exclude this factor.

(Table 3 insert here)

As can be seen from Table 3, the correlation coefficients among the other inputs and outputs are relatively high ($r>0.5$). Some researchers (e.g., Lau 2012) advocated that variables highly correlated with existing model variables are redundant and thus should be removed. However, the use of pairwise correlation should only be seen as a tool for the identification of candidate inputs and outputs and the actual decision should be based on much broader considerations (Dyson et al. 2001; Podinovski and Thanassoulis 2007). In this study, we retained the rest of the variables taking the following reasons into account: (i) correlation results derived from a small sample (twenty FVs) cannot serve for wider reference; (ii) in reality, for instance, the size of an FV does not always correlate with intermodal terminal and warehouse; and (iii) managers may wish to investigate the roles that warehouses and intermodal terminals play in FV efficiency. Table 4 summarizes the variables and provides the corresponding explanation.

(Table 4 insert here)

With respect to the sample size, as a rule of thumb, the relationships among the number of DMUs (n), inputs (m) and outputs (s) should be: $n \geq 2(m+s)$ (Golany and Roll 1989), $n \geq 2m \times s$ (Dyson et al. 2001), $n \geq \max\{m \times s;\ 3(\mathrm{m+s})\}$ (Cooper et al. 2007), respectively. Given $m = 3$ and $s = 3$, the sample size ($n = 20$) used in this study is in line with these recommendations.

4 Empirical results and analysis

The data was evaluated using MaxDEA Pro 6.3 (Chen and Qian 2010), as well as Matlab 2014 and SPSS 21. Both CCR and BCC models were applied due to the lack of precise information on the returns to scale of the FV production function.

4.1 Efficiency value analysis

In this study, an FV is viewed as a DMU and its operating efficiency is broken down into technical (overall), pure technical, and scale efficiencies. Technical efficiency reflects the ability of an FV to obtain maximum outputs given a set of inputs, while scale efficiency reflects the ability of an FV to increase its productivity by achieving its optimal size. The inefficiency of an FV can be attributed to inefficient operation (e.g. too small pure technical efficiency score), disadvantageous exogenous conditions (corresponding to scale efficiency), or both.

Table 5 shows the results obtained from the CCR and BCC models to determine the efficiencies of the FVs under study. As noted previously, the BCC model only identifies pure technical efficiency (PTE), while the CCR model measures overall technical efficiency (TE), which is the combination of PTE and scale efficiency (SE). Hence, the BCC model yields higher technical efficiency values than the CCR model, with respective average values of 0.840 and 0.710. Of the twenty FVs, 60% are found to be technically efficient by the BCC model, while the remaining eight are identified as technically inefficient and their efficiency scores lie between 0.3534 and 0.8897. The

CCR efficiency scores range from 0.2632 to 1, with an overall mean and standard deviation of 0.71 and 0.27, respectively. According to the CCR model, 35% of the FVs are overall technically efficient, whilst 84.61% have efficiency scores below the mean score of 0.71.

(Table 5 insert here)

Notably, seven FVs are overall, pure technically and scale efficient. Bremen, Dresden, Venezia, PLAZA and TVT are efficient in the BCC model, but far away from the CCR frontier. This suggests that the inefficiencies assigned to these five FVs are scale-based. Eight overall inefficient FVs fall short due to technical inefficiency because their PTE scores are lower than their SE scores. This suggests that managers should focus first on removing the technical inefficiency, and then on improving scale efficiencies. Almost half of the scale inefficient FVs (45%) are characterized by IRS followed by CRS (35%). Only 20% of them operate at DRS. In sum, 13 FVs are found to be scale inefficient, implying that 65% of FVs are in an unbalanced status of scale. As can be seen from Table 5, the lowest scale efficiency is calculated for the TVT (0.2632), followed by PLAZA (0.2821). FVs found to be operating under an IRS may prefer to expand their operations in the future. By contrast, for those operating at DRS, their scale sizes need to be decreased for efficiency improvement.

Figure 2 depicts the TE, PTE and SE values for all DMUs. Preliminary observation shows that the technical efficiency scores resulting from the CCR and BCC model show almost the same trend (two lines overlapped), which is slightly different with scale efficiency score. An analysis of variance (ANOVA) of the TE score and PTE score ($F=2.716$) indicates that the technical efficiency measures calculated using CCR and BCC model are not significantly different at the 5% level (with a critical value of 4.098). The Spearman's rank order correlation coefficient between the efficiency rankings derived from CCR and BCC analyses is 0.486, implying that the rank of each FV derived from applying the two different models is similar.

(Fig. 2 insert here)

4.2 Slack analysis

Slacks provide vital information pertaining to the areas in which an inefficient DMU needs to improve its drive towards attaining the status of an efficient one (Kumar and Gulati 2008, p.558). As can be seen from Table 6, "Total area", "No. companies settled" and "Number of jobs" do not require much adjustment. However, utilization is poor for the "Intermodal area" and "Warehouse area" and "Annual load handling" can be improved on the output side. Overall, eleven FVs have non-zero slacks for "Intermodal area", while nine have non-zero slacks for "Warehouse area" and one has non-zero slack for "Total area". Specifically, only Europark has to decrease "Total area" by 137.367 ha to become efficient. Bremen and PLAZA have the greatest excesses in the input variable "Intermodal area" and "Warehouse area". With respect to output slacks, only 50% of FVs have an "Annual load handling" slack equal to zero. This indicates that the other 50% of FVs do not obtain satisfying results on this aspect. In particular, PLAZA, the largest platform in Europe, requires the greatest increase of 157.84 tons in "Annual load handling". In addition, Europark, Novara and Padova need to increase their output for "Number of jobs", while Bremen, Dresden and Novara should attract more companies.

(Table 6 insert here)

4.3 Sensitivity analysis

This section reports the sensitivity analysis results. In order to avoid redundancy, only BCC efficiency scores were scrutinized with the methods described in Sections 2.2.1 und 2.2.2.

4.3.1 Removal of variables

At first we verify whether the efficiency score of an FV is affected if only one input or output is omitted from the data set. Since efficiency never increases upon removal of variables (Pahwa et al. 2003), inefficient FVs in the full BCC model remain inefficient in a new model without a certain variable. Accordingly, attention should be focused on those efficient FVs in the base model which become inefficient after omitting one input or output. As can be seen from Table 7, Quadrante Europa and Eurocentre Toulouse receive identical efficiency values in all situations, implying that their efficiency results are robust. However, all other efficient FVs experience some variation in efficiency scores when variables are omitted. Notably, some FVs were sensitive to a variable change and rapidly become inefficient if only a few variables are dropped. For instance, without "Total area", the scores of Venezia, Dresden and TVT drop by 91.7%, 44.1% and 74%, respectively. This indicates that the "Total area" factor is critical for the efficiency of these three FVs. Therefore, these FVs are not genuinely efficient and their role as benchmarks is questionable.

