Towards better warehouse efficiency distinction through cross-efficiency measurement

Master Thesis
Erasmus University
Rotterdam School of Management
Supply Chain Management

Author: Christian Kaps
Student Number: 454394

Supervisor: Prof. Dr. M.B.M. de Koster
Co-Reader: Prof. Dr. B.M. Balk
External Advisor: Prof. Dr. J.L. Zofío

June 11, 2017

Acknowledgements

This thesis marks the culmination of my Master in Supply Chain Management and as such hopefully facilitates the continuation of my academic career towards a PhD.

I want to thank Prof. Dr. René de Koster for his time, continued feedback and challenging spirit throughout the thesis trajectory. Furthermore, I would like to thank Prof. Dr. Bert Balk for introducing me to the concepts of data envelopment analysis as well as cross efficiency and his scrutiny in assessing the mathematical foundation of this thesis. Also, I am grateful to Prof. Dr. José Zofío for his enthusiasm, conceptual discussions and his advice on this thesis' technical implementation.

Assistance provided by evofenedex, TLN and Dr. Nynke Faber, made the data collection of this thesis possible and is greatly appreciated.

Last but by no means least, I would like to thank my family and friends for their unwearying support, especially my sister Britta, who has accompanied and guided me since the very beginning.

Abstract

Today's customer expects every item to be on stock, personalizeable, shipped within hours and if possible delivered for free. At the same time, the internet has made products and services comparable and allows individuals and companies to shop on a transparent market. To accommodate these demands, warehouses increase floor space, their assortments, broaden their service offering and rely increasingly on technology to improve delivery performance. Among all this change, what constitutes efficient operations? Is outsourcing to a logistics provider the best option? Which industry has the best-in-class warehouse performance and how can all this be measured?

This thesis compares the technical efficiency of warehouses in the Netherlands and Belgium to answer what factors drive operational efficiency in the sector. It looks into differences among product categories, ownership types, value chain positions as well as changes in the industry between 2012 and 2017. The main objective is to find out how automation technology impacts warehouse efficiency and subsequently whether and by how much floor space, assortment size and workforce correlate with efficiency. Additionally, two implementations of Sexton, Silkman, and Hogan (1986), a multiplicative model by Cook and Zhu (2014) and Liang et al. (2008) game theory approach are compared to find the most applicable cross-efficiency method to facilitate such a benchmark.

In order to analyze the research questions, operational data from 102 warehouses was gathered, focused on technical and quantifiable inputs and outputs. Based on this empirical data, cross-efficiency analysis is performed on the entire set and clustered subsets who are then analyzed to find the input mix and warehouse size that allows the most cross-efficient production.

It is found that the input factors assortment (-0.78) and automation (-0.66) exhibit the strongest negative correlation with cross-efficiency across both years. Floor space (-0.43) and workforce (-0.37) are less strongly, but still negatively correlated. These signs and magnitudes of correlations are found across all industry clusters and both observation periods and are significant. Besides the input correlation, it is found that construction and engineering warehouses are less efficient than the overall sample and that no statistically significant differences exist between ownership types and value chain positions.

Yet, these observed inefficiencies are not only of operational nature, as scale is another relevant factor of warehouse performance. This thesis identifies lower and upper thresholds for inputs sizes, beyond which scale efficiency decreases, such as assortment size between 500 - 60,000 SKUs and floor space between 1,280 - 90,000m^2.

Among the cross-efficiency methods, the Sexton's approach, in its original ratio formulation, is found to be best suited for the cross-efficiency comparison, given its methodological proximity to simple DEA, robustness under data modification and ease of implementation. At the same time, the multiplicative framework and the game theory model tested, both revealed pitfalls in application and scoring, rendering them less feasible for studies on real-life data sets.

Despite the aim to make the study as representative as possible, a larger and more granular data set would aid the statistical significance of findings for sub-group analyses with only few observations. Also, the repeated collection of comparable data over time to build a larger panel data set would enable more diverse investigation of temporal efficiency changes. Further research should concentrate on devising a holistic warehouse automation index to offer a scoring framework for future quantitative studies. Lastly, the fields of variable returns to scale cross-efficiency as well as panel-data cross-efficiency offer ample opportunities to advance cross-efficiency calculation and performance benchmarking to more closely resemble the evaluated processes.

Contents

List of Figures

List of Tables

1. Introduction

1.1 Motivation

How competitive are our warehouse operations? Where do we rank within our industry and to what extent is automation technology driving our performance? Although the warehousing and storage industry in the EU accounted for 73 billion € in 2015 and the sector grows faster than the EU's GDP, there is a remarkable research gap in analyzing high-level warehouse efficiency based on real-life data, making it arduous for businesses to answer the above stated questions (Eurostat 2017)(Gu, Goetschalckx, and McGinnis 2007). This thesis aims to fill this void, by exploring to what extent automation technology and other factors impact overall warehouse efficiency.

While ample research has been dedicated to solving specific warehousing problems, such as the optimal number of zones in a pick-and-sort order picking system (de Koster, Le-Duc, and Zaerpour 2012) or the effect of picker personality on picking performance (de Vries, de Koster, and Stam 2016), their scope is almost always limited to a single aspect of a warehouse. Innumerable publications on how to improve single operational decisions have been published, yet besides their narrow focus, they often draw on simulation results (van der Gaast 2015), small empirical samples (Larco et al. 2016) or readily available data that had been collected for a different original purpose and consequently might not be as reliable as desired.

Two papers from the last decade, devoted to warehouse efficiency analysis, are Johnson and McGinnis (2011) on warehousing performance in the United States and De Koster and Balk (2008) benchmarking international warehouse operations. However, as both of these studies use Data Envelopment Analysis *(DEA)* for their analysis, they find 23% and 45% of warehouses to be efficient. From a managerial perspective this renders it difficult to draw conclusions on competitive positioning and best practices.

To overcome this issue, multiple approaches to advance efficiency calculation beyond simple DEA have been suggested, such as cross-efficiency and Stochastic Frontier Analysis *(SFA)*, with the former being applicable in a warehouse efficiency setting, as no production function is known ex-ante (Bogetoft and Otto 2011). Since the emergence of cross-efficiency with Sexton, Silkman, and Hogan (1986), only the adaption by Doyle and Green (1994) has found considerable application in managerial studies, while other authors' methods were often only published and tested on small data-sets, without anyone ever comparing the merits across the individual approaches.

1.2 Research objective

This thesis follows a bi-objective approach. First, it intends to provide empirical insights on the factors that drive warehouse efficiency. A focus hereby will be on the input factor automation, with other factors including warehouse footprint, assortment size and employee count. The analysis is conducted on a cross-sectional level initially and then extended to panel data, comparing the same decision making units' *(DMU)* efficiency between today and 5 years ago.

Second, this thesis will compare four cross-efficiency methods with each other. Primarily it answers, how the methods vary in approach, computational requirements, sensitivity to various changes and most importantly obtained efficiency rankings.

Neither a warehousing benchmark with a tie-breaking method between simple efficient DMUs, nor a comparison of cross-efficiency methods based on real-life data has been conducted yet. In the following the research questions will be introduced.

1.3 Research question

1.3.1 Main research questions

The main managerial research question of this study is: *"What input factors drive warehouse technical efficiency?"*

The main methodological research question of this study is: *"What cross-efficiency method should be used in the future by practitioners when performing efficiency analysis?"*

1.3.2 Sub research questions

For the managerial research question, these sub-questions have been formulated:

- *What are the effects of automation technology on warehouse efficiency?*
- *Which of the other inputs and outputs correlate the most with warehouse efficiency?*
- *How has relative warehouse efficiency changed over the last five years?*

For the methodological research question, these sub-questions have been formulated:

- *Do the different cross-efficiency methods result in statistically different results?*
- *In which aspects do the methods deviate from one another?*
- *Which of the methods' property sets is most conducive for business-context applications?*

1.4 Contribution to practice

For professionals in the warehousing industry, this thesis aims to provide a framework to decide on input factor usage for optimal operational productivity. Especially the focus on automation technology and its development and impact over time is of utmost importance for todays warehouse managers (Bogue 2016). With e-commerce on the rise and the concurrent continued pressure on margins, delivery times, and demand for additional value added service, investing into the right assets has become a core competitive factor in the industry. Given the clustering of warehouses into industry groups, managers can even identify best-in-class input factor allocations for their respective type of operations (Emmett 2005)(Bowersox, Closs, and Cooper 2013). Moreover, the returns to scale analysis offers a chance for managers to gauge the range of competitive warehouses' inputs.

1.5 Contribution to theory

For academics in the fields of efficiency research or productivity bench-marking, this thesis seeks to advance the practice of cross-efficiency usage by comparing the applicability of existing methods to real-life data (Sexton, Silkman, and Hogan 1986)(Liang et al. 2008)(Cook and Zhu 2014). From this, a method recommendation for cross-efficiency usage and consequently unique efficiency ranking is derived. To facilitate future cross-efficiency research, programming implementations for the four considered methods will be publicly provided as part of a DEA MATLAB toolbox (Álvarez, Barbero, and Zofío 2016).

The following chapter contains the literature review and mathematical derivation of the applied concepts. Readers with a primarily managerial focus may skip this part, as it exclusively deals with the intricacies of cross-efficiency calculations and their (non)linear implementations.

2. Literature review

The literature review is split into two parts. First, it briefly discusses the existing research on warehouse efficiency and performance measurement. Afterwards, the cross-efficiency methods used in this paper are introduced in detail.

2.1 Warehouse performance

Johnson and McGinnis (2011) distinguish two separate areas of research to analyze the efficiency of warehouses and their operations. First, there is the line of research focusing on sub-problems within a specific process of the operations. The second area of "logistics benchmarking" is concerned with overall performance comparison and although often applied for companies in general, seems to have found only limited application in the warehouse sector.

2.1.1 Warehouse sub-problems

For the literature devoted to warehousing sub-problems, Gu, Goetschalckx, and McGinnis (2007) provide an extensive overview, clustering the problems into two main fields. The first category is warehouse design, comprising problems on the overall warehouse structure, sizing and dimensioning, department layout, equipment selection and operational strategy. The second category is warehouse operations, which divides into receiving and shipping, storage (SKU-department assignment, zoning, storage location assignment) and order picking (Batching, routing and sequencing and sorting).

Predominantly, literature related to one of these problems, aims to improve the status quo of current operational procedures, potentially even solving them to optimality. Consequently, these publications also increase warehouse performance, but are not concerned with aggregated efficiency of all dimensions interlinked (Gu, Goetschalckx, and McGinnis 2007).

2.1.2 Warehouse benchmarking

The warehouse benchmarking literature is nowhere near as extensive as the above mentioned category. Cohen, Zhen, and Agrawal (1997) were one of the first scholars to benchmark (spare-part) logistics operations. Especially when focusing on recent literature, the results in research databases are meager. De Koster and Warffemius (2005) and De Koster and Balk (2008) compare European distribution centers of international companies using DEA, Johnson and McGinnis (2011) study US-based facilities using the same method. While both models draw on large empirical data sets, the lack of discrimination between DMUs in a simple DEA makes a detailed efficiency comparison impossible. Andrejić, Bojović, and Kilibarda (2013) use principal component analysis *(PCA)* - DEA to analyze Serbian distribution centers. Yet, despite the PCA treatment of data, the limited sample size of seven, renders the results inconclusive.

Banaszewska et al. (2012) develop a framework for measuring efficiency level for depots, justified by attesting warehouses the highest importance for performance in a distribution network. This opinion is shared by the academics in the field - yet also this paper only resorts to applying a DEA approach, which is then extended by ranking efficient DMUs based on the number of times they serve "*as a referent DMU for inefficient DMUs*".

Regarding the employed methods, the current literature favors the non-parametric multi-input, multi-output approach of DEA, coupled with its linear programming properties for warehouse benchmarking, over parametric approaches such as SFA or other non-parametric methods like Free Disposal Hull *(FDH)*. This thesis follows the common practice of DEA-based benchmarking in so far as that its method of choice, cross-efficiency, employs DEA as a starting point of calculations.

Also, this thesis will use the measures and the pre-tested questionnaire from De Koster and Balk (2008) with some modifications, to allow for an increased focus on automation technology's relation to warehouse efficiency. Four cross-efficiency DEA methods will each be applied to the data set in order to break the tie between equally efficient DMUs.

2.2 Efficiency methods

Calculating the efficiency of DMUs in various situations can be achieved through numerous methods. This section gives a short overview of commonly used approaches in operations research, arguing why eventually DEA/cross-efficiency has been chosen for this thesis.

Fried et al. (2008) define efficiency measurement as "a comparison of actual performance with optimal performance located on the relevant frontier". Because this optimal frontier is rarely know in real life, one has to approximate it, using different methods. The broadest classification for these methods provided by the above mentioned authors is into *parametric* and *non-parametric*.

2.2.1 Parametric methods

For parametric methods, a core prerequisite is the existence of a production- or cost function. Estimating these is common practice in the fields of econometrics, which is why these approaches are also often called econometric. With that optimal production frontier at hand, parametric methods then try to distinguish which part of observed deviations from this frontier are statistical noise and which are attributable to a DMU's inefficiency. A prominent example of this approach is the Stochastic Frontier Analysis (Kumbhakar and Lovell 2003). Because there exists no readily available warehouse production frontier specification, parametric methods are infeasible in the context of this thesis.

2.2.2 Non-parametric methods

In contrast to the parametric approach introduced above, non-parametric methods avoid specifying an optimal frontier ex-ante, but rather create Production Possibility Sets *(PPS)*, a set of feasible input and output combinations based on the available data (Cooper, Seiford, and Tone 2007). The most commonly used non-parametric method, DEA, owes its name due to its enveloping property of the dataset's efficient DMUs.

Based on this general idea of assessing efficient DMUs purely based on empirical observations, several extensions of DEA have been developed. When dropping the assumption of convexity, i.e. that the marginal rates of substitution between inputs and outputs are positive, one arrives at the Free Disposal Hull method (Fried et al. 2008). Bogetoft (2000) states that FDH is useful for benchmarking, precisely because the efficient frontier consists of only observed DMUs. However, FDH also excludes input and output vector combinations from the PPS, simply because they were not part of the empirical sample but might be achievable through a (convex) linear combination of DMUs (Green and Cook 2004). Hence, for this benchmarking purpose with a small sample of all warehouses and especially the cross-efficiency application, DEA was given preference over FDH.

Márquez and Lev (2015) also mention Analytical Hierarchy Process *(AHP)* and Technique of Order Preference by Similarity to the Ideal Solution *(TOPSIS)* as common methods for multi-attribute decision making in business settings. AHP was out of scope, as it requires a ranking of importance of goals and alternatives - the opposite of what this thesis aims for. This benchmark is based on the idea that each DMU is able to select its own weights and shadow prices, because multiple input and output vector combinations can be competitively used in warehousing operations.

TOPSIS has been used in conjunction with DEA by Zeydan and Colpan (2009), to yield a preferred order of efficient DMUs after simple DEA application. However, this thesis's benchmark searches for optimal input and output combinations and does not concern itself with a second goal of avoiding the worst-in-sample performance, hence opted against a (fuzzy) TOPSIS approach (Márquez and Lev 2015). Additionally, cross-efficiency is closer to the original idea of DEA and can be considered a natural extension.

In summary, this thesis uses DEA-based cross-efficiency, because this method does not require any ex-ante production frontier, external weights or other assumptions, while uniquely ranking DMUs by proximity to an empirically measured efficiency frontier.

2.3 DEA - Data Envelopment Analysis

2.3.1 Basic DEA concept

In an one-input, one-output scenario, efficiency is merely the ratio of output/input and comparing several entities based on it is trivial. However, when adding more inputs or outputs the efficiency computation can become a complicated endeavor.

Building on (Farrell 1957), Charnes, Cooper, and Rhodes (1978) paved the way for DEA's broad academic and practical adaptation (Fried et al. 2008)(de Koster, Balk, and Nus 2009). According to Cooper, Seiford, and Zhu (2011), DEA's popularity stems from the few required assumptions that nevertheless allow to solve complex cases with unknown input to output relationships. Charnes, Cooper, and Rhodes (1978) defined the objective function to find DMU$_j$'s efficiency (θ_j) as:

$$\max \quad \theta_j = \frac{\sum_{m=1}^{M} y_m^j u_m^j}{\sum_{n=1}^{N} x_n^j v_n^j}, \tag{2.1}$$

where the DMU$_j$'s known M outputs $y_1^j, ..., y_m^j$ are multiplied by their respective weights $u_1^j, ..., u_m^j$ and divided by the N inputs $x_1^j, ..., x_n^j$ multiplied by their respective weights $v_1^j, ..., v_n^j$.

The efficiency score θ_j is sought to be maximized, under the constraints that using those weights on each DMU$_k$ $\quad k = 1, ..., K$, no efficiency score exceeds one:

$$\frac{\sum_{m=1}^{M} y_m^k u_m^j}{\sum_{n=1}^{N} x_n^k v_n^j} \leq 1 \qquad k = 1, ..., K, \tag{2.2}$$

and all inputs, outputs and weights have to be non-negative.

This concept was expanded by various scholars and papers during the subsequent years and decades. Some of the aspects and additions that were made to the original model are introduced in the following subsections.

2.3.2 Input and output orientation

Input-oriented DEA differs from output-oriented DEA in that input-orientation focuses on the reduction of the input vector by a scalar amount. A 70% efficient DMU in an input-oriented model can reduce its inputs by (1-0.7)=30% while keeping the outputs constant to be efficient.

Consequently, in an output-oriented model, the focus is on expanding the output-vector. Under constant returns to scale, the formerly mentioned DMU is still 70% efficient, but the result is interpreted differently. In the output oriented case, the 70% efficient DMU can increase its outputs by 142% = (1/0.7), using the same inputs to be efficient (Daraio and Simar 2006). Which method is used, largely depends on the economic reasoning behind the observed case and whether input and outputs are exogenous or endogenous. There are also ratio-approaches to consider input and output changes at the same time, they are however less popular in practice (Chambers, Chung, and Färe 1996)(Färe et al. 2015).

2.3.3 Technical and allocative efficiency

The difference between technical and allocative efficiency arises from different perspectives on how inputs and outputs can be used and interchanged (Fried et al. 2008). From a technical point of view, a DMU with an efficiency score of less than 1 is inefficient because it fails to generate enough outputs with given inputs or uses too many inputs to create a given output. A technical focus is dominant in situations, where inputs and outputs are largely interchangeable. Allocative efficiency on the other hand, focuses on how well inputs and outputs are allocated in respect to prices. Inefficiencies therefore arise, not because there are not enough outputs generated or too many inputs used, but because the existing inputs are employed in the wrong mix or in the right mix, but thereby achieving a wrong mix of outputs with respect to their market prices. This perspective is mostly found in situations in which a clear economic objective function exists, such as cost minimization or scenarios with customer preferences.

2.3.4 Returns to scale

As Fried et al. (2008) phrase it, "returns to scale relate to how [...] the average product would be affected by scale size". In this thesis, returns to scale *(RTS)* are assumed to be constant, which in real life would mean that by doubling all inputs, one produces twice as many outputs. This perspective is the one propagated in Charnes, Cooper, and Rhodes (1978) and following the authors' names the constant return to scale *(CRS)* model is often abbreviated as *CCR* in the literature. In contrast to CRS, one can allow for diminishing or increasing returns to scale, which for a lot of real-life applications is a warranted assumption. Banker, Charnes, and Cooper (1984) expand the CCR model to allow for variable returns to scale by including another parameter u_0^*, which represents the returns to scale and can be interpreted as the y-axis intercept of a supporting hyper-plane on the efficiency frontier. This variable returns to scale *(VRS)* model is often abbreviated as *BCC*.

Section 5.1.2 introduces RTS calculations in further detail, while section 5.1.3 investigates the RTS distribution in the warehouse data set.

2.3.5 Disposability of inputs and outputs

In the standard DEA model, one assumes strong free disposability. This means that if the input vector x produces the output y, a DMU is free to also produce less output with the same input vector x. Conversely, a DMU is free to produce the output vector y with more inputs than vector x (Fried et al. 2008).

For many applications, this assumption holds, as in a warehouse one can, for example, perfectly well increase employee headcount while keeping the output constant. However, the disposability assumption is challenged in settings with input congestion or undesirable outputs (Cooper, Seiford, and Zhu 2000). A classic example for input congestion is an excess number of miners in a coal mine, blocking each other, thereby reducing output. Undesirable outputs are often part of energy- or sustainability related studies, where pollution is an undesirable output that cannot be avoided (Yang and Pollitt 2009). The interested reader is here referred to Fried et al. (2008), where multiple ways of incorporating such "bad" inputs or outputs are discussed. However, in light of the pre-tested model and the warehouse benchmarking goal of this thesis, the presented models below assume strictly positive data with positive relations between inputs and outputs.