(Table 7 insert here)

We are now in a position to classify the FVs according to the criteria listed in Table 1. According to those classification criteria, 45% of the FVs are identified as marginally efficient, followed by 40% distinctly inefficient ones. Three FVs are categorized as robustly efficient and no FV is classified as marginally inefficient or significantly inefficient. Table 8 also indicates for each FV the variable that results in a drop of efficiency when being removed. For example, Nürnberg has a BCC score of 1, which becomes 0.472 when excluding "No. companies settled". This implies that "No. companies settled" is a strength of this FV. Overall, "Total area" heavily influences the

BCC score for most FVs, with a changing rate of up to 50%, followed by “No. companies settled” (45%). This means that these two variables are strengths for most FVs and can play a significant role in efficiency improvement. Under output orientation, efficiency can be increased by attracting more companies or by reducing the total area, with the first measure being easier to accomplish in the short run. The Pearson correlation coefficients between the full BCC model and the changed models range from 0.591 to 0.991, implying that the results are robust in general.

(Table 8 insert here)

4.3.2 Removal of efficient DMUs

Twelve additional DEA analyses were performed to test the robustness of the DEA results with regard to stability of the reference set and the presence of outliers. The results in Table 9 show that the average BCC efficiency values vary between 0.7886 and 0.8873 with standard deviations of 2.0267 to 2.6491. Deleting Rovigo and Quadrante Europa has the largest impact. However, in 11 out of 12 cases the reference set remains unaltered. Furthermore, the Spearman’s rank correlation coefficient was used to gauge the similarity of the efficiency ranking between the model with full DMUs and those based on removing one efficient DMU at a time. Table 9 shows that these coefficients range from 0.828 to 1.0 and are significant at 99%. The high rank correlation coefficients indicate that the rankings computed are stable with regard to the removal of efficient FVs, further confirming the robustness of the efficiency analysis.

(Table 9 insert here)

4.4 Benchmark analysis

We will now investigate the role that an efficient FVs can play when benchmarking inefficient FVs. Zhu (2000) recommended two measures of the importance of a DMU for benchmarking: (i) the number of times an efficient unit acts as reference DMU; (ii) the benchmark share measure.

4.4.1 Number of peer count

The peer count number measures the extent to which the performance of an efficient unit can be useful for inefficient FVs (Mostafa 2007). The efficient units that are in the vicinity of the inefficient unit, in other words the efficient units that dictate the projected input of inefficient units, are called peers of the inefficient unit (Pahwa et al. 2003). The efficient unit that appears in the reference set of most of the inefficient units indicates the optimal input-output mix for the inefficient DMUs (Jha and Shrestha 2006). An FV that frequently appears in the reference set is likely to be a genuinely efficient unit and is probably an exemplary efficient performer. On the other hand, FVs that seldom appear in the reference set of other FVs are likely to have a very uncommon input/output mix and are thus not suitable benchmarks.

Based on the reference frequencies of the efficient FVs when both CCR and BCC models are applied, 15 FVs are regarded as the reference set for the inefficient ones (see Fig. 3). Quadrante Europa appears most frequently as a peer in both the CCR (13 times) and BCC model (10 times), followed by Eurocentre Toulouse in the CCR (9 times) and the BCC model (3 times). Seven FVs are treated as a reference set in both the BCC and CCR models: Quadrante Europa, Bologna, Padova, Marche, Verona, Nürnberg, and Eurocentre Toulouse. Here, it is worth noting that although some FVs such as GVZ Bremen and TVT have an efficiency score equal to one, there is no reference from other units to these FVs.

(Fig.3 insert here)

4.4.2 Benchmark share measure

Using the benchmark share measure, we can identify the variable for which a particular efficient FV provides the best benchmark (Yadav et al. 2011). As there are twelve pure technical efficient FVs and six variables, we can compute 72 benchmark shares. As can be seen from Table 10, only 18 out of these 72 are greater than 10% and 4 are greater than 50%. Quadrante Europa, which is a highly technically efficient FV, has the biggest benchmark share in job creation (67.76%). As far as other input/output factors are concerned, Quadrante Europa is still relevant, but not as a leader. In addition, Eurocentre Toulouse also has outstanding benchmark shares in terms of "goods handled" and "number of companies", with benchmark shares of 56.89% and 57.62%, respectively. As far as input "total area" is concerned, Venezia has the highest benchmark share. Rivalta Scrivia (36.75%) and Marche (58.06%) have a leading role in terms of "intermodal terminal" and "warehouse". These benchmarks may offer a first guideline for the efficiency improvement of other FVs.

(Table 10 insert here)

(Fig.4 insert here)

Fig.4 uses a pie diagram to show the benchmark share of pure technically efficient FVs for "No. companies settled". Eurocentre Toulouse alone refers to over half of the potential improvement in attracting companies on site (57.62%). Interporto Quadrante Europa and GVZ Nürnberg have benchmark shares of more than 10%. By contrast, the remaining efficient FVs cannot exert much influence on inefficient FVs. A similar picture can be observed in the case of the two other output variables.

As can be seen from Fig. 4, FVs like PLAZA, TVT, Bremen and Rovigo have benchmark shares below 10% and for most of the inputs and outputs the share is 0%. Although these FVs are efficient, they are too different in the input/output space either to be a reference to other units, or to be referenced. Thus, these FVs are termed self-evaluators. This finding is in accordance with the analysis of the number of peers in Section 4.4.1.

According to the average ranking order, Eurocentre Toulouse, Quadrante Europa and Nürnberg are the three top benchmarks in Europe in terms of all inputs and outputs. Most notably, these three FVs also take a leading role in the number of peer counts, as reported in Section 4.4.1.

4.5 Hypothesis testing

In this section, we will test (i) whether the CCR efficiency scores depend on the FV's region and (ii) whether the size of the FV will affect the CCR efficiency score.

4.5.1 Regional differences in efficiency scores

The FVs in our study come from different countries and 80% of them play a leading role in their home country. We are therefore in a position to investigate whether the efficiency scores vary across different countries. In a first step, the twenty FVs were grouped into five subgroups according to their locations, resulting in Table 11. As there is just one FV each in France, Spain and Portugal, we will restrict ourselves to the comparison of FVs in Germany and Italy. As the tested efficiency scores are not normally distributed, two non-parametrical tests, Kruskal-Wallis and Mann-Whitney, are applied to test whether the efficiency scores differ between these two countries. According to Table 12, the p-values of 0.389 indicate that there are no reasons for rejecting the null hypothesis that there is no difference in efficiency scores of FVs between German and Italy with a significance level of 0.05.

(Table 11, 12 insert here)

4.5.2 Freight villages' size and efficiency score

Previous studies showed that there are differences in efficiency scores between small and large distribution systems or warehouses (Andrejić et al. 2013; Banaszewska et al. 2012; Hamdan and Rogers 2008). We therefore investigate whether there are differences in efficiency scores among FVs of different size. As there is no standard classification of FV size, two approaches are applied to classify FV size: (i) FVs less than 150ha are small, and those over than 150ha are large; (ii) small FVs are less than 100ha, large FVs are over 250ha and a size between 100 and 250 is considered as medium (see Table 13). Since the efficiency scores do not fit within a standard normal distribution, the Mann-Whitney U-test is adapted in the context of two groups. Obviously, with a p-value much larger than 0.05, we cannot reject the null hypothesis of no difference in efficiency between large and small FVs. The case with the three groups is presented in Table 14. Here, a Kruskal-Wallis test was run, and with significance at 0.05 level, we also cannot confirm that there is significant difference among three subgroups of FVs.

(Table 13, 14 insert here)

4.6 Comparing SWOT-based benchmarking with DEA-based benchmarking

Finally, we compare our findings with those of the SWOT-based study in Koch et al. (2010). As can be seen from Table 15, both similarities and differences can be identified between two studies. Both studies investigated the FVs in Europe with a benchmarking perspective, aiming to identify best practice of FVs and to provide references for FV development in Europe. The SWOT-based benchmarking surveyed about 100 locations and took 78 FVs into the analysis. Obviously, our research sample is rather small covering 20 FVs only. The reason for this discrepancy is that the SWOT-based study was conducted by the FVs associations EUROPLATFORM EEIG and Deutsche GVZ-Gesellschaft mbH (DGG). The resulting sample size shows the development status of FVs in Europe more comprehensively and the research results have been broadcasted widely in FV industry especially by those “best in class”.

Major differences between the two studies exist in terms of the methodology adopted and the research results. The DEA method used in our research is a purely data driven method that does not require prior knowledge about the production function and corresponding weights of the different factors. To some extent, it is an objective method and it is only the data that identify the best practice performers. The analysis in Koch et al. (2010), on the other hand, is based on 29 assessment criteria (key performance indicators) and a weighting scheme developed by DGG (the FVs association in Germany). While our research only identifies efficient and inefficient FVs, the outcome of the SWOT-based study is a ranking. According to that ranking, Italian “Interporto Bologna”, German “Güterverkehrszentrum Bremen” and Spanish Transport Centres take a leading role in the European Freight Village landscape, and set the standards of good performance (Koch et al. 2010).

In our study, Eurocentre Toulouse, Quadrante Europa and Nürnberg appear as the three top benchmarks in Europe. GVZ Bremen is the oldest and largest example of an FV developed in Germany. Its operators have 8000 employees – an outstanding figure for Europe. However, this advantage is not reflected in our study, as this FV reveals scale inefficiency and operates in decreasing returns to scale. This demonstrates that, by handing multiple inputs and outputs, the DEA can provide more information on efficiency assessment and improvement. Interporto Bologna, one of the leading FVs in Italy also deserves more attention as it operates on IRS. According to our research, it should expand its scale for future efficiency operation. In fact, in line with our finding according to Interporto Bologna's long-term planning, over 200 hectares of land are to be developed.

(Table 15 insert here)

5 Discussion and Conclusions

This study attempts to provide a compelling answer to the problem of assessing the relative efficiency levels of FVs in Europe. With the application of DEA model, twenty FVs have been analyzed in terms of relative efficiency scores, slack analysis, sensitivity analysis, benchmark analysis and hypothesis testing.

The analysis shows that seven FVs are inefficient in both the CCR and BCC models, while eight FVs suffer from pure technical and scale inefficiency. The mean pure technical efficiency score is found to be 84.03%, and twelve FVs are technically efficient. Only 35% of the FVs operate at constant returns to scale and the rest needs to adjust their operating scale for efficiency improvement. The slack analysis shows that most of the inefficient FVs need to reduce their intermodal and warehouse area and to augment the amount of goods load handling to move closer to the efficiency frontier. Based on reliability tests, we found that our results are robust both in terms of variable and DMU selection. A benchmark analysis was conducted to identify FVs that can serve as benchmarks for certain variables for several other FVs. Comprehensively, Interporto Quadrante Europa and Eurocentre Toulouse dominate the benchmark share and are frequently referenced by inefficient FVs. In the last step, statistical tests were applied to investigate whether differences in efficiency scores exist among countries and the size of FVs. No significant differences were found. We also compared our results to a SWOT-based benchmark study and found notable differences in the resulting top benchmark FVs in Europe.

The contribution of this paper is to enhance the state of the art of FV efficiency assessment. For the first time, this paper introduces the DEA method in the context of FVs for efficiency measurement in a systematically manner by (i) the extension of input and output variables and sample size; and (ii) providing useful insights for FV benchmarking from multiple perspectives. Taking the advantage of DEA and the complexity of FVs into account, this paper showed that DEA is a feasible benchmarking approach for FVs.

It should be noted that this research is an exploratory study; the purpose is not to achieve definitive results (e.g. a ranking of FVs) for the direct use of management.

Rather, it draws attention to the value of benchmarking in an effort to measure the efficiency of FVs and serve as a management tool. In the future, some extensions can be envisaged. First, in view of the limited number of FVs analysed and the relatively small set of inputs and outputs used in the present analysis, further studies are recommended to maximize the sample size and consider a wider range of inputs and outputs. However, DEA is a methodology which relies on accessible information. Indeed, if more data were available, FV efficiency could be more thoroughly explored and detailed. It would be extremely helpful if governments would standardize the data collection and openly publish data, as this would enable fair and transparent comparisons. For increased strategic relevance and reliable results, future research in FVs measurement should also strive to cover longer time spans.

Second, instead of output-orientation standard models, input-oriented models and other extensions can also be utilized to measure more subtleties in reality. Network DEA would be suitable for opening the black box of FVs for further investigation, too. Third, to further confirm the comparability of FVs, future research can divide FVs into various clusters in terms of size, facilities and function, and only FVs belonging to the same cluster are included and compared. Last but not least, other decision-making tools such as AHP or techniques for dealing with missing and fuzzy data should be involved to assist the application of DEA to FVs.

Acknowledgments

The authors would like to thank Dr. Claudio di Ciccio for the questionnaire translation and all of the participants who kindly provided useful information for our survey. The authors are grateful to two anonymous referees for their very constructive suggestions for improving the paper.

Compliance with Ethical Standards

We express our appreciation for the State Scholarship Fund of China Scholarship Council (Grant no.2011613013). We herewith state that there are no potential conflicts of interest (financial or non-financial) and that the submission complies with the Ethical Rules of the Central European Journal of Operations Research (CJOR).

Reference

Andrejić M, Bojović N, Kilibarda M (2013) Benchmarking distribution centres using Principal Component Analysis and Data Envelopment Analysis: A case study of Serbia. Expert Syst Appl 40(10):3926–3933

Adler N, Liebert V, Yazhemsky E (2013) Benchmarking airports from a managerial perspective. Omega 41(2):442–458

Amirteimoori A (2011) An extended transportation problem: a DEA-based approach. Central Europ. J. Operations Res 19(4):513–521

Ballis A (2006) Freight villages: warehouse design and rail link aspects. Transp Res Rec 1966(1):27–33

Ballis A, Golias J (2002) Comparative evaluation of existing and innovative rail–road freight transport terminals. Transp Res Part A Policy Pract 36(7):593–611

Banaszewska A, Cruijssen F, Dullaert W, Gerdessen JC (2012) A framework for measuring efficiency levels–The case of express depots. Int J Prod Econ 139(2):484–495

Banker RD, Charnes A, Cooper WW (1984) Some models for estimating technical and scale inefficiencies in data envelopment analysis. Manag Sci 30(9):1078–1092

Banker RD, Thrall RM (1992) Estimation of returns to scale using data envelopment analysis. Eur J Oper Res 62(1):74–84