2.3.6 Drawbacks of DEA

In this subsection, the most-commonly cited drawbacks of DEA applications and methods to circumvent those are described.

When performing DEA, as mentioned above, almost always multiple DMUs receive an efficiency score of 1. This inadvertent result stems from each DMU being able to optimize its efficiency score solely focused on its own inputs and outputs and their respective weights. These weights represent the existing substitutability between inputs and outputs combinations, characterizing the production technology. Consequently, a common observation are extreme weight settings, often weighing several inputs and outputs at 0 and heavily focusing on one input and output with the highest competitiveness. This problem only exacerbates the more inputs and outputs are analyzed, as it yields more degrees of freedom (more weight combinations) for a particular DMU to choose from.

For two reasons, multiple "efficient" DMUs are not a desirable outcome:

1. The whole process of DEA is undertaken to differentiate efficiency among entities - yet a result where commonly multiple DMUs receive the same (and the highest possible) score, fails its intended purpose.

2. Outliers can often choose weights that make them appear efficient, although they might only capture a specific niche without being actually efficient.

To remedy this problem, Thompson et al. (1986) suggested to restrict the set of admissible weights, or similarly disallow zero-weights. This however, directly contradicts the original intention behind DEA - to have DMUs rank themselves without any external adjustments and is frowned upon by some DEA pundits.

Cross-efficiency solves precisely the above mentioned issue without imposing any artificial restrictions. Through the peer-appraising nature of cross-efficiency, the resulting scores observe two properties:

1. Cross-efficiency breaks the tie between the multiple DMUs with a DEA efficiency score of 1, thereby allowing a unique ranking of DMUs.

2. Because a DMU's score is dependent on all other peers, maverick DMUs are penalized and receive relatively lower efficiency scores, shifting the focus away from efficient, but unrepresentative entities (Doyle and Green 1994).

2.4 Cross-efficiency

2.4.1 Different cross-efficiency methods

While the merits of cross-efficiency have been introduced above, the method does not come without its own limitations. Most notably, it was found that the weights derived from DEA are non-unique. In other words, for each DMU there can be multiple sets of weights that satisfy the DEA's constraints and still result in the same DEA efficiency. While unpleasant in a DEA case, this becomes problematic for cross-efficiency, as the peer-appraisal scores are derived from exactly these weights.

There have been multiple approaches in the literature to find non-arbitrary cross-efficiency weights, by adding secondary goals to the cross-efficiency computation. Although they all promise to yield unique and reproducible results, it is investigated for four common, but methodologically different approaches, whether the obtained results are the same (or at least similar) and how robust they are.

For this analysis, this thesis compares two implementations of Sexton, Silkman, and Hogan (1986) approach with Cook and Zhu (2014) multiplicative method and with Liang et al. (2008) game theory framework.

2.4.2 Basic cross-efficiency concept

In the following, the basic cross-efficiency concept as well as its core deficiency will be introduced. Subsequently, the four approaches to remedy this issue are presented.

The basic cross-efficiency concept builds on the notion of CRS DEA, however it uses the obtained weights of DMU_j, $j = 1, ..., K$ and multiplies them by each input and output set (x^k, y^k) of DMU_k, $k = 1, ..., K$ to generate peer-appraisal (efficiency) scores for every DMU_j, DMU_k pair (Doyle and Green 1994). This peer-appraisal efficiency score is denoted as:

$$\theta_{j,k} = \frac{\sum\limits_{m=1}^{M} y_m^k u_m^j}{\sum\limits_{n=1}^{N} x_n^k v_n^j} \tag{2.3}$$

In its most basic form, the cross-efficiency then is calculated as:

$$e_k = \frac{1}{K}\sum_{j=1}^{K} \theta_{j,k} \qquad \text{or} \qquad e_k = \frac{1}{K-1}\sum_{j \neq k} \theta_{j,k}, \tag{2.4}$$

depending on whether one chooses to include the self-appraisal score $\theta_{k,k}$ in the analysis or not. A DMU_k's self-appraisal score, evidently, is the same as the simple DEA efficiency score of DMU_k (Balk 2017).

After performing these calculations, one is left with a $K * K$ matrix of peer-appraisal scores, from which the cross-efficiency scores for every DMU can be derived. Although the above mentioned calculation uses the arithmetic mean to aggregate those scores to the cross-efficiency output, Aczél and Roberts (1989) note that a comparison based on cross-efficiency scores is only meaningful using the geometric mean, suggesting the usage of:

$$e_k = \left(\prod_{j=1}^{K} \theta_{j,k}\right)^{\frac{1}{K}} \qquad \text{or} \qquad e_k = \left(\prod_{j \neq k} \theta_{j,k}\right)^{\frac{1}{K-1}}, \tag{2.5}$$

when comparing cross-efficiency scores.

2.4.3 Non-uniqueness problem

Although this cross-efficiency method breaks the tie among all formerly "efficient" DMUs, it uses the non-unique weights of the DEA model. Because these weights are non-unique, there may be more than one set of weights (x^j, y^j) for DMU_j that all result in the same maximal efficiency θ_k^*. As long as these weights are only used for self-appraisal this does not constitute a major problem, but for cross-efficiency purposes these weights impact each DMU's score, rendering the situation problematic. Depending on the implementation, the non-uniqueness property potentially leads to different sets of weights and consequently different cross-efficiency scores (Sexton, Silkman, and Hogan 1986).

Faced with this conundrum, it has to be noted that this is not merely a mathematical oddity, but a considerable practical limitation, because these weights represent the shadow prices of inputs and outputs. With them being arbitrarily selectable, any interpretation of results or peer selection is questionable at best and futile at worst.

2.4.4 Four approaches to remedy the non-uniqueness problem

Approach 1: Linear aggressive or benevolent approach Expanding on Sexton, Silkman, and Hogan (1986), Doyle and Green (1994) was (and until today is) the main source for academics and practitioners alike to expand the standard cross-efficiency model to obtain unique results[1].

The general idea of the agressive or benevolent approach is to solve the cross-efficiency model with an objective to minimize or maximize the sum of all peer-appraisal scores, subject to the restrictions that the self-appraisal scores remain equal to the results of the simple DEA and no peer-appraisal score is >1. Because optimizing over the sum of ratios creates a non-linear problem, Sexton, Silkman, and Hogan (1986) use a "linear surrogate", for which they sum over all DMU$_j$'s $\quad j \neq k$ inputs and outputs, multiplied by DMU$_k$'s weights:

$$\text{Aggregate outputs:} \qquad \sum_{j \neq k} \sum_{m} y_m^j u_m^k, \tag{2.6}$$

$$\text{Aggregate inputs:} \qquad \sum_{j \neq k} \sum_{n} x_n^j v_n^k \tag{2.7}$$

Substracting the aggregated inputs from the aggregated outputs then creates the following linear problem that is solved as a second step, subsequently to the regular DEA optimization:

$$\max B_k = \sum_{j \neq k} \sum_{m} y_m^j u_m^k - \sum_{j \neq k} \sum_{n} x_n^j v_n^k, \tag{2.8}$$

subject to:

$$v_m^k \text{ and } u_n^k \qquad \geq 0, \tag{2.9}$$

$$\theta_{j,k} \leq 1 \qquad \text{for all DMUs } j \neq k, \tag{2.10}$$

$$\sum_{n} x_n^k v_n^k = 1, \tag{2.11}$$

$$\sum_{m} y_m^k u_m^k - \theta_{k,k} \sum_{n} x_n^k v_n^k = 0 \tag{2.12}$$

In Doyle and Green (1994), three slightly different objectives are mentioned, all aiming at minimizing or maximizing the peer-appraisal scores, but with slightly different approaches to stating the objective. The above presented equations are based on the linear, second approach.

[1]While current literature unanimously names Doyle and Green (1994) as the originators of the aggressive or benevolent approach, this is a miss-attribution of credit to the disadvantage of Sexton, Silkman, and Hogan (1986). Sexton, Silkman, and Hogan (1986) not only coin the term cross-efficiency and introduce the method, they also already use the terminology of aggressive and benevolent and last but not least apply it to a data set. This thesis' author conjectures that the general unavailability of this book-chapter in electronic databases caused pundits to propagate the opinion that the more readily disseminated Doyle and Green (1994) is the original source of the method. This thesis wishes to break with this habit and calls the methods Sexton_Classic and Sexton_Ratio from here on.

The models in which this objective is minimized are called "aggressive", while the maximization models are called "benevolent". One may average the results obtained from both approaches, for a less extreme set of cross-efficiencies. One year after the original article, the same authors published another paper, in which they added a fourth formulation of the objective. They also tested these different approaches and found "extremely high correlations [...] between all four ways", which is why in this paper only the second approach will be implemented, without co-implementing the others (Doyle and Green 1995). Until today, the aggressive and benevolent approaches find the most real-life application of all cross-efficiency methods. Examples of that in the supply-chain sector are: Faber, De Koster, and Smidts (2013) on warehouses management or Yu, Ting, and Chen (2010) on information sharing in supply chains.

Approach 2: Ratio aggressive or benevolent approach The formerly introduced method, despite its common application in practice, is based on a model from 1986 and works based on a linear approximation. This thesis diverges from the common scholarly path and also implements the original non-linear approach. Because of the ratio-form objective, this implementation is closer to the pure idea of minimizing the peer-appraisal scores and requires no linear approximation. The author therefore believes the resulting cross-efficiency scores to more closely resemble the methodological DEA basis. The two main reasons that most likely exist(ed) for choosing the linear surrogate over the non-linear original were:

1. The computational power of the average CPU in 1986 was less than 1/10,000 of today's computers, thereby rendering non-linear computational infeasible for most data set sizes.
2. A linear program always finds the same reproducible, optimal solution - non-linear programs do not share the same property. The solution is dependent on solver algorithms, implementation, starting values and solution sensitivity. They might find the global optimum or converge to a local optimum and then stop analyzing the search space further.

With the advance of technology over the last 3-4 decades and the increase of available computational power, the first problem does not pose an issue any longer (Waldrop 2016). For the second issue, the author believes that using the original DEA weights as initial weights for the cross-efficiency optimization function should serve as a natural starting point for the solver. Also up to 10^{50} iterations were performed per model run, to avoid the algorithm prematurely selecting a local optimum as the globally optimal solution.

The idea of the second approach is to also maximize or minimize the peer-appraisal scores as a secondary constraint, but instead of using the linear surrogate as introduced above, this time the actual sum of ratios will be used:

$$\max B_k = \sum_{j \neq k} \frac{\sum_m y_m^j u_m^k}{\sum_n x_n^j v_n^k} \tag{2.13}$$

The constraints in expressions (2.9) and (2.10) can be applied analogously for this method. However, there are two ways to implement the constraint that the newly chosen cross-efficiency weights yield the same efficiency for DMU_k. In the linear model, this was achieved by constraint (2.11) in conjunction with constraint (2.12). As this approach left the realms of linear programming, this could also be stated as:

$$\frac{\sum_m y_m^k u_m^k}{\sum_n x_n^k v_n^k} = \theta_k \tag{2.14}$$

At first, it seems appealing to shift to ratio-based formulation as much as possible, mostly because it avoids the constraint of the input denominator to equal 1. Still, in this thesis it was decided to not employ the ratio constraint and opt for the variant shown in constraints (2.11) and (2.12). The two reasons for this decision are: First, the DEA model employed to derive the simple efficiency score and core component of this constraint is obtained by using a linear approach that constrains the denominator exactly as constraint (2.11) does. Thus, a linear constraint is more closely related to the original DEA idea. Second, by virtue of this standardization to 1, the resulting weights are less erratic as in the non-limited ratio score case.

Therefore, in this model the ratio objective will be used in combination with the linear constraints previously introduced.

Approach 3: Multiplicative approach The third approach is based on a multiplicative approach to DEA, originally introduced by Charnes, Cooper, Seiford, et al. (1982) that was extended by an e^w term, to produce results that remain invariant, when changing an input's or output's unit of measure (Charnes, Cooper, Seiford, et al. 1983). Hence, DMU$_k$'s simple efficiency following this multiplicative approach is calculated as:

$$\theta_k = \max_{u,v,\omega} \left(e^{\omega} \frac{\prod_{m=1}^{M} (y_m^k)^{u_m}}{\prod_{n=1}^{N} (x_n^k)^{v_n}} \right), \tag{2.15}$$

subject to:

$$e^{\omega} \frac{\prod_{m=1}^{M} (y_m^j)^{u_m}}{\prod_{n=1}^{N} (x_n^j)^{v_n}} \leq 1 \qquad j = 1, ..., K, \tag{2.16}$$

$$u, v \geq 1, \tag{2.17}$$

$$\omega \geq 0 \tag{2.18}$$

This measure was extendend by Cook and Zhu (2014) to a cross-efficiency logic, which is similar to the benevolent approach introduced in the previous section. Yet, following this multiplicative DEA uses the weights as exponents rather than factors and multiplies inputs and outputs with each other rather than summing over them. The cross-efficiency score for DMU$_k$ following Cook and Zhu (2014) is therefore computed as:

$$E_k = \max \left(\prod_{j=1}^{K} \frac{e^{\eta_j} \prod_{m=1}^{M} (y_m^k)^{u_m^{j*}}}{e^{\xi_j} \prod_{n=1}^{N} (x_n^k)^{v_n^{j*}}} \right)^{\frac{1}{K}}, \tag{2.19}$$

subject to:

$$E_{k,j} = \frac{e^{\eta_j} \prod_{m=1}^{M} (y_m^k)^{u_m^{j*}}}{e^{\xi_j} \prod_{n=1}^{N} (x_n^k)^{v_n^{j*}}} \leq 1 \qquad j = 1, ..., K; \quad k = 1, ..., K, \tag{2.20}$$

$$E_{k,k} = \frac{e^{\eta_k} \prod_{m=1}^{M} (y_m^k)^{u_m^{k*}}}{e^{\xi_k} \prod_{n=1}^{N} (x_n^k)^{v_n^{k*}}} = \theta_k^* \qquad k = 1, ..., K, \tag{2.21}$$

$$v_m, x_n \geq 1, \tag{2.22}$$

$$\xi_k, \eta_k \geq 0, \qquad k = 1, ..., K \tag{2.23}$$

Constraint (2.21) hereby limits the cross-efficiency self-appraisal score to be equal to the DEA efficiency result (θ_k^*) for that DMU$_k$. It shall be noted that constraint (2.20) iterates over j and k, which is due to the maximum efficiency approach implemented. In the above mentioned paper, the authors first introduce a standard efficiency model, which follows Sexton's models' logic, but uses multiplication rather than addition. In the maximum efficiency model, this is replaced by each DMU$_k$ individually maximizing the peer-appraisal score it receives from each DMU$_j$. This requires K^2 constraints for each DMU, but therefore results in slightly higher efficiency scores, as the maximum peer-appraisal scores per DMU-pair are used, rather than utilizing the maximum sum overall all peer-appraisals of each DMU.

This entire model can be linearized, by taking logarithms of the input and output values, which transforms the products to sums. By doing this, even large DMU sets with many input and outputs can be solved in very limited time, using linear programming techniques. Cook and Zhu (2014) claim that "the attractive feature of the proposed multiplicative approach is that the resulting cross-efficiency score is uniquely determined".

However, in this thesis, the less complex notation from Charnes, Cooper, Seiford, et al. (1982) that excludes the variable returns to scale (and is thereby better comparable with Sexton's approaches) will be employed. Consequently, for the CRS efficiency, ω is excluded:

$$\theta_k = \max_{u,v} \left(\frac{\prod_{m=1}^{M} (y_m^k)^{u_m}}{\prod_{n=1}^{N} (x_n^k)^{v_n}} \right), \tag{2.24}$$

and for the cross-efficiency calculation η and ξ are excluded:

$$E_{k,k} = \frac{\prod_{m=1}^{M} (y_m^k)^{u_m^{k*}}}{\prod_{n=1}^{N} (x_n^k)^{v_n^{k*}}} = \theta_k^* \qquad k = 1, ..., K \tag{2.25}$$

The constraints are used analogously. While this updated model does not exhibit unit-invariance anymore, it is more comparable to the other models, as it follows a CRS logic and for the purpose of this thesis unit-invariance is a not a necessary model property.

Approach 4: Game theory approach The three approaches introduced so far were "pure" cross-efficiency methods, in that they follow the basic logic of Charnes, Cooper, Seiford, et al. (1982) and Sexton, Silkman, and Hogan (1986) in calculating the DEA efficiencies first and subsequently solving a second problem based on the original solution to obtain unique weights and thereby reproducible results. The game theory approach, introduced by Liang et al. (2008) approaches cross-efficiency differently. In an initial step one solves the simple DEA problem. Then, using those weights obtained for DMU$_k$ one calculates a *game cross-efficiency* α for DMU$_j$ in period 0:

$$\alpha_{k,j}^0 = \frac{\sum_{m=1}^{M} u_{mj}^k y_{mj}}{\sum_{n=1}^{N} v_{nj}^k x_{nj}}, \qquad k = 1, ..., K, \tag{2.26}$$

where u_{mj}^k and v_{nj}^k "*indicate that DMU$_j$ is permitted only to choose weights that will not deteriorate*" the currently estimated efficiency of DMU$_k$. In the following, this game cross-efficiency is transformed to α_k , by averaging across all $\alpha_{k,j}$. This way, each DMU$_k$, $\quad k = 1, .., K$ has only one associated α_k value per period, which is updated iteratively as shown in equation (2.33). The DEA results are used to derive α_k for the starting period 0, which is then updated by solving the following program for each DMU$_j$:

$$\max \sum_{m=1}^{M} u_{mj}^k y_{mj}, \tag{2.27}$$

subject to:

$$\sum_{n=1}^{N} v_{nj}^k x_{nl} - \sum_{m=1}^{M} u_{mj}^k y_{ml} \geq 0, \qquad l = 1, ..., K, \tag{2.28}$$

$$\sum_{n=1}^{N} v_{nj}^k x_{nj} = 1, \tag{2.29}$$

$$\alpha_k \sum_{n=1}^{N} v_{nj}^k x_{nk} - \sum_{m=1}^{M} u_{mj}^k y_{mk} \leq 0, \tag{2.30}$$

$$u_{nj}^k \geq 0 \qquad n = 1, ..., K, \tag{2.31}$$

$$v_{mj}^k \geq 0 \qquad m = 1, ..., K \tag{2.32}$$

After having solved this model for each DMU_k, DMU_j pair, one derives the α_k^{t+1} values in the following manner:

$$\alpha_j^{t+1} = \frac{1}{K} \sum_{k=1}^{K} \sum_{m=1}^{M} (\alpha_k^t) u_{mj}^{k*} y_{mj} \tag{2.33}$$

With these new game cross-efficiency scores, one re-iterates through the model until the α values converge. Liang et al. (2008) prove that this is the case, which gives this model the property of finding unique cross-efficiency scores.

2.4.5 Application to a real-life data set

Three out of the four aforementioned methods have all been introduced by their authors given numerical examples, and especially Sexton, Silkman, and Hogan (1986) method (arguably because of its implementational simplicity) has found considerable application in practice. Yet, they have rarely (if never) been compared against each other on a real-life data set.

This research gap coincides with the above mentioned purpose of this study of analyzing warehouses in the Netherlands and Belgium for their technical efficiency. All observations have specifically been collected for a (cross)-efficiency application to guarantee accuracy, comparability and recentness of the underlying data. In gathering this new data set, the author hopes to minimize non-method-attributable differences in the results.