Bhanot N, Singh H (2014) Benchmarking the performance indicators of Indian Railway container business using data envelopment analysis. Benchmark Int J 21(1):101–120

Cezar-Gabriel C (2010) Performance assessment in operating dry ports. Ann Faculty Econ 1(2):934–938

Chakraborty PS, Majumder G, Sarkar B (2011) Performance measurement of distribution center combining data envelopment analysis and analytic hierarchy process. Adv Ind Eng Manag 6(2):117–128

Charles V, Kumar M, Irene Kavitha S (2012) Measuring the efficiency of assembled printed circuit boards with undesirable outputs using data envelopment analysis. Int J Prod Econ 136(1):194–206

Charnes A, Cooper WW, Rhodes E (1978) Measuring the efficiency of decision making units. Eur J Oper Res 2(6):429–444

Chen G, Qian Z (2010) MaxDEA LP Manual. http://www.maxdea.cn/. Accessed 14 June 2014

Coelli TJ, Rao DSP, O'Donnell CJ, Battese GE (2005) An introduction to efficiency and productivity analysis. Springer, New York.

Cooper WW, Seiford LM, Tone K (2007) Data envelopment analysis: A comprehensive text with models, applications, references and DEA-Solver software. Springer, New York.

Cooper WW, Seiford LM, Zhu J (2011) Handbook on data envelopment analysis. Springer Science and Business Media, London.

Cook WD, Tone K, Zhu J (2014) Data envelopment analysis: prior to choosing a model. Omega 44:1–4

Cullinane K, Wang TF, Song DW, Ji P (2006) The technical efficiency of container ports: comparing data envelopment analysis and stochastic frontier analysis. Transp Res Part A Policy Pract 40(4):354–374

de Carvalho CC, Lima Jr OF (2010) Efficient logistic platform design: the case of Campinas Platform. http://www.abepro.org.br/biblioteca/enegep2010_TI_ST_113_741_17234.pdf. Accessed 20 February 2013

De Witte K, Marques RC (2010) Designing performance incentives, an international benchmark study in the water sector. Cent Eur J Oper Res 18(2):189–220

Dyson RG, Allen R, Camanho AS, Podinovski VV, Sarrico CS, Shale EA (2001) Pitfalls and protocols in DEA. Eur J Oper Res 132(2):245–259

Europlatforms. E.E.I.G. (2004) Logistics centres directions for use, Europlatforms EEIG.http://www.unece.org/fileadmin/DAM/trans/main/eatl/docs/EN-REV-What is a Freight VillageFinalcorretto.pdf. Accessed 12 March 2012

Farrell MJ et al (1957) The measurement of productive efficiency. J R Stat Soc Series A(General) 3(120):253–290

George SA, Rangaraj N (2008) A performance benchmarking study of Indian railway zones. Benchmark Int J 15(5):599–617

Golany B, Roll Y (1989) An application procedure for DEA. Omega 17(3):237–250

Grewal D, Levy M, Mehrotra A, Sharma A (1999) Planning merchandising decisions to account for regional and product assortment differences. Journal of Retailing 75(3):405–424

Gronalt M G, Posset M, Benna T (2008). Assessing performance indicators for inland terminals. Retrieved from *http://simcont.net/fileadmin/simcont/downloads/Pub_PI.pdf* Accessed 3 May 2010

Hamdan A, Rogers KJ (2008) Evaluating the efficiency of 3PL logistics operations. Int J Prod Econ 113(1):235–244

Haralambides H, Gujar G (2012) On balancing supply chain efficiency and environmental impacts: An eco-DEA model applied to the dry port sector of India. Marit Econ Logis 14(1):122–137

Ho CT, Zhu DS (2004) Performance measurement of Taiwan's commercial banks. Int J Prod Perform Manag 53(5):425–434

Hung SW, Lu WM (2008) The comparative productivity efficiency of Taiwan’s integrated circuits packaging/testing firms. INFOR: Information Systems and Operational Research 46 (3):189–198

Jacobs R, Smith PC, Street A (2006) Measuring efficiency in health care: analytic techniques and health policy. Cambridge University Press, New York.

Jha DK, Shrestha R (2006) Measuring efficiency of hydropower plants in Nepal using data envelopment analysis. IEEE Transactions on Power Systems 21 (4):1502–1511

Kapros S, Panou K, Tsamboulas D (2005) Multi-criteria approach to the evaluation of intermodal freight villages. Transp Res Rec 1906(1):56–63

Koch H, Münch S, Nestler S, Nobel N (2010) Ranking of the European Freight Village locations-Benchmarking of the European experience. wvb Wissenschaftlicher Verlag, Berlin.

Kritikos M, Markellos RN, Prastacos.G (2010) Corporate real estate analysis: evaluating telecom branch efficiency in Greece. Appl. Econ. 42(9):1133–1143

Kumar S, Gulati, R (2008) Evaluation of technical efficiency and ranking of public sector banks in India: An analysis from cross-sectional perspective. Int J Prod Perform Manag 57(7):540–568

Kumar Singh S, Kumar Bajpai V (2013) Estimation of operational efficiency and its determinants using DEA: The case of Indian coal-fired power plants. Int. J. Energy Sect 7(4):409–429

Lau KH (2012) Distribution network rationalization through benchmarking with DEA. Benchmark Int J 19(6):668–689

Liang L, Wu J, Cook WD, Zhu J (2008) Alternative secondary goals in DEA cross-efficiency evaluation. Int J Prod Econ 113(2):1025–1030

Lin LC, Hong CH (2006) Operational performance evaluation of international major airports: an application of data envelopment analysis. J. Air Transp Manag. 12(6):342–351

Liu CC (2008) Evaluating the operational efficiency of major ports in the Asia-Pacific region using data envelopment analysis. Appl. Econ 40(13):1737–1743

Liu JH, Dong JQ, Chang Q (2013) Assessing logistics park operational efficiency using dual variables DEA model. Statistics and Decision (8):60–62

Liu J, Zhou Y, Wei X (2010) International studies on logistics parks from 1999 to 2009 : a literature review. Logi. Technol 29(8):43–45

Luo J (2013) Analysis on the operation effects of logistics park based on BP neural network. In: Proceedings of the 19th international conference on industrial engineering and engineering management, pp 1073–1079

Malekmohammadi N, Lotfi FH, Jaafar AB (2011) Data envelopment scenario analysis with imprecise data. Cent Eur J Oper Res 19(1):65–79

Markovits-Somogyi R, Gecse G, Bokor Z (2011) Basic efficiency measurement of Hungarian logistics centres using data envelopment analysis. Periodica Polytechnica Social and Management Sciences 19(2):97–101

Meidute I (2005) Comparative analysis of the definitions of logistics centres. Transport 20(3):106–110

Merkert R, Mangia L (2014) Efficiency of Italian and Norwegian airports: a matter of management or of the level of competition in remote regions? Transp Res Part A Policy Pract 62, 30–38

Mostafa M (2007) Benchmarking top Arab banks' efficiency through efficient frontier analysis. Ind Manage Data Syst 107(6):802–823

Nataraja NR, Johnson AL (2011) Guidelines for using variable selection techniques in data envelopment analysis. Eur J Oper Res 215(3):662–669