3. Methods and conceptual framework

3.1 Cross-efficiency DEA model

For a DEA model, it is pivotal to select as few inputs and outputs as possible, while capturing all relevant factors that determine productivity. Failing to include a valid parameter "will bias the results against efficient users of the input or efficient producers of the output" (Sexton 1986). Simultaneously, adding parameters that do not influence efficiency increases the degrees of freedom for DMUs, and thereby makes certain DMUs appear more efficient than they actually are. To mitigate any risk of such false attributations interfering with the results of this study, a DEA model is used that draws on De Koster and Warffemius (2005), De Koster and Balk (2008) as well as Faber, De Koster, and Smidts (2013), as they have extensively surveyed warehouses and performed research based on this thoroughly tested and proven questionnaire in the past. While De Koster and Balk (2008) have used four inputs and five outputs, Faber, De Koster, and Smidts (2013) reduced this to four inputs and four outputs, combining the formerly separate outputs of *value-added logistics* and *special processes* into one construct, as value-added services by itself did not significantly impact efficiency. Consequently, this thesis uses the following DEA model:

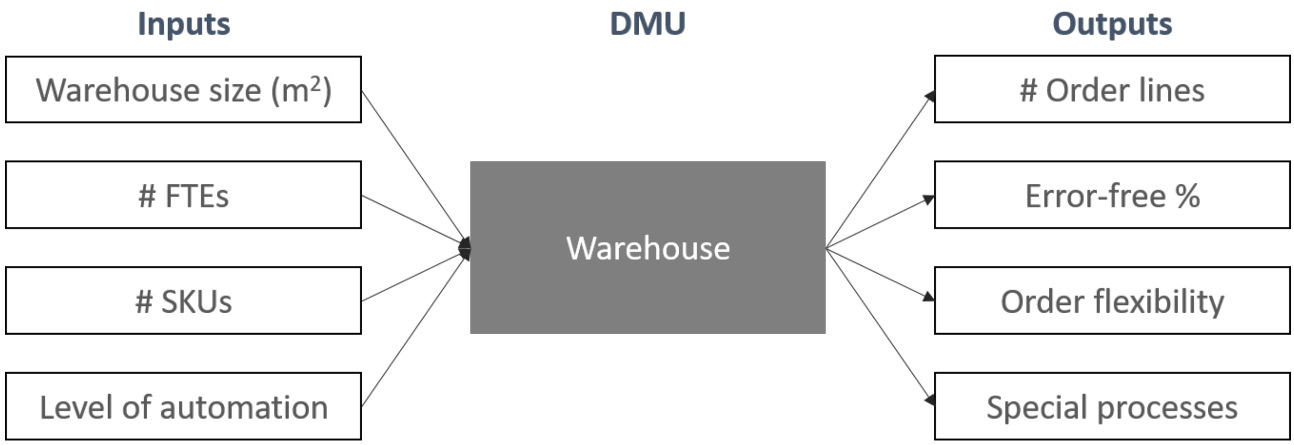


Figure 3.1: DEA input-output model of a warehouse.

3.2 Measurement of variables and constructs

In this section the input and output factors and their measurement are reviewed. For a discussion of the merits of these variables and constructs and their operational impact, please refer to the above mentioned publications. The entire survey questionnaire can be found in appendix B.

3.2.1 Four input factors

- **Warehouse size in m^2:** Measured on a ratio scale, capturing the floor space of the warehouse, including mezzanine floors, in 2012 and 2017 each.
- **Number of FTEs:** Measured on a ratio scale, including employees and temporary workers that are active in the warehouse on a full-time equivalent basis. Due to comparable labor laws and union presence in Netherlands and Belgium, the number of hours worked

per employee only deviates minimally. Direct workforce (active on the warehouse floor) and indirect workforce (i.e. not operational on the floor) are both included, as different levels of automation may require a different split of direct and indirect employees for the warehouse operations.

- **Number of SKUs:** Measured on a ratio scale as the average number of unique articles (SKUs) that are simultaneously stored in the warehouse, in 2012 and 2017 each.

- **Level of automation:** Measured on an ordinal scale, as the combination of an ordinal score for hardware automation and an ordinal score for software automation. Hardware automation is calculated based on how many out of 14 common automation technologies are employed in a warehouse. For the time comparison, each DMU also answered which automation additions it made during the last 5 years. Today's automation score is then calculated as the sum of the score 5 years ago plus all additions. Next to hardware automation, software automation is measured on a six-point ordinal scale, through a question about the warehouse's usage of information systems. The possible answers were: (1) No information system; (2) a standard ERP warehouse module; (3) a standard ERP warehouse module with more than 20% customization; (4) a standard WMS package; (5) a standard WMS package with more than 20% customization or (6) a tailor-made/customized system.

  Because there is no established way on how to weigh software and hardware automation, three different (Sexton's ratio) cross-efficiency implementations with distinct ways of calculating the overall automation score were compared to look at ranking impact of weight differences for software and hardware:

Method 1: The hardware and software scores are added, which allows DMUs to gather up to 28 points (14 in 2012 plus 14 additions over the last 5 years) for hardware automation and 6 for software automation. However, the mean hardware automation score in 2017 was 2.2 (1.3 in 2012), compared to 4 for software. On average, this case weighs software automation at 66%, but leaves ample room for DMUs to increase the hardware weight.

Method 2: Only the hardware score was considered, the software score was excluded.

Method 3: The hardware score and software score were both standardized, by dividing each score individually by the range of observations (Maximum score - minimum score) and then summing over both. Hence, the combined automation score lies between 0 and 2 for each DMU.

The Kendall's ranking correlation for all three methods in 2017 are shown in table 3.1 (2012 values in appendix F):

Automation Method Kendall's τ 2017	**Method 1**	**Method 2**	**Method 3**
Method 1	1***		
Method 2	0,74***	1***	
Method 3	0,86***	0,69***	1***

Table 3.1: Cross-efficiency ranking correlation for different automation score calculations 2017.

The rank correlation between both methods that include software as well as hardware is 0.86 in 2017 and 0.94 in 2012, both at 0.01 significance levels. The ranks, when the software component was excluded, correlate significantly lower with the other two methods.

Because of this robustness under different weighting schemes when both components are included, for the rest of this thesis, the approach summing over both automation components will be used - as it provides the highest freedom for DMUs to differentiate themselves through high levels of hardware automation.

3.2.2 Four output factors

- **Number of order lines:** Measured on a ratio scale. Refers to average daily order lines shipped per day during 2012 and during 2017. Through incorporation of the average, effects of order level seasonality and random fluctuations are mitigated.
- **Error-free order line %:** Measured on an nine-point ordinal scale, with the following values: (1) Not tracked (2) <90%; (3) 90-95%; (4) 95-97%; (5) 97-98%; (6) 98-99%; (7) 99.0-99.5%; (8) 99.5-99.9% and (9) >99.9%. Not tracking this metric (or not having the data available) is penalized, as error-free percentage is (one of) the main quality criterion in warehousing and not observing it renders most internal quality control efforts mute. By providing staggered levels of error-free order-lines, this model can differentiate between a wide array of error-free percentages.
- **Order flexibility:** Measured on a 30-point ordinal scale. To measure this construct, each respondent was asked whether the warehouse could cope with a total of six internal and external changes (1) much worse; (2) worse; (3) equal; (4) better or (5) much better than the competition. When "not applicable" was selected, the respective question was taken out consideration for that warehouse and the score was re-scaled. Re-scaling was achieved, by elimination the "not applicable" questions from a DMU's score and multiplying the remaining score by 6 divided by the number of questions that were applicable.
- **Special processes:** Measured on a 10-point ordinal scale, where respondents selected from a list of ten special, value-added processes that may be performed by the warehouse. The sum of a selections in each year was used as the score.

3.3 Efficiency research hypotheses

In this section, the research hypotheses concerning the cross-efficiency benchmark are introduced. The main hypotheses are related to automation's influence on warehouse performance, for which the author conjectures:

Hypothesis 1a *The correlation between the level of automation and warehouse efficiency is positive.*

Hypothesis 1b *The relative change in the level of automation over time, and the relative change in the warehouse's efficiency rank correlate positively.*

This proposition would be an intuitive explanation for the increase of warehousing automation investments in recent years (Wang, McIntosh, and Brain 2010). A similar propostion is made by Hamberg and Jacques (2012), who believe that as the technical hurdles for some of the most complex problems in warehousing are overcome (unstructured automatic item picking, autonomous vehicle roaming), a transformation toward automated logistics is likely to occur.

Furthermore, it will be analyzed, whether size related input factors: floor space (in m^2), the number of SKUs, or the number of FTEs play a role in efficiency. Because De Koster and Balk (2008) and Hackman et al. (2001) independently found small warehouses (measured in FTEs) to be more efficient, our hypotheses are as follows:

Hypothesis 2 *The correlation between the number of SKUs and warehouse efficiency is negative.*

Hypothesis 3 *The correlation between the number of FTEs and warehouse efficiency is negative.*

These two hypotheses are not built on the general assumption of decreasing returns of scale, as a newly added SKU is not inherently less efficiently handle-able than an existing one. Also, two FTEs can e.g. pick twice the number of order-lines than one FTE can. The assumption of lower cross-efficiency for higher SKUs is based on the fact that more SKUs may make it more difficult to have standardized processes, keep high turnovers of products and lower workers' familiarity with each product. Likewise, more FTEs may lead to a lower degree of identification with the company, lower social pressure, more formalized rules and less specific talent being hired.

However, as there is a constant shift towards facilities with larger footprints, designed to reap the benefits of economies of scale, not all size-related input factors are likely to negatively correlate with efficiency. A report from Onstein et al. (2016) found that especially in the Netherlands, "*the growing demand for very large DCs*" is a dominant phenomenon, most prominently for e-commerce logistics activities. Because there is such a strong trend towards bigger facilities, the hypotheses related to footprint are:

Hypothesis 4a *The correlation between the available floor space (in* m^2*) and warehouse efficiency is positive.*

Hypothesis 4b *The relative change in the available floor space (in* m^2*) over time, and the relative change in the warehouse's efficiency rank correlate positively.*

It is worth mentioning that floor space and SKUs might correlate, but are not logically tied to each other. A bottling warehouse next to a beverage factory will more likely use more floor space for the same number of SKUs than a consumer electronics warehouse. Also, a lower floor space usage for the same number of SKUs and the same type of warehouse could indicate better stock management/higher inventory turnover.

Similarly, floor space and FTEs will likely correlate as well, however there are multiple ways in which a warehouse can eliminate workforce requirements within the same footprint (smart usage of floor space, clever routing, efficient procedures, order batching, shift scheduling etc.), without changing its automation. Therefore, these input factors are initially considered separately, but it will be tested whether the three input factors truly are substitutes or complements of each other.

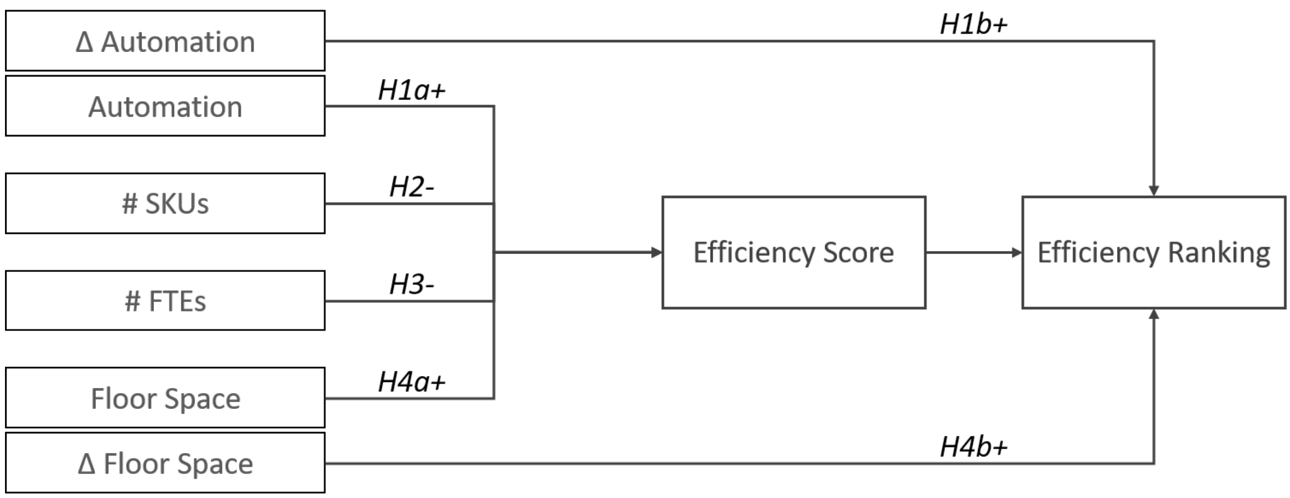


Figure 3.2: Conceptual model.

Certain scholars object to using a conceptual model, in which the dependent variable is calculated based on the explanatory variables of the same model. However, in order to obtain the cross-efficiency score, the dimensional inputs and outputs are transformed extensively, resulting in a dimensionless cross-efficiency score. Because of this transformation and loss of dimensionality, the relation of inputs and outputs to cross-efficiency scores is unproblematic for the conceptual model.

3.3.1 Clustering and sub-group analysis

The initial data set of respondents contains warehouses from all industries, all positions in the value-chain and all ownership types. These have varying effects on the comparability of DMUs. While e.g. foreign companies or logistic companies might result in different levels of management knowledge compared to regular Dutch commercial warehouses, the impact on the technology level is assumed to be marginal.

But the percentage of cold-storage and the warehouse's position within the company's value chain have a large impact on the achievable efficiency, as they are indirectly impacting the technology (Emmett 2005). For each sub-category cluster of warehouses with a significant number of respondents, this thesis provides a separate, individual cross-efficiency analysis. It is expected to observe larger cross-efficiency variations across the entire set, but due to market forces and competitive pressure relatively uniform scores within a sector.

3.4 Method comparison metrics

For the thesis' second objective, to distinguish the individual merits of four cross-efficiency methods, they will be analyzed in respect to the differences in implementation, usage and results. First and foremost, it will be determined whether the different result sets, obtained through the different techniques, are statistically different by performing a Wilcoxon signed rank test on the data. If they happen to be different, based on the metrics of table 3.2, recommendations will be provided on when to use which method.

Metric	Description
Methodological Proximity to DEA	Tests how close the cross-efficiency method follows the logic of the original DEA model. As the DEA efficiency scores are a natural upper limit for the cross-efficiency scores, methodological proximity makes the result more sound.
Implementational Ease	Tests how easily the method can be implemented and the degree of computational strain when solving a large data set with it. Especially, for application in non-academic fields, easy implementation and especially the swift calculation in case of automated or frequent application are relevant.
Extendability	Tests how many modifications to the basic model are described in literature, that allow tweaking the basic method to the requirements of an individual project.

Metric	Description
Discriminatory Properties	Tests how high the relative differences are between the DEA efficient *(CRS score of 1)* DMUs and how much of the overall efficiency range they are assigned to. As cross-efficiency is mainly applied to break the tie between various efficient DMUs, this is of utmost importance to a method's user.
Sensitivity to Changes of Scale	Tests how robust a method's results are to changes in input and output parameters. A low sensitivity to scale changes increases the validity of results in volatile environments.
Sensitivity to Erroneous Data	Tests how robust a method's results are to (random) changes in some DMUs inputs and outputs. A low sensitivity to erroneous data increases the validity of results, when data is subjective/opinion-dependent, based on estimates or exposed to possible human error.
Sensitivity to Dominant DMU Elimination	Tests how robust a method's results are to deleting a DMU that lies on a DEA efficiency hyperplane. A low sensitivity to deleting such DMUs increases the validity of results, especially for industry comparisons, when maverick DMUs might be part of the DMU set.

Table 3.2: Method comparison metrics.

Implementation of the discriminatory properties metric For the calculation of the discriminatory power of the methods, two separate approaches will be followed: Initially, the four methods will be applied in their standard form as described in the literature review. Afterwards, the cross-efficiency scores will be calculated, only factoring in the peer-appraisal scores of DMUs with a simple efficiency score of 1. Hereby, the intuition is that when calculating cross-efficiency scores based on all DMUs, each DMU's weights are valued equally. However, all non-efficient DMUs choose their weights based on the DMU(s)' hyperplane(s) they are dominated by (Fried et al. 2008). Consequently, counting every non-efficient DMU's weights in reality only counts the dominating DMU's weights over and over again. To abstract from this behavior, all non-efficient DMUs will be eliminated from the sample for the second approach, thereby only allowing each hyper-plane to impact the cross-efficiency score once.

4. Data

4.1 Data collection process

This chapter introduces the reach-out- and response collection process as well as the final set of respondents and the correlations between inputs and outputs. In short, the mailing lists of evofenedex, TLN and a third company that wishes not to be disclosed as well as contact data of previous research in the warehousing sector were used to gather respondents. Additionally, extensive personal outreach via Linkedin, RSM's alumni network and online search was conducted. Each contacted individual was introduced to the subject of the study and incentivized by offering them an individual report of their facility's performance compared to the competition. A sample report can be found in appendix C. All questionnaires were administered through *Qualtrics*, an online surveying platform, where the survey was accessible in Dutch and English.

In total, 1,827 individuals were contacted, which resulted in 214 submissions, out of which 131[1] were entirely completed and use-able. The respective response rate is 11.7% (all respondents) and 7.2% (use-able respondents), which is lower than the response rates Muilerman (2001) identifies for studies in the logistics sector. However, Muilerman (2001) was published before the advent of the internet and is therefore not directly comparable to online-administered surveys as in this case. Sauermann and Roach (2013) find in their study that the low costs of surveying with online-tools, has lead to "*oversurveying*", reducing the achievable response rates below earlier levels.

Additional reasons, possibly explaining the low response rate are:

- The high share of individuals reached out to by the three collaboration companies via their mailing lists, which naturally is a contact channel to which warehouse managers devote little attention.

- The non-existence of a database of warehouses in the Netherlands and Belgium, which requires manual investigation and informal outreach via e.g. Linkedin or cold-calling, which results in a less convincing first impression on a warehouse manager than e.g. a bureau of statistics request would have.

- The author's inability to speak Dutch, which made all forms of personalized contact less convenient for the respondents and especially cold-calling less feasible.

- The limited time-frame for data-collection of nine weeks, intersected by public holidays, rendering long-term follow-up or avoidance of busy weeks more difficult as well as ruling out paper-based mailings, given their long response times.

The three most often cited reasons (aside from not reacting in general) were warehouse managers being: 1) too busy 2) not allowed to share the required data 3) not interested in external benchmarking, with the first stated reason accounting for over 90% of reactions.

Detailed description of outreach Given the high share of this thesis' overall workload dedicated to data collection, this section will depict the outreach process in more tangible detail. The subsequent paragraphs may be of special interest for readers, contemplating the

[1] Out of the 131 complete respondents, 102 were used for the analysis in this thesis. The remaining 29 respondents were eliminated because they employed less than 5 FTEs, making them extreme observations and therefore hardly comparable to full-size warehouses.

pursuit of similar research, as for most publications the data-collection is summarized in few sentences, which does not do the required effort for this topic justice.

In order to contact the 1,827 individuals mentioned above, several methods were used:

1. Three data sets from publications and theses from 2009-2015 were obtained, with information dating back to 2004. Over 75% of that contact information was outdated, but person-by-person Linkedin- as well as online follow-ups made it possible to contact approximately 70% of respondents via phone or mail, some of which had transitioned into non-warehouse positions. In combination this method yielded 626 contacts, out of which 432 were reach-able via phone or mail.

2. The author sent 3,173 LinkedIn-network requests to warehouse professionals and site managers in the Netherlands and Belgium, out of which 643 individuals accepted. Each of the LinkedIn contacts then received a message including the survey information and a sample output of the individual report as motivation plus up to three follow-ups in weekly intervals, if they did not react. Including correspondence with contacts, over 6,000 messages were sent. Minus the overlap with the first method, 528 individuals were contacted via LinkedIn.

3. Approximately 110 Dutch supply chain master students were asked to inform their personal network about the study, as connections with a personal background tend to have the highest conversion. This resulted in 28 new connections.

4. The supply-chain-focused websites logistiek.nl, logistiekProfs.nl and logistiektotaal.nl each published articles about the study, for which the total view count cannot be established.

   These four methods combined yielded 105 of the 131 respondents. 55 of these contacted individuals were excluded from the response count, as they replied to not be working a warehouse related function any longer.

5. Evofenedex reached out to 391 members individually and to over 28,000 contacts via a general knowledge newsletter, a total of four times over the course of 6 weeks, out of which 284 looked at the survey and 10 filled it out entirely. TLN contacted 575 unique members, three times, over the same period, resulting in 15 complete respondents. The third company messaged 35 warehouse managers internally twice, with a response rate of 3%.

Overall, it has to be noted that the broad outreach via several channels, especially using the mailing lists and digital newspapers, creates noise in estimating the overlap of methods and thereby might underestimate the actual response rate by multi-counting contacts.

4.2 Response set

Out of the 131 entirely submitted surveys, 29 submissions were eliminated, because the respective companies employed less than 5 FTEs, thus the final response set consists of 102 warehouses. This was done, as De Koster and Balk (2008) and Faber, De Koster, and Smidts (2013) find very small warehouses to be more efficient than other facilities. Since businesses of 1-4 individuals are hardly comparable to full-sized warehousing operations (and in their nature extreme observations), they skew the DEA results and, if included, negatively affected the efficiency comparison.

One can find several guidelines in the DEA literature about the required minimum number of DMUs for good discriminatory power of the analysis, given homogeneity of the sample (Avkiran

2006). This data set fullfils the agressive rule-of-thumb by Dyson et al. (2001) to aim for a minimum number of DMUs equal to the number of inputs multiplied by the number of outputs times two, or 32 in this case ($4 \times 4 \times 2$), while Golany and Roll (1989) threshold recommendation is only twice the sum of input and output factors, or 16 in this case ($(4+4) \times 2$). This issue will be revisited for the clustering discussion.

Out of the 102 warehouses that comprise the final response-set, 82 are located in the Netherlands, 17 in Belgium and 3 in Germany. They store 12 different product categories (all 13 selectable categories, except military/defense), employ over 6,000 FTEs and store 2.2 million SKUs on over 1.8 million square meters, of which 25% are cold-storage. More details, the size distributions, and a clustering are provided in the following subsections.

4.2.1 Product categories and industry breakdowns

Each participant was allowed to select up to two product categories that best classify the warehouse's products. The 102 warehouses selected 145 categories, with logistics, consumer goods and groceries/food stated most often, classifying half of the warehouses. Although logistics exhibits the highest category share, it is also the category which was most often co-selected together with another product category.

A detailed split of the product categories is provided in figure 4.1.

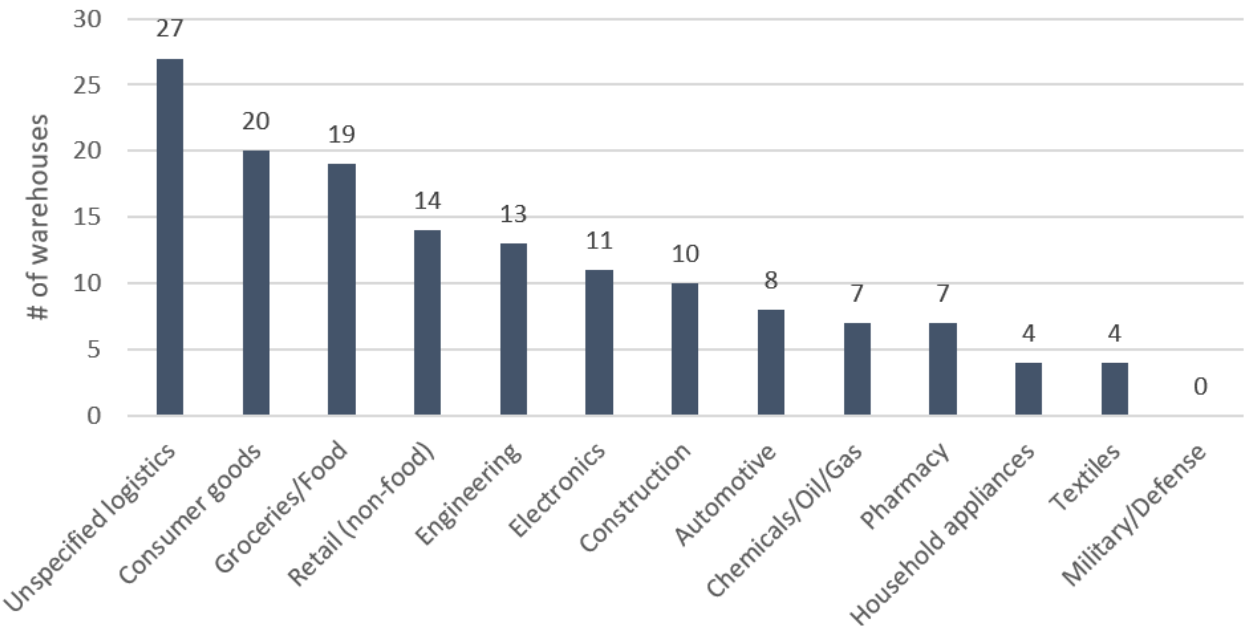


Figure 4.1: Product category split of warehouses in the sample.

4.2.2 Value chain position and ownership breakdowns

57% of warehouse (58/102) are owned and operated in-house, a third (34/102) is operated by a third party logistics provider *(3PL)* in a warehouse facility offering services to multiple customers and only 10% of warehouses (10/102) are dedicated and operated through a provider. 60% of warehouses (61/102) are wholesale warehouses, shipping to B2B partners, 26% (27/110) are production warehouses. Facilities are considered production warehouses, if the production of the products that are stored within the warehouse are on the same premises. This includes raw material and components as well as storing finished products for further distribution. 14% of warehouses (14/102) are retail facilities with direct end-customer contact.

For a matrix overview of the two dimensions, see table 4.1.

Value Chain Matrix	In-House	3PL - Dedicated	3PL - Multiple	Sum
Production Warehouse	19	0	8	**27**
Wholesale Warehouse	31	6	24	**61**
Retail Warehouse	8	4	2	**14**
Sum	**58**	**10**	**34**	**102**

Table 4.1: Value chain position and operations provider of warehouses in the sample.

4.2.3 Size breakdowns

For every size metric the average has grown over the past 5 years. The range of warehouse FTEs increased from 5-312 in 2012 to 5-350 in 2017, floor space (in m^2) ranged from 400-275,000 in 2012 and from 500-275,000 in 2017. In appendix D, the change in floor space size between 2012 and 2017 is depicted, by color-coding the size of a warehouse in 2012 and observing its size development until 2017. Figure 4.3 shows the floor space sizes, without tracking the size shifts on an individual facility level.

For SKUs and order lines, the relative differences in size were larger across the sample, which is explainable through the different nature of e.g. an aircraft spare-part warehouse maintaining stock of all parts over 40 years, versus a fresh-fruit importer, where no item is stored for longer than 24 hours. The SKU range moved from 100-250,000 in 2012 to 100-400,000 in 2017 and order lines from 25-55,000 in 2012 to 54-55,000 in 2017. Detailed, graphical representations of the sample's size metrics can be found in figures 4.2-4.5.

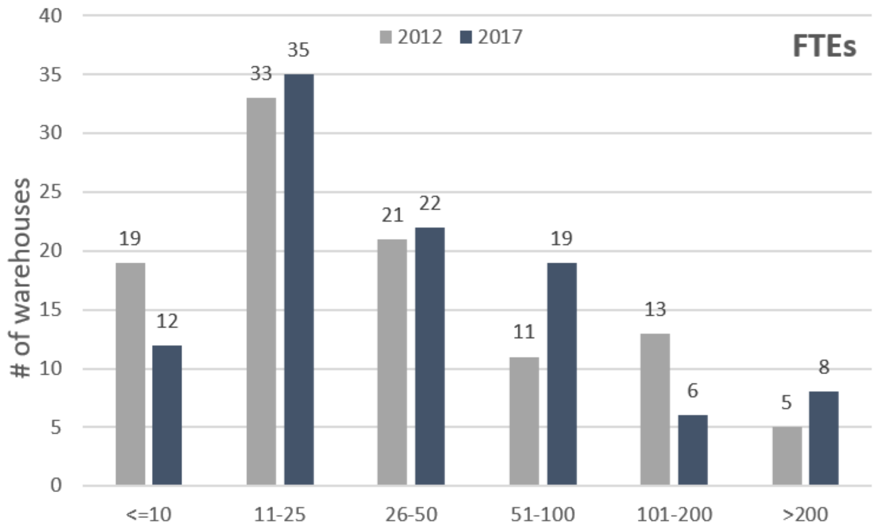


Figure 4.2: Number of FTEs across the warehouse sample in 2012 and 2017.

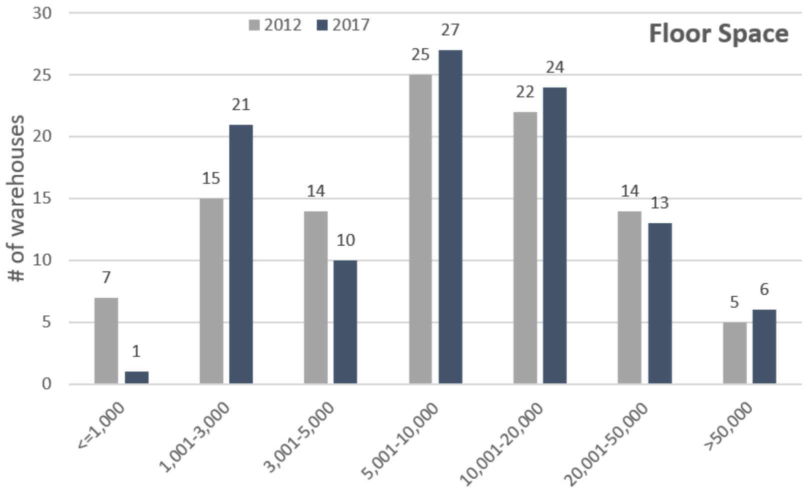


Figure 4.3: Floor space (in m^2) across the warehouse sample in 2012 and 2017.

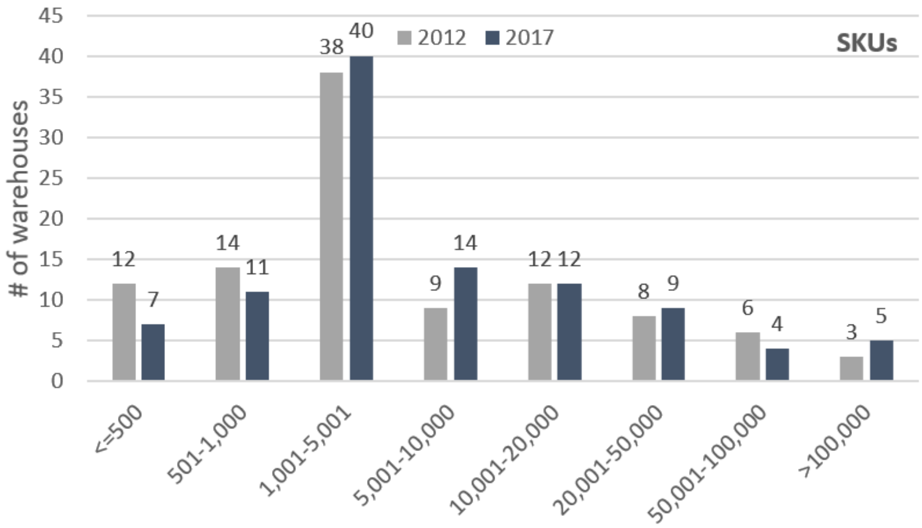


Figure 4.4: Number of SKUs across the warehouse sample in 2012 and 2017.

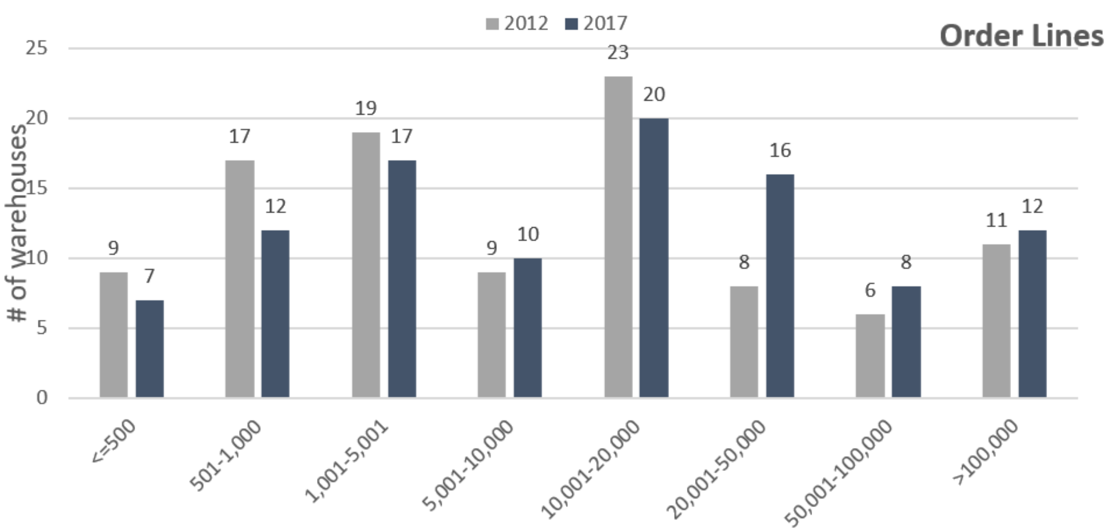


Figure 4.5: Number of order lines across the warehouse sample in 2012 and 2017.

Graphical outputs for the automation score input and the output scores of error free %, order flexibility and special processes can be found in appendix D.

4.3 Clustering of data set

The clustering of warehouses into subsets is done to obtain insights on how efficiency patterns within industries distinguish themselves from when comparing efficiency across multiple sectors. Because of the limited number of observations per industry, no universal applicability for these analyses can be claimed. Nevertheless, this drill-down and the intermediate juxtaposition of homogeneous DMUs was the most requested feature by participating managers.

Taking into consideration the recommendatios for the minimum number of DMUs in a DEA analysis in section 4.2, only clusters are formed with approximately 20 or more observations. This strikes a balance between sufficient discrimination of units and homogeneity of DMUs. One of the three clusters is a combination of two product categories with comparable operations.

The three clusters are:[2]

1. Engineering + Construction *(21 warehouses)*
2. Consumer goods *(20 warehouses)*
3. Groceries/Food *(19 warehouses)*

The analyses in chapter 5 will initially be performed for the entire set and then for each of the clusters individually.

4.4 Correlation of inputs and outputs

The subsequent sections will investigate the correlation between inputs and outputs. The detailed input and output data can be found in appendix E, descriptive statistics for 2017 in the tables 4.2 and 4.3. For a DEA model it is desirable to have inputs that are uncorrelated among each other and outputs that positively correlate with at least one input.

Inputs 2017	**FTEs**	**Floor Space (m^2)**	**SKUs**	**Automation Score**
Minimum	5	500	100	2
Median	30	9.250	4.600	6
Average	59	18.244	21.088	7
Maximum	350	275.000	400.000	16
Std. Dev.	74	32.414	57.393	3

Table 4.2: Descriptive statistics of response set's inputs 2017.

[2]Two DMUs selected both categories, engineering and construction, but only count once each for the joint cluster.

Outputs 2017	**Order Lines**	**Special Process Score**	**Error Free % Score**	**Order Flexibility Score**
Minimum	54	2	1	12
Median	1.200	6	7	22
Average	4.931	6	6	21
Maximum	55.000	10	9	30
Std. Dev.	9.815	2	2	4

Table 4.3: Descriptive statistics of response set's outputs 2017.

For better readability from this point onwards, the inputs will be referred to as FTEs, Floor Space, SKUs and Automation, without stating their measurement scale. Respectively, the outputs will be referred to as Order Lines, Special Processes, Error Free % and Order Flexibility.

4.4.1 Correlation of inputs among each other

To test the ranking correlation, Kendall's τ *(tau)* is calculated for all method pairs. This method was chosen, instead of Spearman's ρ *(rho)*. Both are non-parametric methods to test for ranking correlation, but Kendall's method is more conservative, exhibits more statistical robustness and is more easily interpret-able (Arndt, Turvey, and Andreasen 1999). Kendall's τ signifies how identical the ranks of two paired populations are. A score of 1 stands for perfect correlation whereas a score of -1 represents entirely opposite rankings.

Out of the six relevant pairs to check for correlation (all combinations of selecting two of the four inputs without repetition), only the FTEs & floor space pair (0.47 in 2017 and 0.46 in 2012) shows rank correlations above 0.2 in both years. These correlations are explainable, because larger facilities usually employ more workforce. As the rest of the input pairs are less, or not at all, correlated and the model has been tested in the past, no adjustment (input elimination) to the model is made. The 2017 input correlation coefficients[3] can be found in table 4.4, the 2012 input correlation coefficients in appendix F.

Kendall's τ 2017	**FTEs**	**Floor Space**	**SKUs**	**Automation**
FTEs	1			
Floor Space	0,47***	1		
SKUs	0,19***	0,08	1	
Automation	0,32***	0,17**	0,22***	1

Table 4.4: Input factors' Kendall's τ correlation coefficients 2017.

4.4.2 Correlation of outputs among each other

For the six output pairs, none exhibits a rank correlation above 0.2 in both periods. The highest rank correlation across 2012 and 2017 is error free % and order flexibility, which could indicate that a warehouse with resilient and effective processes is able to be react flexibly to changes as well as ship with a high degree accuracy. The number of order lines and order flexibility exhibited the the highest rank correlation of 0.16 in 2017. The output factor correlation coefficients deem no changes to the model necessary. The 2017 output correlation coefficients are shown in table 4.5, the 2012 correlations in appendix F.

[3] All p values in this thesis are indicated as follows: *$p < 0.10$; **$p < 0.05$; ***$p < 0.01$.

Kendall's τ 2017	Order Lines	Special Processes	Error Free %	Order Flexibility
Order Lines	1			
Special Processes	0,14*	1		
Error Free %	-0,02	0,04	1	
Order Flexibility	0,16**	0,06	0,15**	1

Table 4.5: Output factors' Kendall's τ correlation coefficients 2017.

4.4.3 Correlation of inputs with outputs

For the correlation between inputs and outputs, the main criterion to look for is that each output is at least positively correlated with one input and that negative correlations are minimal, which signifies an underlying model in which a DMU uses or transforms the selected inputs into selected outputs.

The results of this study are in line with findings of prior studies. The highest positive correlation is found for order lines and FTEs, which is explainable, considering that the main objective of a warehouse is order shipment and employees are hired to perform this task. All outputs are positively correlated with at least one input in 2012, and only error free % is not significantly correlated with any input in 2017. No significant negative rank correlations are observed in either 2012 or 2017. The order of magnitude and sign of all significant correlations in 2017 is similar to 2012. The 2017 correlation coefficients are shown in table 4.6, the 2012 coefficients can be found in appendix F.

Kendall's τ 2017	Order Lines	Special Processes	Error Free %	Order Flexibility
FTEs	0,45***	0,13*	-0,03	0,1
Floor Space	0,23***	0,14*	-0,01	0,1
SKUs	0,28***	0,08	0,11	0,1
Automation	0,24***	0,19**	0,05	0,25***

Table 4.6: Input and output factors' Kendall's τ correlation coefficients 2017.

5. Analysis

5.1 DEA results

Before performing the cross-efficiency calculations, the regular DEA was run to obtain the DMU's efficiency scores that limit the weight selection process of the cross-efficiency methods.

5.1.1 DEA constant returns to scale results

Under constant returns to scale assumptions, 26 DMUs are considered efficient in 2017 and 35 in 2012, with average efficiency scores of 0.681 (2017 - standard deviation 0.25) and 0.751 (2012 - standard deviation 0.25). In combination, with the slightly increased minimum efficiency score (0.235 in 2017 vs. 0.232 in 2012), this indicates a moderate shift towards similar input and output mixes and less extreme efficiency scores. This transformation leads to a smaller range of scores and less "maverick" DMUs that receive unity scores, thereby decreasing the average cross-efficiency score. Using these scores however, one cannot conclude how warehouse efficiency itself has developed, as the CRS DEA scores are dependent on the distance from the efficiency frontier, which in turn is highly dependent on the shape of the PPS itself. Score distribution plots for both CRS DEAs can be found in appendix G.

Out of the 26 efficient DMUs in 2017, 11 act as peers for 10 or more DMUs[1], while other DMUs (e.g. DMU006 or DMU040), were not used as a peer by any other DMU. This lets one infer that the operations (the input and output combinations) of the DMUs that are most often selected as peers are more representative. At the same time, one could consider DMUs with very few selections as peer "rogue" DMUs with limited relevance for the rest of submissions:

DMU Number	CRS 2017	Peer Selection
DMU006	1,00	1
DMU008	1,00	7
DMU024	1,00	9
DMU027	1,00	19
DMU028	1,00	37
DMU040	1,00	1
DMU041	1,00	4
DMU045	1,00	6
DMU049	1,00	48
DMU050	1,00	38
DMU052	1,00	11
DMU063	1,00	17
DMU066	1,00	15
DMU067	1,00	9
DMU071	1,00	6
DMU098	1,00	51
DMU099	1,00	9
DMU100	1,00	11
DMU104	1,00	10
DMU106	1,00	5
DMU107	1,00	6

[1]The DMU numbers are based on the 131 DMU data set that still included small (<5 FTEs) warehouses. Because of this, the numeration exceeds 102, although the analyses were performed for 102 warehouses.

DMU Number	CRS 2017	Peer Selection
DMU108	1,00	1
DMU115	1,00	7
DMU118	1,00	2
DMU125	1,00	1
DMU128	1,00	13

Table 5.1: Efficient DMUs and number of selections as peers - CRS 2017.

The 2012 CRS peer selection table can be found in appendix H.

5.1.2 DEA returns to scale calculations

Although it was mentioned in the literature review that the cross-efficiency methods will be run under the CRS assumption, it is still meaningful to study the returns to scale, observed by the individual DMUs. The main reason why VRS is rejected for the later analysis is its over-proportional high share of efficient DMUs. Despite cross-efficiency's properties to break the tie, even among multiple DMUs, the increased efficiency scores through VRS skew the cross-efficiency rankings and may lead to inadvertent results such as negative cross-efficiency scores. Although the implementation of VRS models in cross-efficiency is subject of research, its extent is limited and still in an early and experimental stage (Lim and Zhu 2015).