Nyhan RC, Martin LL (1999) Comparative performance measurement: a primer on data envelopment analysis. Public Perform. Manag. Rev 22(3):348–364

Ondrich J, Ruggiero J (2002) Outlier detection in data envelopment analysis: an analysis of jackknifing. J Oper Res Soc 53(3):342–346

Pahwa A, Feng X, Lubkeman D (2003) Performance evaluation of electric distribution utilities based on data envelopment analysis. Power Systems, IEEE Transactions on 18(1):400–405

Podinovski VV, Thanassoulis E (2007) Improving discrimination in data envelopment analysis: some practical suggestions. J Prod Anal 28(1-2):117–126

Peng Wong W, Yew Wong K (2007) Supply chain performance measurement system using DEA modelling. Ind Manage Data Syst 107(3):361–81

Ross AD, Droge C (2004) An analysis of operations efficiency in large-scale distribution systems. J Oper Manag 21(6):673–688

Ramanathan R (2003) An introduction to data envelopment analysis: a tool for performance measurement. Sage Publ, India

Sarkis J, Talluri S (2004) Performance based clustering for benchmarking of US airports. Transp Res Part A Policy Pract 38(5):329–346

Smith P, Mayston D (1987) Measuring efficiency in the public sector. Omega 15 (3):181–189

Sufian F (2007) Trends in the efficiency of Singapore's commercial banking groups: A non-stochastic frontier DEA window analysis approach. Int. J. Prod. Perf. Manag 56(2):99–136

Tongzon J (2001) Efficiency measurement of selected Australian and other international ports using data envelopment analysis. Transp Res Part A Policy Pract 35(2):107–122

Vrochidis B (2013) Logistics centres as economic drivers of their regions. Rotterdam, Netherlands: Department of Erasmus School of Economics: Erasmus University of Rotterdam. http://thesis.eur.nl/pub/14748.

Wang DZ, Chen JJ (2009) Analysis on logistics park performance appraisal based on fuzzy synthetic evaluation method. Logistics Sci-Tech 10(32):79–82

Winkler H, Seebacher G (2011) Management of freight villages: findings from an exploratory study in Germany. Int J Logist-Res App 14(4):271–283

Wu J, Haasis HD (2013) Converting knowledge into sustainability performance of freight villages. Logis Rese 6(2-3):63–88

Yadav VK, Padhy N P, Gupta HO (2010) A micro level study of an Indian electric utility for efficiency enhancement. Energy 35(10):4053–4063

Yadav V K, Padhy NP, Gupta, HO (2011) Performance evaluation and improvement directions for an Indian electric utility. Energy Policy 39(11):7112–7120

Yu MM, Chen PC (2011) Measuring air routes performance using a fractional network data envelopment analysis model. Cent Eur J Oper Res 19(1):81–98

Zhou G, Min H, Xu C, Cao Z (2008) Evaluating the comparative efficiency of Chinese third-party logistics providers using data envelopment analysis. Int J Phys Distr Log 38(4):262–279

Zhu J (2000) Multi-factor performance measure model with an application to Fortune 500 companies. Eur J Oper Res 123(1):105–124

Zhu J, Shen ZH (1995) A discussion of testing DMUs’ returns to scale. Eur J Oper Res 81(3):590–596

Table1 Classification criteria of sensitivity analysis

Efficient status	Criteria	Priority
Robustly efficient	The DEA efficiency stays at one or decreases slightly if one variable is removed at a time.	★☆☆☆☆
Marginally efficient	The DEA efficiency is one for the base model and remains at one in some situations, but drops significantly in other situations.	★★★☆☆
Marginally inefficient	The DEA efficiency is below one but above 0.9 for the base model and stays in that range during the analysis.	★★☆☆☆
Significantly inefficient	The DEA efficiency is between one and 0.9 and drops to much lower values during the analysis.	★★★★☆
Distinctly inefficient	The DEA efficiency is below 0.9 in all situations.	★★★★★

Note: the more ★, the more attention needs to paid to the respective DMU.

Table 2 Summary statistics of the dataset

Variables	Mean	Median	St. deviation	Minimum	Maximum
Input variables					
Total area	253.21	212.00	273.04	22.00	1311.80
Intermodal area	25.80	12.00	40.92	0.03	180.00
Warehouse area	56.67	33.50	80.61	0.25	315.00
Amount of investment	1190.50	149.00	3374.72	18.00	14376.00
Output variables					
Number of jobs	3131.60	1750.00	3863.28	22.00	13000.00
Annual load handling	18.35	6.00	25.97	0.10	80.00
No. companies settled	101.75	96.50	90.70	2.00	270.00

Table 3 Correlations between variables

Items	Total area	Inter-modal area	Warehouse area	Amount of Investment	Number of jobs	Annual load handling	No. companies settled
Total area	1						
Intermodal area	0.509*	1					
Warehouse area	0.856**	0.545*	1				
Amount of investment	0.01	-0.01	0.01	1			
Number of jobs	0.716**	0.483*	0.609**	-0.05	1		
Annual load handling	0.247	0.534*	0.114	-0.13	0.345	1	

No. companies settled	0.645**	0.319	0.474*	-0.10	0.62**	0.545*	1

Note: * Correlation significant at the 0.05 level. ** Correlation significant at the 0.01 level.

Table 4 Definition of variables

Items	Variables	Description	Units
Inputs	Total area	Total area already currently developed , not including the area for further expansion	Hectares
	Intermodal area	The total area of intermodal terminal	Hectares
	Warehouse area	The total area of warehouse	Hectares
Outputs	Number of jobs	The number of employees of companies that rented facilities are working in FV	Number
	Annual load handling	Annual load traffic generated by the facilities offered by the FV	Million Tons
	No. companies settled	Number of companies on site	Number

Table 5 The results of the CCR and BCC efficiency model

No.	FVs	CCR Technical efficiency (TE)	BCC Pure Technical efficiency (PTE)	Scale efficiency (SE)	$\sum\lambda^*$	Returns to scale
1	Eurocentre Toulouse	**1.0000**	**1.0000**	1**.0000**	1**.0000**	CR
2	GVZ Berlin Süd Großbeeren	0.3105	0.3534	0.8785	1.1759	DR
3	GVZ Bremen	0.8340	**1.0000**	0.8340	1.7163	DR
4	GVZ Dresden	0.6456	**1.0000**	0.6456	0.0988	IR
5	GVZ Europark	0.5889	0.5959	0.9882	0.8722	IR
6	GVZ Nürnberg	**1.0000**	**1.0000**	1**.0000**	1**.0000**	CR
7	Interporto Bologna	0.6600	0.6769	0.9750	0.6950	IR
8	Interporto Novara	0.3607	0.4088	0.8824	0.2863	IR
9	Interporto Padova	0.6808	0.6961	0.9780	0.6800	IR
10	Interporto Parma	0.5049	0.5144	0.9816	0.7011	IR
11	Interporto Rovigo	**1.0000**	**1.0000**	**1.0000**	**1.0000**	CR
12	Interporto Venezia	0.5208	**1.0000**	0.5208	0.0838	IR
13	Interporto Verona	**1.0000**	**1.0000**	**1.0000**	**1.0000**	CR
14	Interporto Marche spa	**1.0000**	**1.0000**	**1.000**0	**1.0000**	CR
15	Interporto Nola Campano	0.6667	0.6701	0.9949	1.0119	DR
16	Interporto Quadrante Europa	**1.0000**	**1.0000**	**1.0000**	**1.0000**	CR