This section explores how returns to scale can be evaluated, for which Banker and Thrall (1992) show that the returns to scale of DMUs can be determined based on the lambdas of the CRS solution. These lambdas are obtained from solving the primal of the dual linear DEA program described in section 2.3.1. The lambdas indicate which peers a DMU has selected. Each DMU$_j$ has K lambdas associated with it, one for each DMU$_k$ that could be one of DMU$_j$'s peers. Hence, most lambdas for a given DMU$_j$ will be 0, indicating that this DMU$_k$ does not act as a peer. Contrarily, for a DMU that is used as a peer, that lambda will be positive. All lambdas combined then form the position on the PPS that the DMU is projected to. For an efficient DMU this means that it uses only itself as a peer and therefore has a sum of lambdas of 1. As in a CRS model, a DMU's sum of lambdas is not constrained, DMU$_j$'s possible returns to scale can be distinguished as shown stated in Fried et al. (2008):

- If $\sum_{k=1}^{K} \lambda_{jk}^{*} = 1$ for at least one optimal solution, then CRS holds locally at DMU$_j$.
- If $\sum_{k=1}^{K} \lambda_{jk}^{*} > 1$ for all optimal solutions, then DRS holds locally at DMU$_j$.
- If $\sum_{k=1}^{K} \lambda_{jk}^{*} < 1$ for all optimal solutions, then IRS holds locally at DMU$_j$.

While this method can be used as an approximation for all DMUs, Banker and Thrall (1992) prove this theorem only for DMUs on the efficiency frontier. In order to definitively expand the RTS categorization to all DMUs, Banker, Chang, and Cooper (1996) introduce a linear program. Its objective function seeks to minimize the sum of lambdas for each DMU$_j$ with $\sum_{k=1}^{K} \lambda_{jk}^{*} > 1$ in its optimal CRS DEA solution. In its DEA optimal solution, this would imply DRS, yet the algorithm checks, whether there are alternative solutions, resulting in the same efficiency, but with a sum of lambdas equal to 1 (and therefore CRS, rather than DRS). Next to the sum of lambdas, the objective function for DMU$_j$ also incorporates an infinitesimal weight ϵ for the slacks, so in the case of two solutions with identical sums of lambdas, the one with more slacks (s^{-}, s^{+}) is used:

$$\min \quad \sum_{k=1}^{K} \lambda_{jk} - \epsilon \left(\sum_{n=1}^{N} s_n^{j-} + \sum_{m=1}^{M} s_m^{j+} \right) \tag{5.1}$$

This is subject to the constraint that all inputs and outputs are identical to the DEA CRS solution, as well as all lambdas being positive and the sum of lambdas greater or equal to one:

$$\theta_j^* x_n = \sum_{k=1}^{K} x^k \lambda_{jk} + s_n^{j+} \qquad n = 1, ..., N, \tag{5.2}$$

$$y_m = \sum_{k=1}^{K} y^k \lambda_{jk} + s_m^{j-} \qquad m = 1, ..., M, \tag{5.3}$$

$$1 \leq \sum_{k=1}^{K} \lambda_{jk}, \tag{5.4}$$

$$0 \leq \lambda_{jk} \qquad k = 1, ..., K \tag{5.5}$$

To test, whether IRS in the DEA solution hold ($\sum_{k=1}^{K} \lambda_{jk}^* < 1$), one needs to adjust the objective function (5.1) in the following way (Cooper, Seiford, and Tone 2007):

$$\max \quad \sum_{k=1}^{K} \lambda_{jk} + \epsilon \left(\sum_{n=1}^{N} s_n^{j-} + \sum_{m=1}^{M} s_m^{j+} \right) \tag{5.6}$$

Also, the sum of lambdas are constrained to be smaller or equal than 1.

5.1.3 DEA returns to scale prevalence

This section uses the methods introduced in section 5.1.2 to investigate the returns to scale distribution in the warehouse data set. Both methods introduced above find the same distribution of returns to scale. This leads one to conjecture that given the multi-dimensional PPS and the high number of DMUs in the sample, the degrees of freedom for a DMU to choose its peers to reach optimal efficiency are low.

The sum of lambdas for the efficient DMUs in 2017 and 2012 indicate the following returns to scale:

DMUs' Returns to Scale	**2012**	**2017**
Increasing RTS	25	22
Decreasing RTS	42	54
Constant RTS	35	26
Total Observations	**102**	**102**

Table 5.2: Warehouse sample's returns to scale in 2012 and 2017.

In both years IRS and DRS are found within the sample, although the more recent data exhibits a higher number of decreasing returns to scale facilities. If one loosens the condition for constant returns to scale to occur for sums of lambdas between 0.9 and 1.1, this would increase the number of CRS warehouses by 22 (in reality exactly constant returns to scale of 1 are hardly found). As however, in the next section the RTS will be analyzed in comparison to warehouse size, the equality constraint ($\sum \lambda = 1$) was chosen for better discrimination among warehouses.

To convey an idea of the returns to scale in the sample, table 5.3 shows the operationally most efficient range of inputs for warehouses, i.e. the range below which only warehouses with increasing returns to scale were found and above which only warehouses with decreasing returns to scale were found. This helps to understand, at which scale of inputs warehouses tend to be most efficient, disregarding the operational strategies they employ:

Optimal Input Ranges	2012 Min	2012 Max	2017 Min	2017 Max
FTEs	8	115	10	243
Floor Space	1.500	26.000	1.260	275.000
SKUs	700	250.000	850	17.600
Automation	3	10	4	14

Table 5.3: Range of optimal inputs scales for 2012 and 2017.

Table 5.3 shows that a wide variety of set-ups are possible to achieve optimal scale size, yet there are thresholds on floor-space, assortment size as well as the level of automation. For FTEs, almost the smallest and largest facilities were within the optimal range.

5.1.4 Decreasing returns to scale for larger warehouses

In the preceding section, it was shown that all return to scale types are present in the sample, based on which a range for optimal input size was provided. Also, it could be shown that decreasing returns to scale facilities have increased in number and that overall efficient warehouses have decreased. This section will therefore investigate, where in the data set which returns to scale occur and whether the trend of increasing warehouse size can explain the observed shift in RTS over time. Previously, studies found large warehouses to be less efficient than small warehouses (de Koster and Balk 2008), but one has to distinguish between two factors that may drive this efficiency gap. First, decreasing returns to scale could be at play, which would make it impossible for larger warehouses to become CRS efficient. Second, operational inefficiencies could be the reason that large warehouses are less efficient.

When the returns to scale, obtained through Banker, Chang, and Cooper (1996) method, are plotted in figure 5.1 with the three inputs FTEs, floor space and SKUs as axes, one can see the decreasing returns to scale for larger warehouses:

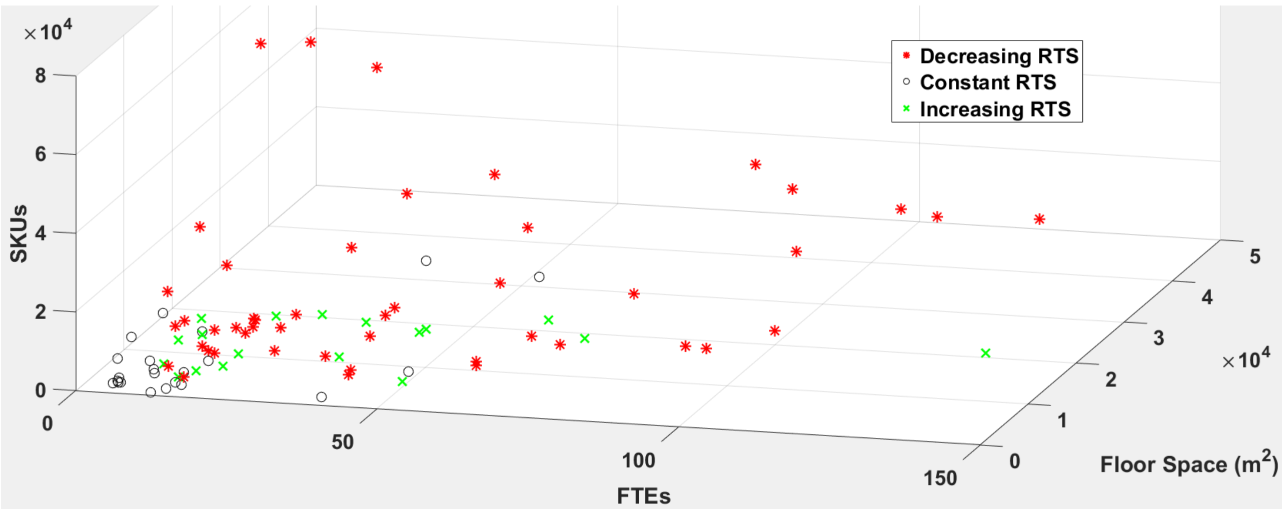


Figure 5.1: Returns to scale compared to input size.

All increasing returns to scale DMUs are found in the vicinity of the origin, while all larger DMUs exhibit DRS. One warehouse, with relatively large FTE headcount and moderate floor space as well as few SKUs is the exception and observes IRS. In the interest of distinguishability, the axes were cut of at 150 FTEs, 50,000m^2 and 80,000 SKUs, which excludes 15 DMUs, 7 of which with decreasing and 4 with increasing returns to scale. Out of the 4 increasing returns to scale warehouses, 3 employ very similar input and output mix, with high FTE-counts, but relatively low other inputs, all with CRS efficiency scores of less than 0.5. Aside from this particular mix of inputs and outputs, all other combinations are subject to decreasing returns to scale, which confirms the intuition by scholars in the field.

5.1.5 DEA variable returns to scale

After providing empirical evidence of the DRS for larger warehouses, it is meaningful, to gauge the magnitude of the resulting scale inefficiency (inefficiency, not attributed to inefficient input usage, but simply sub-optimal input size) in the sample. Next to operational performance, this is the other element impacting efficiency. To calculate scale efficiency, one needs to divide the CRS DEA score of a DMU by its VRS score (Cooper, Seiford, and Tone 2007). Consequently this section introduces the VRS results.

Under VRS, 46 DMUs are considered efficient in 2017 and 51 in 2012, with average efficiency scores of 0.781 (2017 - standard deviation of 0.24) and 0.817 (2012 - standard deviation of 0.23). Also, under VRS it is found that the minimum efficiency score of the sample has increased from 0.239 to 0.248. As with VRS more DMUs are efficient, there are less DMUs that select dominant peers. Because of this, the absolute number of times a DMU acts as a peer in the VRS case, compared to CRS, is reduced - but the same DMUs are found to be most often selected as peers in both cases:

DMU Number	**VRS 2017**	**Peer Selection**
DMU006	1,00	2
DMU008	1,00	6
DMU023	1,00	2
DMU024	1,00	5
DMU027	1,00	20
DMU028	1,00	21
DMU033	1,00	2
DMU034	1,00	2
DMU040	1,00	17
DMU041	1,00	1
DMU045	1,00	8
DMU049	1,00	25
DMU050	1,00	35
DMU052	1,00	6
DMU055	1,00	2
DMU058	1,00	2
DMU059	1,00	3
DMU061	1,00	1
DMU062	1,00	1
DMU063	1,00	7
DMU064	1,00	1
DMU066	1,00	9
DMU067	1,00	9
DMU071	1,00	4
DMU089	1,00	1

DMU Number	VRS 2017	Peer Selection
DMU091	1,00	5
DMU098	1,00	35
DMU099	1,00	11
DMU100	1,00	2
DMU101	1,00	2
DMU104	1,00	7
DMU105	1,00	1
DMU106	1,00	11
DMU107	1,00	8
DMU108	1,00	1
DMU115	1,00	9
DMU116	1,00	2
DMU117	1,00	1
DMU118	1,00	3
DMU122	1,00	2
DMU123	1,00	3
DMU124	1,00	1
DMU125	1,00	2
DMU126	1,00	2
DMU128	1,00	11
DMU130	1,00	2

Table 5.4: Efficient DMUs and number of selections as peers - VRS 2017.

The 2012 VRS peer selection table can be found in appendix H.

5.1.6 Inefficiencies of scale

As mentioned above, to understand the extent of scale inefficiencies impacting the efficiency scores, one draws on the ratio of CRS to VRS efficiency score. The average scale efficiency in 2017 was 0.88 with a standard deviation of 0.17 (0.92 and 0.13 in 2012). The full table can be found in appendix I. The average CRS-, VRS- and scale efficiency scores for both years are shown in table 5.5.

Average Efficiency Score	CRS DEA	VRS DEA	Scale Efficiency
2012	0,75	0,82	0,92
2017	0,68	0,78	0,88

Table 5.5: Average DEA efficiency scores by model and scale efficiency scores.

Based on these values, scale inefficiency is a considerable factor in the overall inefficiency of DMUs. The 0.68 overall CRS efficiency in 2017 is caused by inefficient operations (0.78) and scale inefficiencies (0.88). Because under VRS, each DMU can choose its own optimal scale, any VRS inefficiency results from operational aspects. Likewise, the ratio of CRS/VRS reflects scale efficiency. In 2012, scale efficiency is slightly higher (0.92) and operational efficiency at 0.82. Based on this sample, the inefficiency split between operational and scale inefficiency is approximately 2:1 ((1-0,78):(1-0,88)), a ratio that has not been computed previously for studies applying this input-output DEA model. In other words, a third of the warehouses' inefficiency in this sample stems from operating at non-efficient scales.

5.2 Statistical comparison of four cross-efficiency methods

After analyzing inputs and outputs correlations, DEA results and the returns to scale, the following sections will investigate the cross-efficiency results under the constant returns to scale assumptions.

Running the four cross-efficiency methods (one ratio and one classic implementation of Sexton, Silkman, and Hogan (1986), a multiplicative approach following Cook and Zhu (2014) and Liang et al. (2008) game theory method) on the data set leads to different cross-efficiency score distributions. All models were run in their benevolent setting, or aim to maximize the peer-appraisal scores, which makes results most comparable and also follows the market intuition that competitors maximize their own efficiency given a set of weights and constraints.

While the two Sexton-derived methods resulted in scores for the 2017 data of the same order of magnitude per DMU (as expected given the similar models), the obtained cross-efficiency scores by the multiplicative approach are significantly lower (0.360 and 0.362 average for the first two models, vs. 0.026 for the multiplicative approach with standard deviations of 0.177, 0.181 and 0.111).

Cross-efficiency 2017	**Sexton_Classic**	**Sexton_Ratio**	**Multiplicative**	**Game Theory**
Minimum Score	0,090	0,087	0,000	0,164
Average Score	0,360	0,362	0,026	0,535
Maximum Score	0,894	0,890	0,911	1,000
Std. Dev. Score	0,177	0,181	0,111	0,219

Table 5.6: Results comparison of cross-efficiency methods 2017.

The game theory approach resulted in the highest average score of 0.535 (standard deviation of 0.219) and finished after 32 iterations. In a preliminary run with 80 warehouses, two warehouses were found with cross-efficiency scores of 1. This is surprising, as cross-efficiency is designed to break the tie among efficient DMUs, but as introduced in the literature review, this method does not follow the classic cross-efficiency path. Rather, it iterates through "rounds" in which the DMU find Nash-equilibria for their weights and efficiency scores until those converge over time. Because in the preliminary set, the two DMUs employed mutually exclusive inputs in their optimal solution (and consequently had zero-weights on all inputs the other DMU uses), it is conceivable to find two "efficient" DMUs. Although this did not happen for the final data-set it poses a limitation to the method.

Cook and Zhu (2014) present a calculatory example with only ten DMUs, in which the two highest ranked DMUs appear to have efficiencies of 1.000 and 0.999. When recalculating these values, the MATLAB code used for this thesis found both DMUs from the example to have a cross-efficiency of one. The author conjectures that because of the pairwise equilibria-search, two DMUs using different inputs or outputs can both be found to be cross-efficient.

The results of the comparison for the 2012 data can be found in appendix J.

5.2.1 Normal distributions

After the comparison of magnitude and range of the cross-efficiency scores, the normal distribution of scores is tested for, using one-sample Kolmogorov-Smirnov *(KS)* tests. For all four cross-efficiency methods, the normal distribution assumption can be rejected at $p<0.01$. The KS values, when rounded to 4 decimals were 0.0000 for all four cross-efficiency scores for both periods. The score distributions can be inspected in figure 5.2. Most scores of the multiplicative approach are below 0.01, which most likely is due to the exponential nature of the calculations, heavily emphasizing efficient DMUs over inefficient ones, which leads to stark score differences compared to additive DEA calculations.

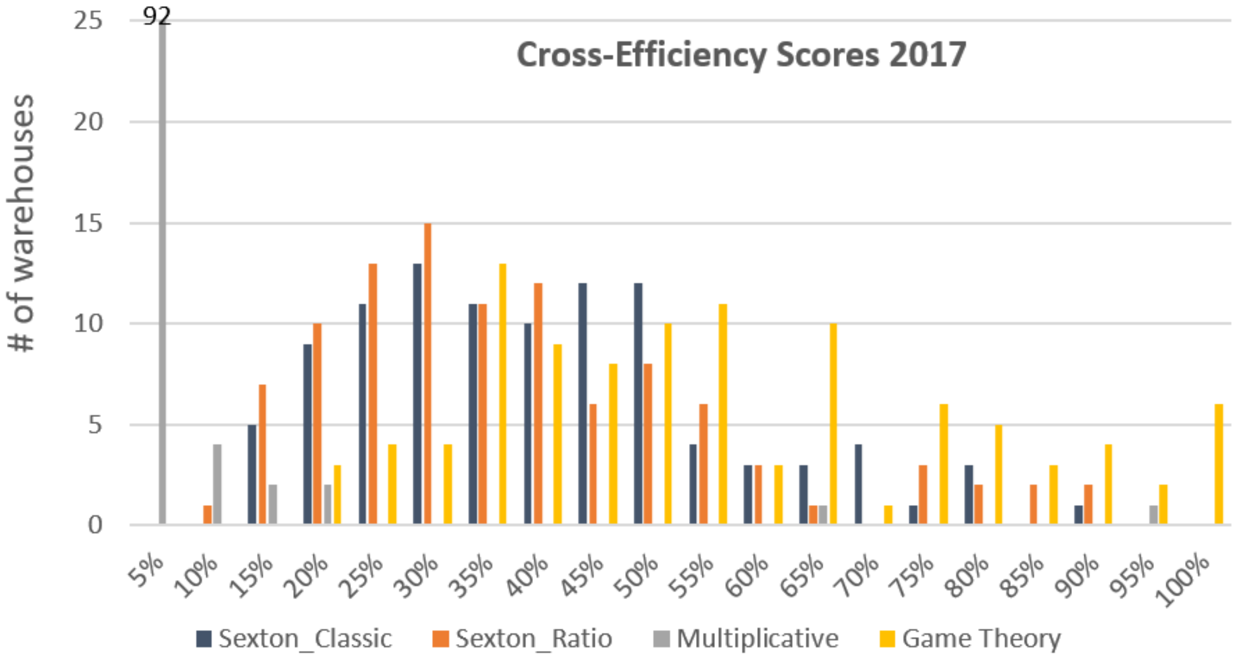


Figure 5.2: Cross-efficiency score distribution 2017 per method.

The 2012 score distribution per method can be found in appendix J.

5.2.2 Score similarity

After comparing magnitude and distribution of the cross-efficiency scores, it is imperative to test, whether the four methods result in the same scores, which is done using the Wilcoxon signed rank *(WSR)* test that checks whether the median difference between two samples is zero.

The 2012 results at a 0.05 significance level are stated in table 5.7.

WSR Results 2012	**Sexton_Classic**	**Sexton_Ratio**	**Multiplicative**	**Game Theory**
Sexton_Classic	0			
Sexton_Ratio	0	0		
Multiplicative	1	1	0	
Game Theory	1	1	1	0

Table 5.7: Wilcoxon rank test results for cross-efficiency method comparison 2012.

In 2017 the hypothesis that the scores are from the same distribution are rejected for all pairs, but as evident from table 5.7, this assumption cannot be rejected when comparing both Sexton methods in 2012. Because of this, the methods are analyzed more deeply.

5.2.3 Ranking similarity

Next to the different scores, the most critical question is whether the methods result in comparable rankings. After all, a manager is less interested whether his cross-efficiency score is 0.6

or 0.7, but where his facility benchmarks and who is best-in-class.

To give an impression of the different rankings, table 5.8 displays a ranking of the Top10[2] DMUs by cross-efficiency score for the four methods in 2017. DMUs in bold signify that they are found in the Top10 by all four methods.

Cross-efficiency Ranking 17	**Sexton_Classic**	**Sexton_Ratio**	**Multiplicative**	**Game Theory**
Top 1	**DMU098**	**DMU098**	**DMU067**	DMU049
Top 2	**DMU050**	**DMU050**	**DMU050**	**DMU098**
Top 3	**DMU028**	**DMU028**	**DMU028**	**DMU050**
Top 4	DMU049	DMU049	**DMU027**	**DMU028**
Top 5	**DMU104**	**DMU104**	**DMU098**	**DMU104**
Top 6	**DMU066**	**DMU066**	**DMU066**	**DMU066**
Top 7	**DMU027**	**DMU027**	**DMU104**	**DMU027**
Top 8	**DMU067**	**DMU067**	DMU006	**DMU067**
Top 9	DMU041	DMU041	DMU115	DMU107
Top 10	DMU107	DMU040	DMU040	DMU041

Table 5.8: Top10 DMUs, by cross-efficiency score and method 2017.

The full rankings for 2017 and 2012 can be found in appendix J. It is noteworthy that already the Top3 are not identical between the methods and 7 DMUs in 2017 and only 5 DMUs in 2012 were ranked in the Top10 by all methods. Still, the two Sexton methods have very similar rankings, with DMUs' ranks mostly differing by usually 1-2 places the most. The evolutionary game theory approach, ranks the Top10 similar to the Sexton approaches (9 out of 10 DMUs constituting the Top10 are the same), whereas the multiplicative approach's ranking differs considerably already in the first few DMUs.

Kendall's τ is calculated for all method pairs[3]. The 2017 correlations are found in table 5.9, the 2012 correlations in appendix J.

Kendall's τ 2017	**Sexton_Classic**	**Sexton_Ratio**	**Multiplicative**	**Game Theory**
Sexton_Classic	1***			
Sexton_Ratio	0,96***	1***		
Multiplicative	0,67***	0,68***	1***	
Game Theory	0,82***	0,82***	0,63***	1***

Table 5.9: Kendall's τ correlations for cross-efficiency rankings 2017.

The Sexton approaches have a correlation score of over 0.94 in both years, indicating almost identical rankings. Similarly, Sexton's methods and the game theory approach correlate with more than 0.82 in both years. The correlations between the multiplicative approach and the other three approaches are considerably lower (0.63-0.68).

5.2.4 Correlation of cross-efficiency methods with DEA and super-efficiency

As a last step, before the cross-efficiency scores are used to study warehouse performance, the cross-efficiency score distributions are compared to the DEA and super-efficiency scores. This

[2]Top10 hereby refers to the 10 DMUs that received the highest cross-efficiency score per method
[3]All p values in this thesis are indicated as follows: *p < 0.10; **p < 0.05; ***p < 0.01.

is done, to develop a feeling for how well the cross-efficiency rankings resemble the DEA scores (or super-efficiency scores, where for each DEA-efficient DMU it is calculated, how far from the PPS this DMU would be, if one excluded it from the sample).

The four CRS cross-efficiency methods that were introduced in the literature review are included in the comparison. Additionally, the Sexton ratio approach was also implemented, only using the CRS DEA efficient DMUs for the cross-efficiency computation. This results in only the DEA-efficient DMUs peer-appraising all other DMUs and therefore avoids that DEA-inefficient DMUs impacting the cross-efficiency scores.

The results for Kendall's τ correlations in 2017 are in table 5.10, the 2012 values are almost identical and can be found in the appendix J.

Kendall's τ 2017	**DEA**	**Super-efficiency**
Sexton_Classic	0,641***	0,636***
Sexton_Ratio	0,643***	0,639***
Multiplicative	0,533***	0,527***
Game Theory	0,8***	0,795***
Sexton_EfficientOnly	0,615***	0,616***
DEA	1***	0,954***
Superefficiency	0,954***	1***

Table 5.10: Kendall's τ correlation between cross-efficiency, DEA and super-efficiency scores 2017.

These correlations allow three findings: First, out of all cross-efficiency methods, the game theory method correlates the strongest with DEA and super-efficiency scores. Second, super-efficiency resembles the DEA score distribution more closely than cross-efficiency, which can be explained by super-efficiency only changing the DEA efficient DMUs' scores. Third, the correlations of the Sexton method only using efficient DMUs are similar to the Sexton method employing all DMUs as peer-appraisers. This indicates that the choice for only DEA-efficient DMUs or the entire sample to calculate cross-efficiency weights has little impact on the overall scores. Hence, for this sample the frequency and weight of the DMUs at the frontier seem to be relatively balanced. Although it is shown in table 5.1 that certain DMUs act more often as peers, their relative weight (λ) seems to be lower. Hence, the combination of (frequency as peer) * (weight as peer) is more equally distributed.

Usage of methods for the hypothesis testing As it was established above that the Sexton methods and the game theory approach lead to similar rankings and the Sexton approaches are computationally less demanding, this thesis will perform the hypothesis testing with the Sexton ratio approach. It resembles the original DEA idea best and does not use a surrogate constraint. Next to that, as shown in figure 5.2, it observes a less uniform score distribution than the game theory model, which allows for better discrimination among the most efficient DMUs. Also, the game theory approach with its limitation, on certain data sets, to result in two DMUs with unity cross-efficiency scores seems less suitable for industry comparison with cross-efficiency scoring. Especially considering applications in cluster analysis on multiple data subsets, mutually efficient DMUs become increasingly likely, given that efficient DMUs employing mutually exclusive inputs and outputs become more prevalent in smaller samples.

The multiplicative approach is dismissed for this analysis, as the warehouse benchmark results are most relevant to practitioners and the multiplicative approach's results of scoring most DMUs below 0.05 and efficient ones at 0.8 and higher renders a meaningful comparison very challenging. The biggest problem would be, to infer changes in cross-efficiency scores based on changes in inputs, if the relative distances between cross-efficiency scores are so different in magnitude. A re-basing of solutions would be possible, but required assumptions on how to best scale scores, which contradicts the cross-efficiency purpose of benchmarking without external inference.

5.3 Cross-efficiency comparison of warehouse subsets

5.3.1 Cross-efficiency comparison of different industries

The following section analyzes the relationship between warehouse-characteristics, other than inputs and outputs, and their correlation with cross-efficiency scores.

Initially, it is tested, whether there are differences between the three different clusters (Construction + Engineering, Consumer Goods, Food/Groceries). To achieve this, a Kruskal-Wallis *(KW)* analysis is performed to see whether the warehouse cross-efficiency ranks of the clusters' warehouses are different. This test was chosen, as it does not require the assumption of normally distributed residuals, works for more than two groups of data of unequal size and is non-parametric.

Rank 1 is assigned to the warehouse with the highest cross-efficiency score, whereas Rank 102 is assigned to the warehouse with the lowest cross-efficiency score. The two overlapping DMUs (DMUs that selected product categories of two clusters), were randomly assigned to one cluster. The test rejects the hypothesis that the three clusters' ranks originate from the same distribution with a p value of 0.04 in 2017 and 0.03 in 2012. The detailed test results can be found in appendix K.

To understand the magnitude of differences among product categories more closely, Wilcoxon signed rank tests were performed for each cluster against the remainder of the panelists, to see where and how the ranks differ. Based on this analysis, it becomes evident that solely the construction+engineering group is statistically different (less efficient) from the rest of the sample. In both years, it observes lower average rankings than both other sectors and the overall sample at (0.02 p-value in 2017, 0.01 in 2012). Although the other two sectors observe slightly higher average rankings than the overall sample, those differences are not significant. The detailed metrics of the clusters' ranks and the test's p values for 2017 are in table 5.11, the table for 2012 in appendix K.

Cluster Ranks 2017	**Overall Sample**	**Construction+ Engineering**	**Consumer Goods**	**Food/ Groceries**
Observations	102	20	19	19
Minimum Rank	1	6	2	4
Maximum Rank	102	96	98	101
Average Rank	51,50	65,00	47,74	45,00
Std. Dev. Rank	29,44	28,19	26,44	26,36
WSR vs. Remainder p-value		0,02	0,54	0,29

Table 5.11: Cluster metrics and Wilcoxon rank test results vs. remaining panelists 2017.

It can by hypothesized that the worse than average cross-efficiency in the construction and engineering sector is partially attributable to the high prevalence of spare-part warehouses in this cluster that in their nature store stock for long times and have lower output levels compared to their size. 5 out of the 20 construction and engineering warehouses are focused on spare-part operations, while neither consumer goods nor food/groceries have spare-part warehouses in their cluster.

5.3.2 Cross-efficiency and value chain position

To analyze whether the value chain position has an impact on efficiency, a Kruskal-Wallis test was performed. It is found that in 2017 there is no statistical difference between production, wholesale and retail warehouses. For 2012, the higher than average score is significant at the 5.8% threshold, yet given the low sample size of only 14 retail warehouses this finding should be verified with a larger data set.

2017 Value Chain's Cross-Efficiency	**Overall Sample**	**Production**	**Wholesale**	**Retail**
Observations	102	27	61	14
Minimum Score	0,09	0,09	0,12	0,18
Maximum Score	0,89	0,89	0,87	0,82
Average Score	0,36	0,39	0,34	0,38
Std. Dev. Score	0,18	0,22	0,16	0,17
KW p-value			0,6704	

Table 5.12: Kruskal-Wallis test for different value chain positions' effect on cross-efficiency scores 2017.

The 2012 value chain position Kruskal-Wallis test result table can be found in appendix K.

5.3.3 Cross-efficiency and ownership type

Similarly to above, in this section, a Kruskal-Wallis test was performed to analyze the impact of ownership type on cross-efficiency scores. It is found that neither for 2017 nor 2012 there is a statistical difference between the self-owned, dedicated contracted and multiple contracted facilities. In both years though, the in-house warehouses observed the highest efficiencies, while 3PL-operated warehouses were less efficient. At the same time, the scores of both 3PL groups observed lower standard deviations. Both findings can be explained, as the less efficient DMUs may be those with a lower degree of operational specialization, as specifically tailored warehouses are not warranted by most logistic contracts with little continuity:

2017 Ownership's Cross-Efficiency	**Overall Sample**	**In- House**	**3PL-Dedicated**	**3PL-Multiple**
Observations	102	58	10	34
Minimum Score	0,09	0,09	0,21	0,12
Maximum Score	0,89	0,89	0,55	0,58
Average Score	0,36	0,39	0,33	0,33
Std. Dev. Score	0,18	0,22	0,10	0,11
KW p-value			0,8949	

Table 5.13: Kruskal-Wallis test for different ownership types' effect on cross-efficiency scores 2017.

The 2012 ownership type Kruskal-Wallis test result table can be found in appendix K. When running the same test, but combining dedicated 3PL warehouses and those with multiple clients into one, the score differences remain insignificant in both years (Kruskal-Wallis p-value of 0.64 in 2017).

5.3.4 Control variable selection

For the subsequent correlation analyses between the non-nominal variables (inputs, outputs,cold-storage%) and cross-efficiency, Multiple Linear Regressions *(MLR)* will be run, weighing the errors with ordinary least squares. Hoff (2007) find that despite the censored nature of the observed cross-efficiency score in the interval of]0;1], OLS predicts ranking for DEA slightly better than three commonly used alternative models (Tobit regression, Papke-Woolridge approach, unit-inflated beta model) for regressing factors on cross-efficiency scores.

A crucial part of deciding on the correct regression model is to include the appropriate control variables. To achieve this, in a first step one regression is run, only using the potential control variables to decide which are significant. In this first model, value chain position, ownership type, industry type and cold-storage percentage were used as explanatory variables for the dependent variable cross-efficiency score. All variables were standardized and the model run three times, for 2012, for 2017 and for the relative change of scores and variables between 2017 and 2012. The model results are displayed in table 5.14.

MLR Coefficients Control Variables	**2012**	**2017**	**Change**
Intercept	0,27	0,05	0,29
Production Warehouse	-0,8*	0	0
Wholesale Warehouse	-0,68*	-0,07	-0,22
Retail Warehouse	0	0,37	-0,61
Self-Owned	0,42	0,3	0,12
Third-Party - Dedicated	0	0	0,57
Third-Party - Multiple	0,19	-0,08	0
Cold-Storage	0,24**	0,26**	0,05
Automotive	0,04	-0,3	-0,48
Chemicals/Oil/Gas	0,26	0,3	0,09
Construction	-0,34	-0,16	0,42
Consumer goods	0,45	0,2	-0,28
Electronics	0,22	0,29	0,22
Engineering	-0,15	-0,38	0,05
Groceries/Food	0,16	0,05	-0,02
Household Appliances	-0,57	-0,96*	-0,18
Logistics	-0,26	-0,44	-0,33
Military/Defense	0	0	0
Pharmacy	0,46	0,28	-0,02
Retail (non-food)	-0,17	-0,61	-0,55
Textiles	0,51	0,15	-0,28
Adjusted R^2	*0,02*	*0,04*	*-0,01*
RMSE	*0,99*	*0,98*	*1*
F-value	*1,14*	*1,24*	*0,95*

Table 5.14: Multiple linear regression results, performed on only control variables.

From these results, one can exclude all variables, except for cold-storage percentage, as control variables, given that no other variable was significant in 2012 and 2017. Also, the model fit as measured by R^2 (0,02 and 0,04) and F-value (1,14 and 1,24) is low. The lack of explanatory power can especially be seen from the non-zero intercepts, despite a standardized dependent variable.

In a next step, the four inputs and four outputs as explanatory variables and cold-storage percentage as control variable are used, to gauge the improvement in model fit. The dependent variable is cross-efficiency and all variables are standardized for comparability of results:

MLR Coefficients	**2012**	**2017**	**Change**
Intercept	0	0	0
Cold-Storage	0,09	0,22***	-0,05
FTEs	-0,49***	-0,48***	-0,2**
Floor Space	-0,32***	-0,2**	0,05
SKUs	-0,34***	-0,26***	-0,54***
Automation	-0,44***	-0,55***	-0,68***
Order Lines	0,7***	0,66***	0,59***
Special Processes	0,24***	0,2***	0,28***
Error Free %	0,06	0,06	0,18**
Order Flexibility	0,23***	0,1	0,24***
Adjusted R^2	*0,55*	*0,53*	*0,52*
RMSE	*0,59*	*0,68*	*0,69*
F-value	*14,40*	*13,70*	*13,10*

Table 5.15: Multiple linear regression results, performed with inputs and outputs, controlling for cold-storage percentage.

The explanatory fit of the models including inputs and outputs (adjusted R^2 of >0.5 and double digit F-values) has increased by a large degree through including the inputs and outputs. Also, each of the inputs and outputs is a significant regression coefficient in at least one of the three analyzed runs (2012, 2017 and relative change of variables and cross-efficiency scores). Due to the aforementioned standardization of variables, the intercept is zero and henceforth will not be displayed in model outputs.

For an empirical data-set, the goodness of fit is satisfactory and the results show a significant relationship between inputs, outputs and cross-efficiency scores. Hence, this model will be used for subsequent analyses.

5.3.5 Partial correlations and multiple linear regression

The MLR model introduced above allows an understanding of the change in cross-efficiency score, based on a change in one of the variables. As the values are all standardized, a regression coefficient of -0.55 (automation and cross-efficiency in 2017) means that a change of the automation score by one standard deviation, reduces the cross-efficiency score by 0.55 standard deviations.

An alternative way of assessing the relation between explanatory and dependent variables is partial correlation. For partial correlation, the correlation between one explanatory- and one dependent variable is calculated, after having controlled for other model variables. If one uses Pearson partial correlation, it finds the same p-values as MLR, but returns classic Pearson correlation coefficients between -1 and 1 (Schroeder and Sjoquist 2011). However, as the dependent and control variables in this case are not continuous, a partial rank correlation is given preference over a partial Pearson correlation.

The Spearman ρ partial rank correlation results between cross efficiency and each input, output as well as cold-storage are found in table 5.16.

Spearman ρ Partial Correlations	**2012**	**2017**	**Change**
Cold-Storage	-0,08	-0,04	0,01
FTEs	-0,2**	-0,37***	-0,23**
Floor Space	-0,46***	-0,43***	-0,14
SKUs	-0,64***	-0,78***	-0,3***
Automation	-0,61***	-0,66***	-0,74***
Order Lines	0,52***	0,58***	0,28***
Special Processes	0,42***	0,56***	0,39***
Error Free %	0,32***	0,23**	0,18*
Order Flexibility	0,26**	0,25**	0,32***

Table 5.16: Spearman's ρ partial rank correlation between cross efficiency and each individual model variable, when controlling for all other inputs, outputs and cold-storage.

For this thesis' hypotheses testing, these correlations are the crucial parameters. However, the detailed, subsequent analyses will be focused on the MLR results from table 5.15. Because of the large number of variables in the model and the fact that all significant factors exhibit the same signs for partial rank correlation as well as multiple linear regression, the MLR coefficients are given preference, as they indicate the magnitude of impact of explanatory variable changes. Also, the model fit of the MLR models can be more easily assessed and interpreted than the combination of multiple partial regression results. Hence, in the following, the reported coefficients will focus on the regression results and additionally state the ranking correlations where deemed informative.

5.3.6 Cross-efficiency comparison of cold-storage facilities

Before investigating the inputs' relationship with cross-efficiency, the control variable cold-storage is analyzed. Initially, it is tested whether warehouses with any degree of cold-storage operations rank differently under cross-efficiency than DMUs with no cold-storage.

Using the Wilcoxon signed rank test on the 41 cold-storage DMUs and the 61 non-cold-storage DMUs, the average rank for cold-storage facilities is 5 ranks better in 2017 than without cold-storage. Despite the similar standard deviations in both groups, the WSR test attests no statistically significant difference. In 2012 the rankings were virtually identical, confirmed by a p-value of 0.89:

Cold-Storage	**No Cold-Storage 2017**	**Cold-Storage 2017**	**No Cold-Storage 2012**	**Cold-Storage 2012**
Observations	61	41	61	41
Minimum Rank	2	1	3	1
Maximum Rank	99	102	99	102
Average Rank	53,59	48,39	51,61	51,34
Std. Dev. Rank	28,71	30,24	29,83	28,86
WSR vs. Remainder p-value	0,29		0,89	

Table 5.17: Wilcoxon test comparison between cold-storage and non-cold-storage DMU ranks.

However, this analysis does not incorporate the percentage of cold-storage yet. In a next step, the relation between cross-efficiency score and cold-storage percentage (measured as percentage of available floor space in m^2) is looked at.

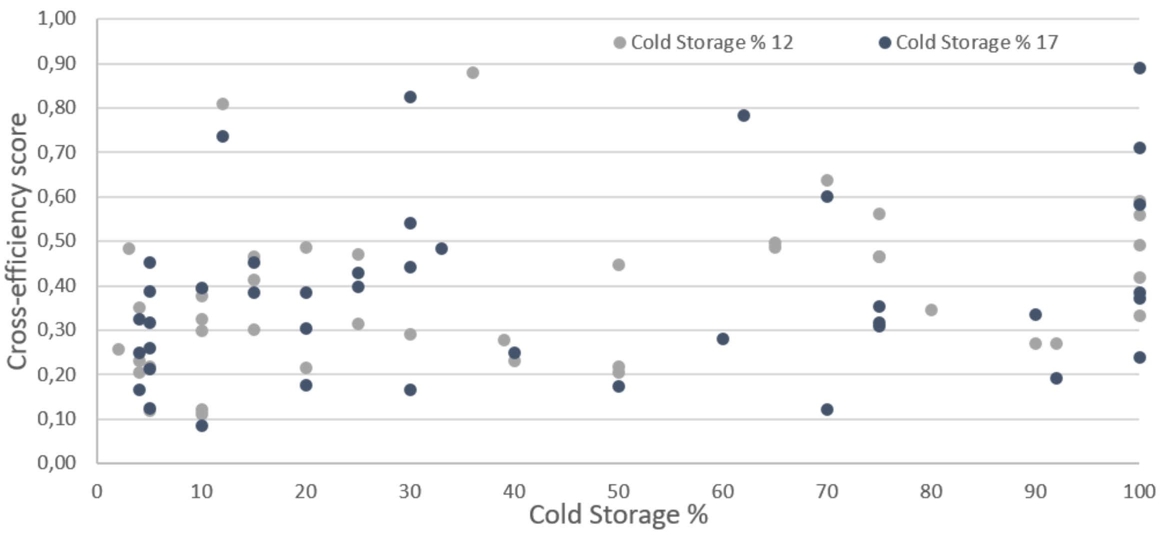


Figure 5.3: Plot of cold-storage percentage and cross-efficiency score.

When looking at figure 5.3, it is difficult to establish a clear idea of the relation between the two parameters. But running a MLR and controlling for the DEA input and output variables finds a positive regression coefficient for cold-storage percentage and cross-efficiency. Whether one looks at all DMUs or only at those with a cold-storage percentage >0%, the regression coefficient is positive in both years, but only significant in 2017:

	All DMUs		**Cold-Store only**	
Cold-Storage Regression Coefficient	**2012**	**2017**	**2012**	**2017**
Cold-Storage Coefficient	0,09	0,2229***	0,1209	0,1978*
Adjusted R^2	*0,55*	*0,53*	*0,73*	*0,74*
RMSE	*0,59*	*0,68*	*0,52*	*0,51*
F-value	*14,40*	*13,70*	*10,20*	*10,70*

Table 5.18: Regression coefficient between cold-storage percentage and cross-efficiency score, after controlling for inputs and outputs.