17	Interporto Rivalta Scrivia	**1.0000**	**1.0000**	**1.0000**	**1.0000**	CR
18	Interporto Torino	0.8818	0.8897	0.9911	0.9178	IR
19	PLAZA	0.2821	**1.0000**	0.2821	4.8398	DR
20	TVT	0.2632	**1.0000**	0.2632	0.3473	IR
	Mean	0.7100	0.8403	0.8608	1.0213	
	SD	0.2714	0.2264	0.2388		

Notes: IR-increasing returns to scale; CR-constant returns to scale; DR-decreasing returns to scale; $\sum\lambda^*$ sum of optimized value of λ

Table 6 CCR slack analysis of inefficient FVs

Freight Villages	CCR TE	slack values: Total area	Intermodal area	Warehouse area	Number of jobs	Annual load handling	No. companies settled
BerlinSüd	0.310	0	-24.340	-203.700	0	40.523	0
Bremen	0.834	0	-158.345	-64.552	0	0	14.748
Dresden	0.646	0	-5.449	0	0	0	8.287
Europark	0.589	-137.37	0	0	2323.837	8.613	0
Bologna	0.660	0	-1.958	-26.508	0	30.695	0
Novara	0.361	0	-11.106	0	691.898	0	40.245
Padova	0.681	0	-20.720	-10	277.440	47.952	0
Parma	0.505	0	-6.173	-42.243	0	35.396	0
Venezia	0.521	0	-7.583	-9.225	0	1.630	0
NolaCampano	0.667	0	-2.321	-15.702	0	32.095	0
Torino	0.882	0	0	-28.446	0	14.436	0
PLAZA	0.282	0	-27.996	-233.242	0	157.843	0
TVT	0.263	0	-7.005	0	0	1.290	0
No. DMUs with slacks		1	11	9	3	10	3

Note： Negative value means suggest reduction of input parameters.

Table 7 Sensitivity analysis results by removing of variables

No	DMUs	BCC PTE	Efficiency value without input: Total area	Intermodal area	Warehouse area	Efficiency value without output: Number of jobs	Annual load handling	No. companies settled
1	Eurocentre	1.000	1.000	1.000	1.000	1.000	1.000	1.000

	Toulouse							
2	**Berlin Süd**	0.353	0.322	0.353	0.353	0.2037	0.353	0.308
3	**Bremen**	1.000	1.000	1.000	1.000	**0.963**	**0.798**	1.000
4	Dresden	1.000	**0.559**	1.000	1.000	1.000	1.000	1.000
5	**Europark**	0.596	0.596	0.571	0.168	0.596	0.596	0.191
6	Nürnberg	1.000	1.000	1.000	1.000	1.000	1.000	**0.472**
7	Bologna	0.677	0.444	0.673	0.677	0.637	0.677	0.321
8	Novara	0.409	0.329	0.409	0.360	0.409	0.191	0.409
9	Padova	0.696	0.463	0.696	0.696	0.696	0.696	0.143
10	Parma	0.514	0.334	0.514	0.514	0.480	0.514	0.213
11	Rovigo	1.000	1.000	1.000	**0.075**	1.000	1.000	1.000
12	Venezia	1.000	**0.083**	1.000	1.000	1.000	1.000	1.000
13	Verona	1.000	1.000	1.000	1.000	1.000	**0.593**	1.000
14	Marche	1.000	1.000	1.000	0.656	1.000	1.000	1.000
15	Nola Campano	0.670	0.657	0.670	0.670	0.648	0.670	0.449
16	Quadrante Europa	1.000	1.000	1.000	1.000	1.000	1.000	1.000
17	Rivalta Scrivia	1.000	1.000	**0.382**	1.000	1.000	1.000	1.000
18	Torino	0.890	0.815	0.851	0.890	0.874	0.890	0.385
19	PLAZA	1.000	1.000	1.000	1.000	0.926	1.000	**0.923**
20	TVT	1.000	**0.259**	1.000	1.000	1.000	1.000	1.000
Average		0.840	0.693	0.806	0.753	0.822	0.799	0.691
No. of efficient DMUs		12	9	11	10	10	10	10
Changing rate			50%	20%	20%	35%	15%	45%

Note: Changing rate=Total number of changing DMUs (compared to basic BCC model)/Total number of DMUs (20)*100%

Table 8 Classification of FVs based on sensitivity analysis

No.	FVs	BCC PTE	Classification of FVs	Variables to considered for efficiency improvement/Strength
1	Eurocentre Toulouse	1.000	Robustly efficient	
2	Berlin Süd	0.353	Distinctly inefficient	Number of jobs
3	Bremen	1.000	Marginally efficient	Annual load handling
4	Dresden	1.000	Marginally efficient	Total area
5	Europark	0.596	Distinctly inefficient	Warehouse area, No. companies settled
6	Nürnberg	1.000	Marginally efficient	No. companies settled
7	Bologna	0.677	Distinctly inefficient	No. companies settled
8	Novara	0.409	Distinctly inefficient	Annual load handling

9	Padova	0.696	Distinctly inefficient	No. companies settled
10	Parma	0.514	Distinctly inefficient	No. companies settled
11	Rovigo	1.000	Marginally efficient	Warehouse area
12	Venezia	1.000	Marginally efficient	Total area
13	Verona	1.000	Marginally efficient	Annual load handling
14	Marche	1.000	Marginally efficient	Warehouse area
15	Nola Campano	0.670	Distinctly inefficient	No. companies settled
16	Quadrante Europa	1.000	Robustly efficient	
17	Rivalta Scrivia	1.000	Marginally efficient	Intermodal area
18	Torino	0.890	Distinctly inefficient	No. companies settled
19	PLAZA	1.000	Robustly efficient	Number of jobs
20	TVT	1.000	Marginally efficient	Total area

Table 9 Results of the jack-knifing analysis

FVs removed from analysis	Mean PTE	SD.	NE DMUs	Coefficient	New DMUs in the reference set
Eurocentre Toulouse	0.8794	2.4345	12	0.910**	None
Bremen	0.8319	2.2295	11	1.000**	None
Dresden	0.8336	2.2289	11	1.000**	None
Nürnberg	0.8382	2.4456	12	0.963**	None
Rovigo	0.7886	2.0267	10	0.987**	None
Venezia	0.8325	2.2293	11	0.987**	None
Verona	0.8322	2.2294	11	1.000**	None
Marche	0.8336	2.2290	11	1.000**	None
Quadrante Europa	0.8873	2.6491	13	0.828**	Torino, Europark
Rivalta Scrivia	0.8319	2.2295	11	1.000**	None
PLAZA	0.8321	2.2294	11	0.963**	None
TVT	0.8319	2.2295	11	1.000**	None
Full BCC model	0.8403	0.2264	12		

Note: (i) NE: the number of efficient DMUs; (ii) ** Correlation is significant at the 0.01 level (2-tailed)

Table 10 Benchmark shares of 12 efficient FVs

	Output factors		Input factors				*Average*
DMUs	Number of	Annual load	No.	Total	Intermodal	Warehouse	*rank*

	jobs (%)	handling (%)	companies settled (%)	area (%)	area (%)	area (%)	
Quadrante Europa	**67.76 (1)**	3.55 (4)	1.77 (6)	13.11 (4)	30.78 (2)	23.41 (2)	3.17
Marche	0.00 (10)	0.00 (10.5)	2.86 (5)	0.00 (9.5)	0.00 (9.5)	**58.06 (1)**	**7.58**
Rovigo	1.71 (5)	1.09 (6)	1.55 (7)	0.00 (9.5)	1.64 (5)	0.39 (5)	6.25
Venezia	7.93 (3)	0.00(10.5)	17.91 (2)	**39.20 (1)**	0.00 (9.5)	0.00 (9)	5.83
Rivalta Scrivia	0.00 (10)	0.48 (7)	0.29 (8)	0.00 (9.5)	**36.75 (1)**	0.00 (9)	7.42
Verona	0.00 (10)	25.47 (2)	0.00 (10.5)	0.38 (6)	0.00 (9.5)	0.00 (9)	7.83
Dresden	0.12 (7)	9.92 (3)	3.39 (4)	25.43 (2)	13.98 (4)	0.00 (9)	4.83
Bremen	0.00 (10)	0.12 (8)	0.00 (10.5)	0.00 (9.5)	0.00 (9.5)	0.00 (9)	9.42
Nürnberg	17.56 (2)	2.47 (5)	14.60 (3)	3.17 (5)	15.27 (3)	5.93 (4)	3.67
PLAZA	0.30 (6)	0.00 (10.5)	0.00 (10.5)	0.00 (9.5)	0.00 (9.5)	0.00 (9)	9.17
TVT	0.00 (10)	0.00 (10.5)	0.00 (10.5)	0.00 (9.5)	0.00 (9.5)	0.00 (9)	9.83
Eurocentre Toulouse	4.62 (4)	**56.89 (1)**	**57.62 (1)**	18.70 (3)	1.57 (6)	12.21 (3)	3.00
Total	100	100	100	100	100	100	

Note: ranks are given in parenthesis, and ties are assigned mid-rank.

Table 11 Average efficiency scores according to countries

Countries	France	Germany	Italy	Spain	Portugal
Number of units	1	5	12	1	1
Average efficiency score	1.000	0.6758	0.7730	0.2821	0.2632

Table 12 Results of the Kruskal-Wallis and Mann-Whitney tests for differences between Germany and Italy

Kruskal-Wallis test (α=0.05)	Results
Chi-square	0.839
df	1
p-value	0.389

Mann-Whitney test (α=0.05)	Results
U	21.5
Z	-0.916
Exact Sig. (2-tailed)	0.389

Table 13 Average efficiency scores according to the size of Freight Villages

	Testing approach

	2 groups	*(ha)*	*3 groups*	*(ha)*	
Group	Small	Large	Small	Medium	Large
Criteira	<150	>150	<100	[100,250]	>250
Number of units	7	13	6	8	6
Average Efficiency	0.6753	0.7606	0.7278	0.7689	0.8562

Table 14 Results of the Mann-Whitney and Kruskal-Wallis tests for differences between FVs of different size

Two groups Mann-Whitney test(α=0.05)		Three groups Kruskal-Wallis Test(α=0.05)	
U	42	Chi-Square	0.763
Z	-0.283	df	2
Exact Sig. (2-tailed)	0.813	P-value	0.683

Asymp. Sig. (2-tailed) =0.777

Table 15. Comparison SWOT-based benchmarking and DEA-based benchmarking analysis

Item	**SWOT-based benchmarking**	**DEA-based benchmarking**
Content	Ranking der europäischen GVZ Standorte – Benchmarking der europäischen Erfahrungen (2010)	Efficiency Analysis of European Freight Villages – Three Peers for Benchmarking (2015)
Researcher	EUROPLATFORMS and Deutsche GVZ Gesellschaft (DGG)	Research team
Perspectives	Industrial/Practical; Investigation among members of EUROPLATFORMS	Academic; DEA application
Purpose	To assess the level of development of the European FVs, rank FVs on an European level	To examine the efficiency of sampled FVs in Europe at the macro-level, bringing forth scopes of improvement through DEA application and shedding light on efficiency measurement
Methods	Survey (29 key performance indictors); Score by experts; SWOT analysis	Survey (10 research questions); DEA
Results	Ranking of FVs on a European level Management suggestions References for potential customers	Identification of relatively efficient and inefficient FVs; Arouse interest for FV efficiency measurement

Management application	Help to provide more transparency to the market segment of international logistics centers; aim to give a positive impulse to the further successful European development of sustainable macro logistics concepts; allow networking among FVs Europe-wide, supported by better knowledge and access to significant information of the market position and strategies of the individual FVs	Illustrate that DEA is a feasible approach for benchmarking FVs; Help to identify best practice as well as provide improvement directions for inefficient FVs;
Other features	Large sample size (78 FVs/9countries)/ High response rate/ Expert review	Small sample size (20 FVs/5 countries)/Low response rate/ Purely data driven analysis