Although initially counter-intuitive that labor-intensive cold-storage drives cross-efficiency, one can conjecture that establishing cold-storage capabilities requires standardized processes with high adherence to not endanger the products. This operational accuracy might then translate into overall efficiency.

5.4 Effect of automation on cross-efficiency

After the previous sections explored the variances among the several warehouse characteristics in relation to cross-efficiency, this section is devoted to analyzing the impact of individual input factors on cross-efficiency.

The standardized regression coefficients for 2017 are summarized in table 5.19 and discussed in detail in the next subsections, the 2012 results and coefficients for change of inputs and change of cross-efficiency can be found in appendix L.

MLR Coefficients 2017	**All DMUs**	**Construction+ Engineering**	**Consumer Goods**	**Food/ Groceries**
Cold-Storage	0,22***	-0,1	0,06	0,29**
FTEs	-0,48***	-2,69**	-0,72	-0,15
Floor Space	-0,2**	-0,36	0,07	-0,29
SKUs	-0,26***	-0,35*	-0,26	-0,18
Automation	-0,55***	-0,12	-0,89**	-0,66***
Order Lines	0,66***	2,87**	1,24*	0,69***
Special Processes	0,2***	0,83***	0,29	0,16
Error Free %	0,06	0,12	0,41*	0,04
Order Flexibility	0,1	-0,17	0,2	0,26**
Adjusted R^2	*0,53*	*0,49*	*0,52*	*0,89*
RMSE	*0,68*	*0,71*	*0,69*	*0,33*
F-value	*13,70*	*3,04*	*3,19*	*17,90*

Table 5.19: Multiple linear regression coefficients of inputs and outputs on cross-efficiency scores 2017.

For the next analyses, it be noted that the adjusted R^2 and F-values for the regression on DMU clusters, in part, exceed the levels observed for the regressions on entire data-set. This better fit may be explained by the more homogeneous nature of facilities within the same industry.

5.4.1 Automation's effect in 2012 and 2017

Automation negatively regresses on cross-efficiency with -0.55 in 2017 and -0.44 in 2012, both at a 1% significance level. Likewise, the rank correlation between the two variables is -0.61 in 2012 and -0.66 in 2017. This contradicts the assumption made at the beginning of the thesis that more automation leads to higher efficiency, but could be caused by the fact that larger warehouses tend to have a higher level of automation but in previous studies were found to be less efficient than smaller warehouses.

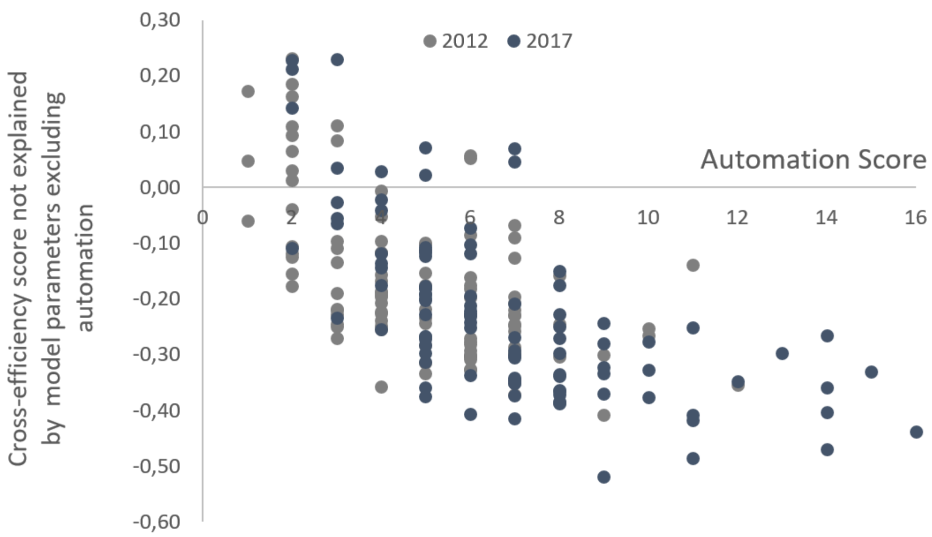


Figure 5.4: Regression of automation on cross-efficiency score, controlling for the other inputs, outputs and cold-storage.

5.4.2 Automation change's effect on cross-efficiency scores

The same negative relation holds true when comparing the change of automation

$$\left(\frac{Automation \quad Score \quad 2017 - Automation \quad Score \quad 2012}{Automation \quad Score \quad 2012}\right)$$

against the change in cross-efficiency score. The regression coefficient of Δ automation on Δ cross-efficiency score between 2017 and 2012 is -0.68.

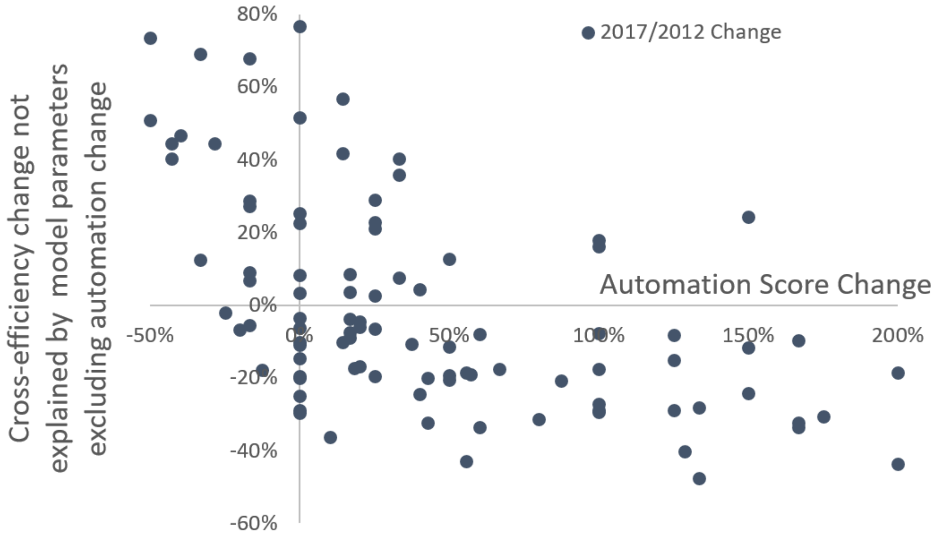


Figure 5.5: Regression of change in automation score on change in cross-efficiency score.

Hence, automation and cross-efficiency negatively correlate in both years and in the time-comparison between 2017 and 2012. This finding supports the hypothesis that large warehouses use more automation, because they could not stay competitive otherwise. Simultaneously, they are less technically efficient, as they are more likely subject to decreasing returns to scale. This interpretation would explain why automation is evidently increasingly used, although it does not seem to positively correlate with cross-efficiency. It is more efficient than the non-automated set-up at a larger scale, but has a minimum scale threshold before which it is being employed. Further research is necessary to investigate this conjecture.

5.4.3 Automation cluster analysis

Despite the overall negative correlation between automation and cross-efficiency, there are certain differences among the different clusters, most notably the groceries group. While the consumer goods' average automation score is slightly below the average score (6.58 vs 6.83), grocery warehouses have a mean automation score of 8.21. Consumer goods had the highest negative regression coefficient of -0.89 at an alpha of 0.01 and also the highest automation change, as well as the highest number of DMUs that changed their automation level. One could hypothesize that automation is rather a symptom than the cause of inefficiency.

The author visited the largest food warehouse in the sample and discussed reasons for automation investment decisions with the warehouse management team. According to them, the high level of automation installed in their facility is necessary, because the tens of thousands perishable units they receive and ship every day could not be handled by purely manual labor at competitive costs. But even more important, conveyor belts, automated stackers, labeling machines etc. ensure better quality and higher consistency. In such a sensitive environment, where slight mistakes render the product obsolete and hygiene regulations dictate process design, automation technology is the best option. In this particular case, the high quality focus and exotic assortment leads to a low cross-efficiency ranking, yet all-manual operations would likely rank worse.

Because of this, one can conjecture that the reduced cross-efficiency caused by automation is not a causal relationship, but hints to operational necessities of specific industries and warehouse sizes.

Automation Statistics	**Overall Sample**	**Construction+ Engineering**	**Consumer Goods**	**Food/ Groceries**
Average Score 2017	6,83	6,95	6,58	8,21
Std. Dev. Score 2017	2,96	2,78	3,03	3,64
Regression Automation & Cross-efficiency 2017	-0,55***	-0,12	-0,89**	-0,66***
Average Automation Score Change 2012-2017	65%	49%	106%	65%
% of DMUs with Change	84%	75%	95%	89%

Table 5.20: Cluster analysis metrics of automation's effect on cross-efficiency ranking.

Overall, all sectors increased their average automation scores and almost 85% of facilities across the entire sample have changed soft- or hardware technology during those 5 years.

5.5 Effect of floor space on cross-efficiency

5.5.1 Floor space's effect in 2012 and 2017

The negative coefficients that were observed between cross-efficiency scores and automation are also prevalent for floor space at -0.32 in 2012 at 0.01 significance. The 2017 regression coefficient is -0.20 (0.05 significance). The rank correlation between the two variables is -0.46 in 2012 and -0.43 in 2017. This also rejects the hypothesis of larger warehouses being more efficient.

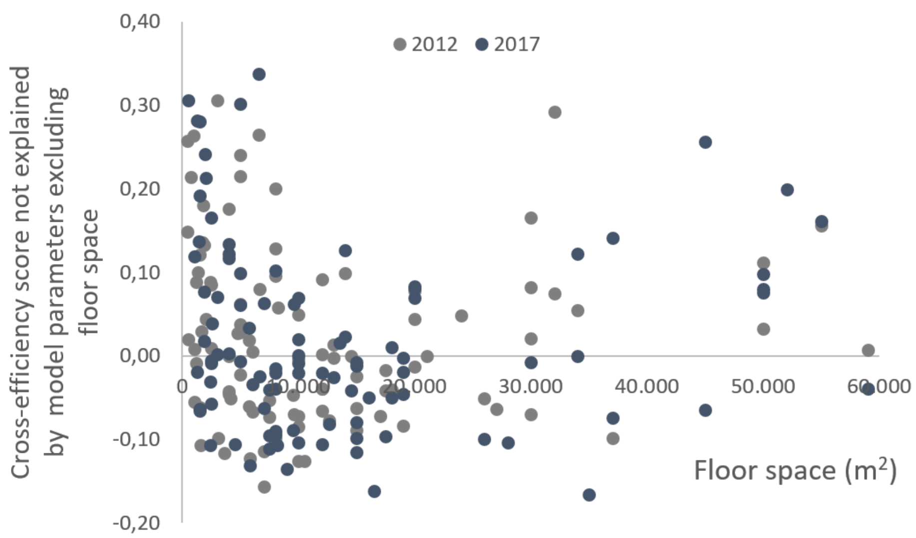


Figure 5.6: Regression of floor space on cross-efficiency score, controlling for the other inputs, outputs and cold-storage.

Seeing that increasingly large warehouses are built especially for e-commerce applications, this is not intuitive, but reasons for larger-scale warehouses may lie beyond the scope of operational warehouse efficiency, for example in assortment availability. The operational inefficiencies could also be compensated for, by having financial benefits through centralization and economies of scale in in- and outbound shipment and consolidation of products. A detailed investigation of this matter however, is beyond the scope of this thesis.

5.5.2 Floor space change's effect on cross-efficiency scores

Contrary to the automation change, it is observed that floor space change is not significantly correlated with cross-efficiency. Floor space was the only input or output, whose change is not correlated with a change in cross-efficiency:

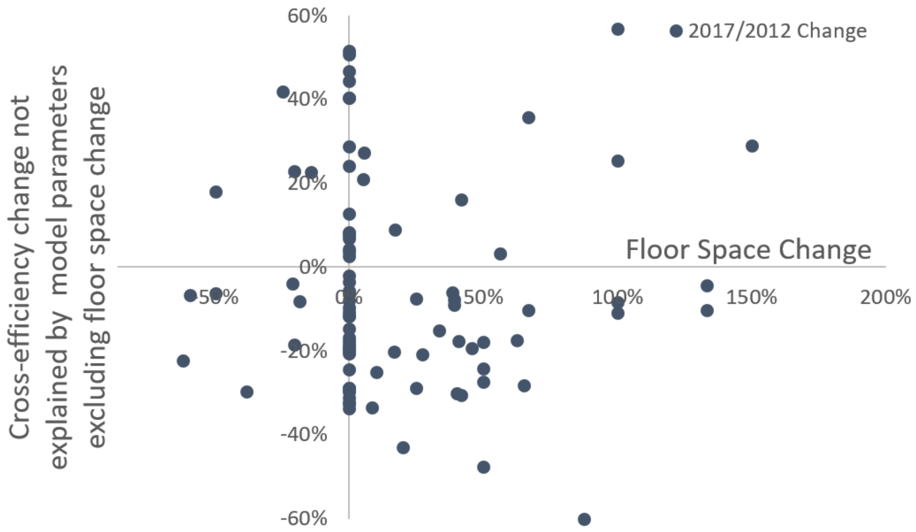


Figure 5.7: Regression of change in floor space on change in cross-efficiency score.

5.5.3 Floor space cluster analysis

In the cluster analysis, it becomes evident, that consumer goods warehouses use marginally more floor space than the overall sample (21,000m^2 against 18,000m^2), while construction+engineering warehouses on average are only 16,000m^2 in size. Although the different industry clusters have diverging floor space correlations, they are not significant. Facilities in the food/groceries sector have shown the largest average increase in floor space over the past years of +79%, although only 47% of warehouses changed their size, indicating a very high growth for those warehouses who did expand. The overall increase in space was 36% over 5 years, which fits the general observation of warehouses growing in size. Yet, the standard deviation of floor space for the entire set as well as for the construction+engineering and consumer goods clusters is higher than average - representing a high dispersion of facility sizes.

Floor Space Statistics	**Overall Sample**	**Construction+ Engineering**	**Consumer Goods**	**Food/ Groceries**
Average Floor Space 2017	18.244	16.136	20.721	19.533
Std. Dev. Floor Space 2017	32.414	20.823	31.770	16.267
Regression Floor Space & Cross-efficiency 2017	-0,20*	-0,36	0,07	-0,29
Average Change Floor Space 12-17	36%	12%	32%	79%
% of DMUs with Change	55%	70%	58%	47%

Table 5.21: Cluster analysis metrics of floor space's effect on cross-efficiency ranking.

5.6 Effect of SKUs and FTEs on cross-efficiency

5.6.1 SKUs and FTEs's effect in 2012 and 2017

The regression coefficients for assortment size and workforce are similar in sign and magnitude to automation:

MLR Coefficients	**2012**	**2017**
SKUs	-0,34***	-0,26***
FTEs	-0,49***	-0,48***

Table 5.22: Correlation coefficients of SKUs and FTEs with cross-efficiency scores.

5.6.2 SKUs and FTEs changes' effect on cross-efficiency scores

Analogous to the negative correlations between cross efficiency and SKUs as well as FTEs in the specific years, the change between 2012 and 2017 also negatively correlates with cross-efficiency change. In total, 88 DMUs changed assortment and 92 their employee count during the period:

MLR Coefficients	**Change**
Change in SKUs	-0,54***
Change in FTEs	-0,20**

Table 5.23: Regression coefficients of change in FTEs as well as SKUs and cross-efficiency score change.

The rank correlation results for 2012, 2017 and the change between 2017 and 2012 show identical signs and are all significant as well.

5.6.3 SKUs and FTEs cluster analysis

For the SKU analysis it is noteworthy that the standard deviation is larger than the average number of SKUs for the entire sample as well as all clusters, except consumer goods - indicating a large variability in assortment size. The construction+engineering cluster hereby grew slowest, with an average increase in SKUs of 16%, while the overall sample's assortment grew on average by 58% and the food assortments even by 143%. Regardless of growth, the groceries cluster has by far the lowest number of SKUs (4k on average, compared to 21k in the overall sample), most likely due to the perishable nature of the products eliminating long-stored stock. The overall sample observes a regression coefficient of -0.26 of SKU count on cross-efficiency score. This magnitude is also found for the individual clusters, at varying levels of statistical significance:

SKU Statistics	**Overall Sample**	**Construction+ Engineering**	**Consumer Goods**	**Food/ Groceries**
Average Number of SKUs 2017	21.088	53.361	10.336	3.677
Std. Dev. SKUs 2017	57.393	85.000	8.750	3.816
Correlation SKUs & Cross-efficiency 2017	-0,26***	-0,35*	-0,26	-0,18
Average Change SKUs 12-17	58%	16%	37%	143%
% of DMUs with Change	86%	90%	74%	89%

Table 5.24: Cluster analysis metrics of SKUs' effect on cross-efficiency ranking.

Similar to the standard deviation/average ratio of SKUs, all groups also observe standard deviations equal or greater than the average of their FTE headcount. On average, 59 employees work per facility, with construction and engineering facilities employing most people on average (78). The regression coefficient between cross-efficiency score and workforce is -0.48 for the overall sample, and even -2.69 for the construction and engineering sector. 90% of DMUs in the entire sample observed changes in their FTE-count, with an average increase in FTEs of 32% over the past 5 years. Food and grocery warehouses outgrew the average workforce with an average increase of 43%.

FTE Statistics	**Overall Sample**	**Construction+ Engineering**	**Consumer Goods**	**Food/ Groceries**
Average Number of FTEs 2017	59	78	61	55
Std. Dev. FTEs 2017	74	103	89	54
Correlation FTEs & cross-efficiency 2017	-0,48***	-2,69**	-0,72	-0,15
Average Change FTEs 2012-2017	32%	7%	25%	43%
% of DMUs with Change	90%	95%	89%	84%

Table 5.25: Cluster analysis metrics of FTE's effect on cross-efficiency ranking.

The negative correlation of SKUs and FTEs with cross-efficiency is in line with the hypotheses of this thesis, prior research findings and the decreasing returns to scale results.

5.7 Running DEA and cross-efficiency for sectors individually

As a last test, the entire DEA and Sexton ratio cross-efficiency models were run with DMUs from only one cluster at a time, which changes the input and output parameter weights, as the overall PPS changes. This was done to see, whether choosing more homogeneous DMUs (from only one industry), would increase the cross-efficiency scores, assuming that competition within the warehousing sector would force similar facilities to employ similar input and output mixes to stay competitive. Hence, within one industry most warehouses would ideally be clustered around the sector's best practice input and output mix.

Table 5.26 displays descriptive statistics of the cross-efficiency scores per cluster in 2017 (2012 data in appendix M). It is notable that the average and the minimum value per cluster are higher than in the overall set of DMUs, but standard deviations per cluster are higher as well, indicating a higher variability of scores. The variability might however be attributable to the relatively smaller range of observations per cluster.

Efficiency Scores 2017	**Construction+ Engineering**	**Consumer Goods**	**Food/Groceries**
Minimum Score	0,29	0,18	0,17
Maximum Score	0,95	0,89	1,00
Average Score	0,60	0,50	0,54
Std. Dev. Score	0,18	0,21	0,21
Observations	20	19	19
Adjusted R^2	*0,59*	*0,47*	*0,92*
RMSE	*0,64*	*0,73*	*0,28*
F-value	*4,03*	*2,76*	*25,40*

Table 5.26: Descriptive statistics of the cross-efficiency score distribution for calculations with individual clusters 2017.

Additionally, the regression analysis for the individual clusters' cross-efficiency scores and the input and outputs was performed again, resulting in coefficients of the same sign and order of magnitude as for running cross-efficiency and the regression on the entire data set. The adjusted R^2 values for the 2012 and 2017 regression models reach higher values (0.47-0.92) on average for all individual industry runs. Likewise, the F-values are slightly higher for the individual runs as well. This indicates a better fit of the models, when running the DEA cluster by cluster, which cannot be explained by smaller sample sizes, as the measures were adjusted for number of observations. The detailed coefficient tables can be found in appendix M.

As will be discussed in detail in section 6.2, the Sexton ratio method retains its discriminatory power on smaller data-samples and distributes the cross-efficiency scores over a similar range as when using the entire sample. With the increased models' fit, one can argue in favor of using only a smaller subset of DMUs to rank specific industries. However, the industry-only rankings are not identical to the originally calculated ones. Hence, careful reflection on the merits and drawbacks of each approach is mandatory - before opting for either.