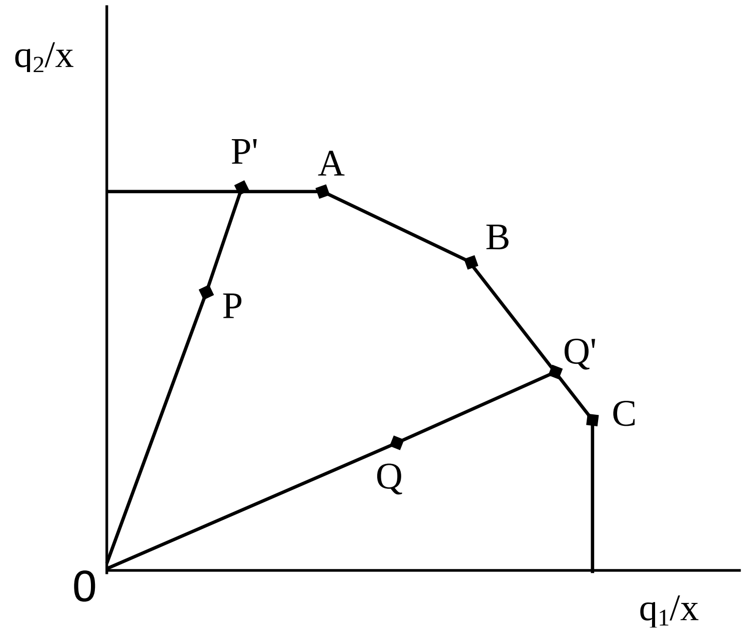

Fig 1. Output-oriented slacks

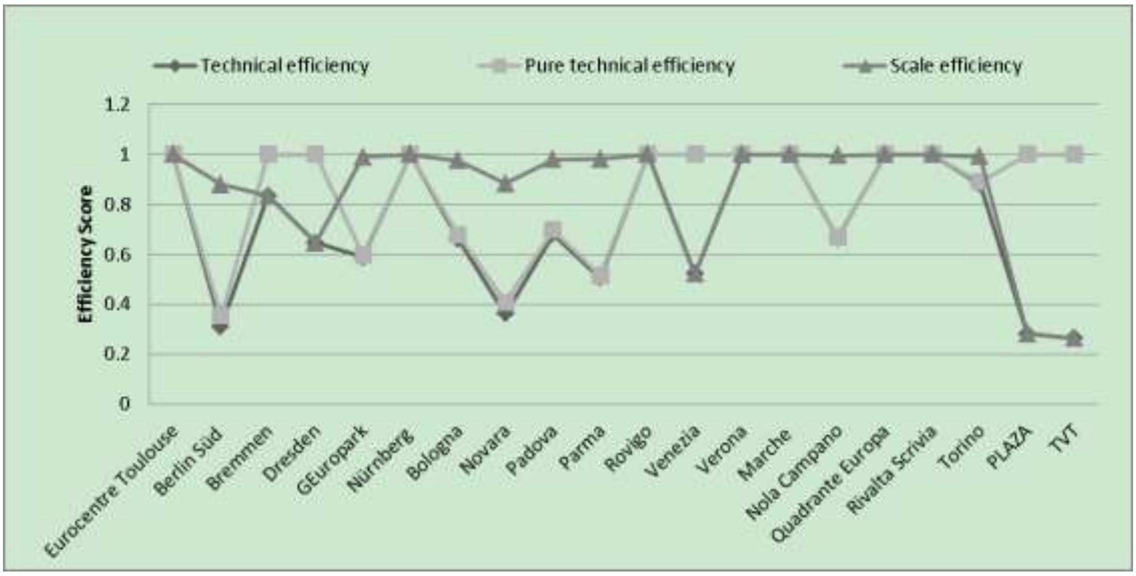

Fig 2. The line chart of TE, PTE and SE

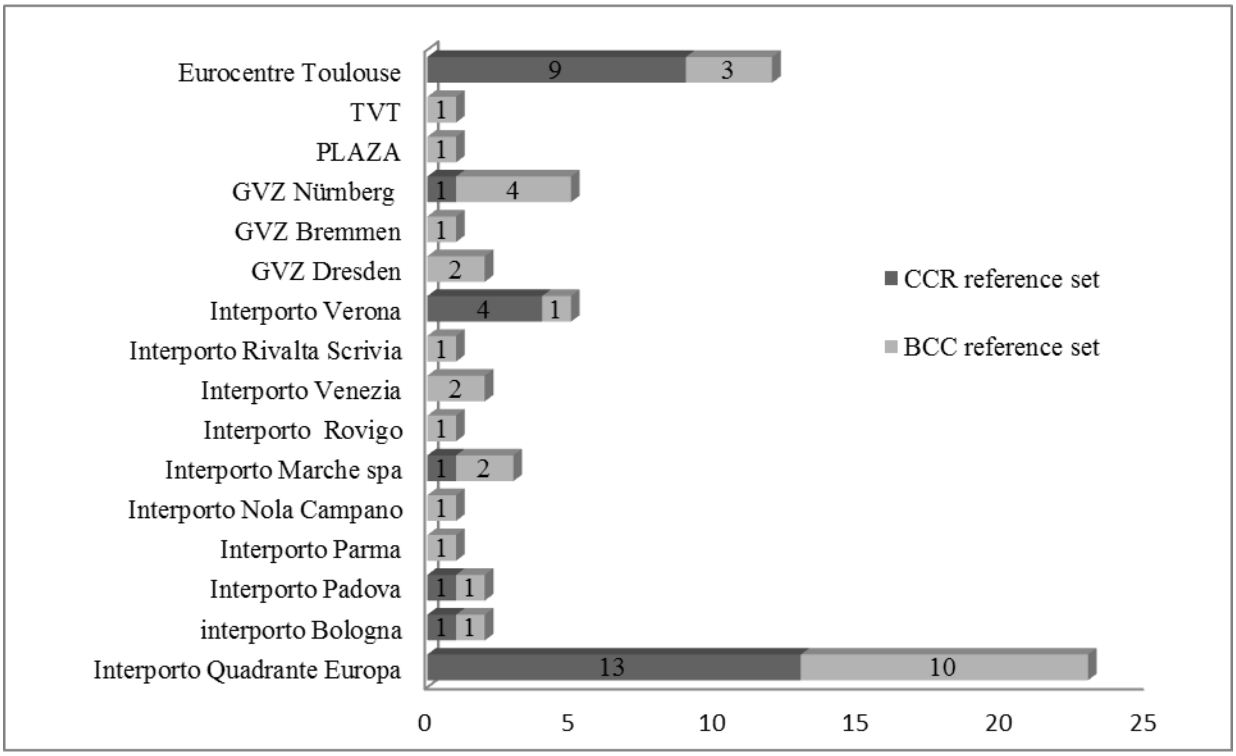

Fig. 3. Reference set frequencies under the CCR and BCC models

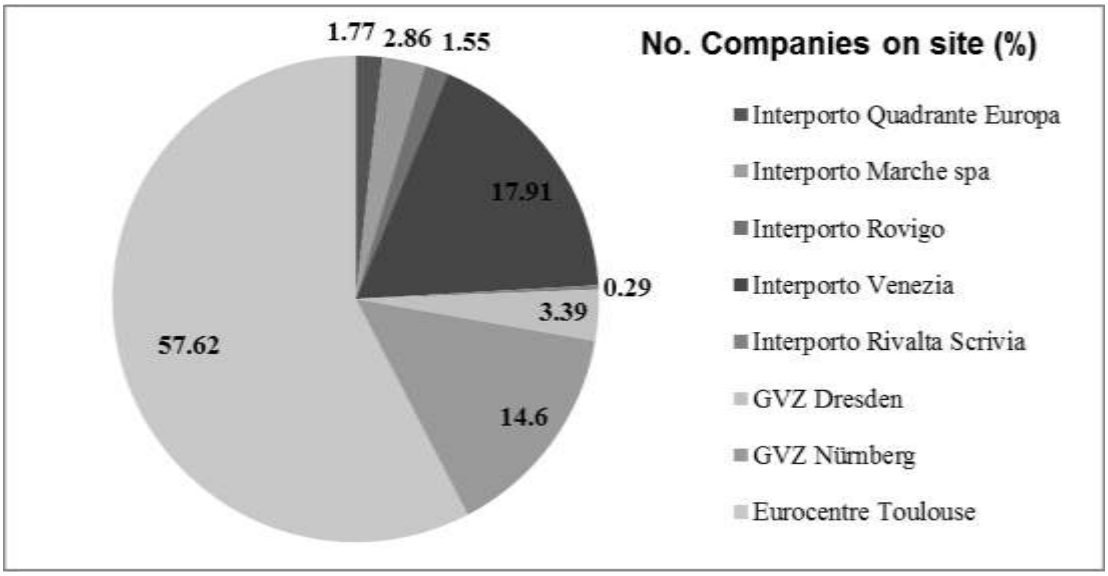

Fig.4. Share of efficient Freight Villages for efficient improvement