5.8 Managerial impact on cross-efficiency

This section deviates from the previous ones, in that it tries to explain part of the observed cross-efficiency deviations by decisions and strategies employed by the warehouse management. For this purpose, each warehouse manager was asked ten questions about internally and externally oriented management elements, such as pre-shift meetings, training plans and customer involvement programs. For each of the questions the respondents could choose the degree to which the activity was performed in the warehouse from 1 to 5:(1) Never; (2) Rarely; (3) Sometimes; (4) Frequently and (5) Very Frequently. As these questions were not previously tested, Horn's Parallel Analysis *(PA)* is performed as a first step, to see how many different concepts are measured through the questions. This method was chosen, given its low sensitivity to data distribution and its property to exhibit little under- or overestimating of the number of components (Dinno 2009).

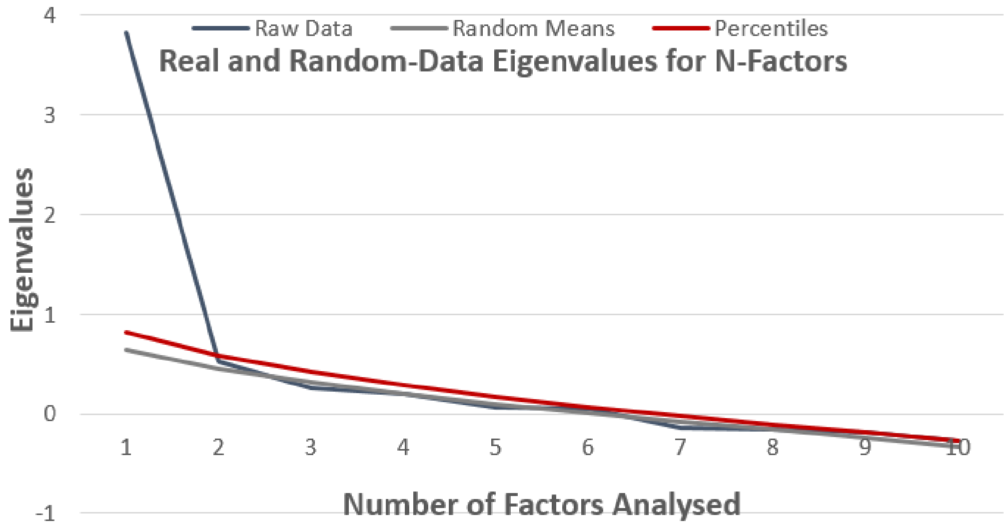


Figure 5.8: Parallel Analysis for management data 2017.

A Monte-Carlo based PA run of 5,000 iterations with an alpha threshold of 0.05 shows that only one factor reaches the required significance. As the 2012 management data PA neither found the second factor to exceed the required significance, it was decided to combine all measures to one factor. Subsequently, the factor loadings for this construct were tested for both years.

Factor Loadings	**2012**	**2017**
Pre Shift Meetings	0,42	0,60
Individual Training Plans	0,56	0,59
Incentive Programs	0,46	0,63
Strategies to Empower	0,63	0,60
Worker Rotation	0,55	0,45
Internal Information Visibility	0,60	0,53
Shop Floor Walks	0,54	0,60
5S Measures	0,69	0,83
Employee Involvement	0,57	0,68
Customer Involvement	0,40	0,59

Table 5.27: Factor loadings for management impact questions.

Given that these questions are of subjective nature, the lower factor loading in 2012 can be explained by the warehouse managers guessing the frequency of activities to a higher degree. The Cronbach alpha of the entire question set is 0.855. As no factor loaded lower than 0.3 and each of the ten question decreases Cronbach's alpha when excluded, all ten factors were used for the final analysis. The descriptive statistics of the combined management score variable can be found in table 5.28.

Management Score Statistics	**2012**	**2017**
Observations	102	102
Minimum Score	10	12
Average Score	22,5	33,3
Median Score	22,5	34,0
Maximum Score	34	46
Standard Deviation Score	5,9	7,2

Table 5.28: Descriptive statistics for the combined management score variable in 2012 and 2017.

After performing the linear regression, the management score observes a significant negative regression coefficient, yet at a low magnitude of less than -0.01 in both years. At the same time, the models' fit is minimal in both years as can be seen from the very low R^2 of 0.04 and 0.03. The model results and a scatter plot, evident of the low correlation between management score and cross-efficiency, are found in table 5.29 and figure 5.9.

MLR Coefficients Management	**2012**	**2017**
Intercept	0,5***	0,52***
Management Score	-0,0062**	-0,0048*
Adjusted R^2	*0,04*	*0,03*
RMSE	*0,18*	*0,18*
F-value	*4,27*	*3,84*

Table 5.29: Multiple linear regression coefficients of management score on cross-efficiency scores.

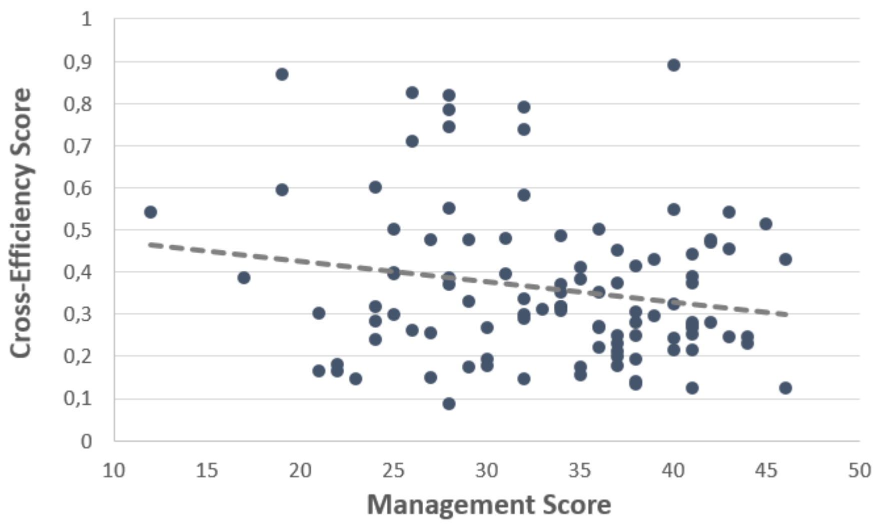


Figure 5.9: Management score and cross-efficiency score of DMUs 2017.

Although the ten questions appear to be measuring one management construct, it has no bearing on the efficiency of warehouses. More research would be necessary, to devise a measurable management construct that explains cross-efficiency to a larger extent. It is noteworthy that when performing MLR with the cross-efficiency score as dependent variable and the ten management questions as individual explanatory variables, the continuous improvement framework, 5S, observes the highest positive correlation with cross-efficiency. Likewise, the employee involvement program score shows the highest negative correlation with cross-efficiency.

To conclude the management section, a Kendall's rank correlation analysis is performed, to see whether the management score correlates with the outputs and the basic warehouse productivity measures of order lines per employee:

Kendall's τ Correlation with Management Score	**2012**	**2017**
Order Lines	0,03	0,22***
Special Processes	0,05	0,15**
Error Free %	0,24***	0,18**
Order Flexibility	0,16**	0,3***
Order Lines / FTE	-0,05	0,12*

Table 5.30: Kendall's τ rank correlation coefficients between management score, outputs and order lines per FTE.

Also for this analysis, significant but low correlations are found between the management score and the outputs, with order flexibility in 2017 showing the highest coefficient of 0.3. Based on these results, the management score cannot be closely linked to cross-efficiency or any of the outputs, but warehouse output and management score are loosely correlated.

5.9 Summary analysis

In this chapter, it was found that variable returns to scale are prevalent in the warehousing sector and an indication of optimal input size was provided. A split between operational inefficiencies and scale inefficiency of approximately 2:1 was empirically derived. Furthermore, it was established that three cross-efficiency methods' rankings correlate with >0.9, while the multiplicative approach differed from these three rankings, but also observed a very skewed score distribution, with few DMUs exhibiting high cross-efficiency scores and over two-thirds of DMUs with scores below 0.05. Hence, the analyses of this chapter were conducted using the Sexton ratio-approach. A detailed break-down of the methods' properties is performed in the next chapter.

When more closely investigating the three clusters' (construction + engineering, consumer goods, food/groceries) cross-efficiency rankings, they do differ compared to each other in 2017, but not in 2012 and those differences when comparing individual clusters against the remainder of responses are not statistically significant. Cold-storage warehouses were found to be slightly more efficient than non-cold-storage warehouses, while value chain position and ownership type did not impact the cross-efficiency scores.

Moreover, across the entire set and each cluster, it was found that all input factors automation, floor space, SKUs and FTEs negatively correlate with the cross-efficiency scores in both years. These results did not change, when running the cross-efficiency models per cluster individually. The implications for the hypotheses will be discussed in the following chapter.

6. Hypotheses and method findings

6.1 Hypotheses results

The hypotheses are evaluated based on the overall sample of 102 DMUs. As shown in table 6.1, the regression coefficients of automation on cross-efficiency scores for the entire sample set are significant at a 0.01 level for both years and the time-line comparison. The original hypothesis of a positive correlation was based on the idea that despite being an input, automation might have a much lower (standardized) negative impact than the other inputs. This could have been interpreted as automation being the "input of choice" when wanting to increase inputs and retain efficiency. The results are supported by the multiple linear regression results and the partial rank correlations alike.

The hypotheses 1a and 1b that automation and efficiency as well as the change in automation and the change in efficiency positively correlate have to be rejected. Also, the hypothesis 4a that floor space and efficiency positively correlate has to be rejected. In all three cases the observed correlation was negative. The hypotheses 2 and 3 that SKUs and FTEs negatively correlate with efficiency is supported by this sample. Hypothesis 4b that floor space change and efficiency change are positively correlated has to be rejected based on the warehouse data as there was no statistical correlation between the two (refer to table 5.16 for all correlation coefficients).

A summary of the regression coefficients[1] of the inputs, outputs and cold-storage on cross-efficiency for both time periods and the change over time is provided in table 6.1.

MLR Cross-efficiency Coefficients	**2012**	**2017**	**Change**
Cold-Storage	0,09	0,22***	-0,05
FTEs	-0,49***	-0,48***	-0,2**
Floor Space	-0,32***	-0,2**	0,05
SKUs	-0,34***	-0,26***	-0,54***
Automation	-0,44***	-0,55***	-0,68***
Order Lines	0,7***	0,66***	0,59***
Special Processes	0,24***	0,2***	0,28***
Error Free %	0,06	0,06	0,18**
Order Flexibility	0,23***	0,1	0,24***
Adjusted R^2	*0,55*	*0,53*	*0,52*
RMSE	*0,59*	*0,68*	*0,69*
F-value	*14,40*	*13,70*	*13,10*

Table 6.1: Multiple linear regression coefficients of inputs, outputs and cold-storage on cross-efficiency scores - 2012, 2017, change '12-17. The 2012 and 2017 coefficients are based on input and outputs values, the change coefficients on the relative change of values.

This answers the managerial research question from the beginning. Based on this data set of 102 warehouses and the applied Sexton ratio cross-efficiency method, all inputs negatively correlate with cross-efficiency, but floor space exhibits the lowest magnitude, followed by assortment size. Contrary to the conceptual model presented in the methodology part, automation is negatively correlated with cross-efficiency - so is the number of FTEs.

[1]All p values in this thesis are indicated as follows: *p < 0.10; **p < 0.05; ***p < 0.01.

6.2 Cross-efficiency method evaluation

All methods will be scored on all metrics with a score between 1 and 5, with 5 being the maximum score. This procedure is used, so one can quickly compare methods across single dimensions, but not intended to be used to compute an overall score per method, as different situations require different methods and put varying emphasis on particular characteristics. A recommendation of when to use which method will be provided as well.

- **Methodological proximity**

Sexton_Classic: **3** - The approach is very closely related to the regular DEA, but sums over inputs and outputs, to linearize the objective function. Yet, the cross-efficiency results of this data set are only correlated with CRS DEA by approximately 0.6.

Sexton_Ratio: **4** - The approach is fully derived from the logic of simple DEA. Yet, the cross-efficiency results of this data set are only correlated with CRS DEA by approximately 0.6.

Multiplicative: **2** - The approach is based on DEA weights, but employs exponential transformation of the objective function. The cross-efficiency results of this data set are only correlated with CRS DEA by approximately 0.5.

Game Theory: **4** - Only the starting point of DEA scores and the idea of comparing DMUs remains the same, but iterative process and pair-wise optimization are used instead of global optimization. However, the score-distribution observes the highest correlation with CRS DEA of all methods (0.8).

- **Implementational ease**

Sexton_Classic: **5** - Simple linear program to be directly implemented with Excel or mathematical programming tools - almost instant solving time due to linear properties.

Sexton_Ratio: **4** - Simple, but non-linear program that solves within ten seconds in this case of 102 DMUs with four inputs and outputs.

Multiplicative: **4** - Program with quadratic growth of constraints, rendering set-up difficult, but solves within 12 seconds despite thousands of constraints, due to linear properties.

Game Theory: **2** - Iterative nature, plus pairwise comparison requires linear solver to be called thousands of times, which took over 70 minutes for this model on a quad-core 3.8 GHz CPU.

- **Extendability**

Sexton_Classic: **4** - Given the limited community of DEA scholars in general, a wide variety of extensions to the model is readily available and tested.

Sexton_Ratio: **4** - Given the limited community of DEA scholars in general, a wide variety of extensions to the model is readily available and tested - the ratio set-up of constraints should not impair the usage of the extensions.

Multiplicative: **2** - Relatively new model (this thesis used an adapted version of the scale-invariant one published in 2014), for which possible extensions are mentioned in the original paper and build on more popular non-exponential models, but only a limited number of citations can be found in online databases.

Game Theory: **3** - Considerable number of publications applying the model (mainly in Asia) and some papers extending it. Most extensions are co-authored by one or more of the original authors, but the general game theory approach does find application among DEA pundits.

- **Discriminatory properties** As introduced in section 3.4, in the following it will be analyzed how the cross-efficiency score distribution per method changes, when performing cross-efficiency for all DMUs compared to only including the CRS efficient DMUs. In table 6.2, the range and standard deviation of cross-efficiency scores for the CRS efficient DMUs is shown, when performing the cross-efficiency calculations across all DMUs:

Simple Efficient DMUs - All DMU DEA	**Sexton_ Classic**	**Sexton_ Ratio**	**Multiplicative**	**Game Theory**
Average score range 12/17	0,64	0,65	0,92	0,53
Average std.dev. 12/17	0,17	0,17	0,20	0,14
Average kurtosis 12/17	0,20	0,05	2,28	-0,16

Table 6.2: Comparison of cross-efficiency score distributions of simple-efficient DMUs, all DMUs runs.

In comparison, in table 6.3, one can see the resulting score distributions, when each cross-efficiency model was run, only including the CRS efficient DMUs as peers:

Simple Efficient DMUs - Efficient DMU DEA	**Sexton_ Classic**	**Sexton_ Ratio**	**Multiplicative**	**Game Theory**
Average score range 12/17	0,63	0,65	0,88	0,36
Average std.dev. 12/17	0,17	0,18	0,20	0,10
Average kurtosis 12/17	-0,34	-0,43	3,01	0,14

Table 6.3: Comparison of cross-efficiency score distributions of simple-efficient DMUs, efficient DMUs only runs.

For the all-DMU inclusive setup, it has to be noted that all four methods have a large enough range to compare scores and also standard deviations that allow for sufficient distribution. But only when looking at the kurtosis, one sees why the Sexton approaches and the game theory approach are better suited than the multiplicative one - the multiplicative's high kurtosis indicates a strong concentration around the mean. Instead of spreading the simple-efficient DMUs across the cross-efficiency score range, the majority of observations is clustered into one dense area. One issue, not captured here is the game theory's propensity to rate up to two DMUs as cross-efficient, which reduces the differentiating ability of the model.

When running exclusively CRS efficient DMUs in DEA and cross-efficiency models, both benevolent Sexton approaches retain range and standard deviation, but reduce their kurtosis. The multiplicative and game theory approach have very similar range and standard deviations, but the kurtosis increases for both methods.

Application of "simple-efficient-DMUs-only" Performing cross-efficiency only with the CRS efficient DMUs, increased the spread of the cross-efficiency scores across the range of observations, while retaining a similar standard distribution for the two Sexton approaches. If one performs cross-efficiency solely for tie-breaking reasons, a Sexton model only including efficient DMUs can therefore be worth considering. It has to be noted though, that the rankings obtained by the two approaches are not identical - giving a recommendation on which ranking represents reality more accurately is beyond the scope of this thesis and subject to personal preference. However, practitioners may contemplate both approaches, before selecting one for their analysis.

Sexton_Classic: **5** - Excellent discriminatory properties and, if necessary, with a reverseable secondary objective, for increased discriminatory options.

Sexton_Ratio: **5** - Excellent discriminatory properties and, if necessary, with a reverseable secondary objective, for increased discriminatory options.

Multiplicative: **2** - Discriminatory properties, but with a very high kurtosis, rendering practical comparisons difficult.

Game Theory: **4** - Good discriminatory properties and low kurtosis (even negative in the full data set but with a chance to have one equally efficient pair of DMUs.

- **Sensitivity to scale changes** DEA as an extreme point method, generally is sensitive to change of DMUs on the efficiency frontier. A change in the scale of inputs or outputs for such an efficient unit has implications on DMUs that considered the changed DMU as an efficient peer. Additionally, other DMUs that were previously dominated by different peers can switch to using the changed DMU as peer. A non-efficient DMU's change (given it does not become efficient afterwards) has only a local effect on that particular DMU.

  Because cross-efficiency uses the DEA's efficiency scores, it therefore logically changes its scores and ranking based on the DEA outcomes. For this sensitivity analysis, the main focus is therefore to find the method whose scores change least in the given scenarios.

  For the scale sensitivity test, 10% of all inputs were randomly scaled by factor 0.1, 0.2, 0.5, 2, 5 or 10. This can be of special interest in industries with volatile inputs, or models expressing inputs in prices. For each factor 100 runs were performed per model, except for the game theory approach, where due to computational reasons, 10 runs were performed. The average, relative score deviations were then compared (numeric results in appendix N):

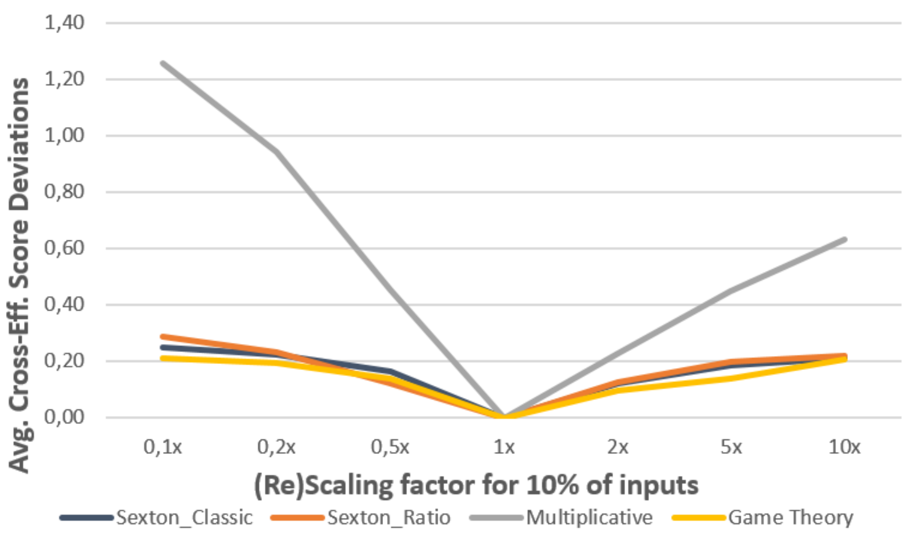


Figure 6.1: Average relative score deviations per method, based on re-scaling of 10% of inputs by a given factor.

The two Sexton approaches, and the game theory approach all show low sensitivity to scale changes. The multiplicative method's scores change three to four times as heavily as the other three methods.

Sexton_Classic: **5** Highly stable scoring across all scaling factors.

Sexton_Ratio: **5** Virtually identical sensitivity to the classic approach. Observes slightly higher deviations for the 0.1 factor, but lower changes for 0.5 factor.

Multiplicative: **2** Highest deviations of all methods as well as over 30% average changes at factor 2 input changes.

Game Theory: **5** Most stable scoring of all approaches.

- **Sensitivity to erroneous entries** For the scale sensitivity test, 5%, 10%, 25%, 50%, 75% and 100% of inputs were scaled by a random value between 75% - 125%. This is supposed to test for situations such as this thesis, when data is collected through questionnaires and subject to human error. Depending on consistency checks during the data collection processes, the required granularity and type of data requested, the received data may vary in exactness - this test sees how the methods handle such imprecisions. The average, relative score deviations were then compared (numeric results in appendix N):

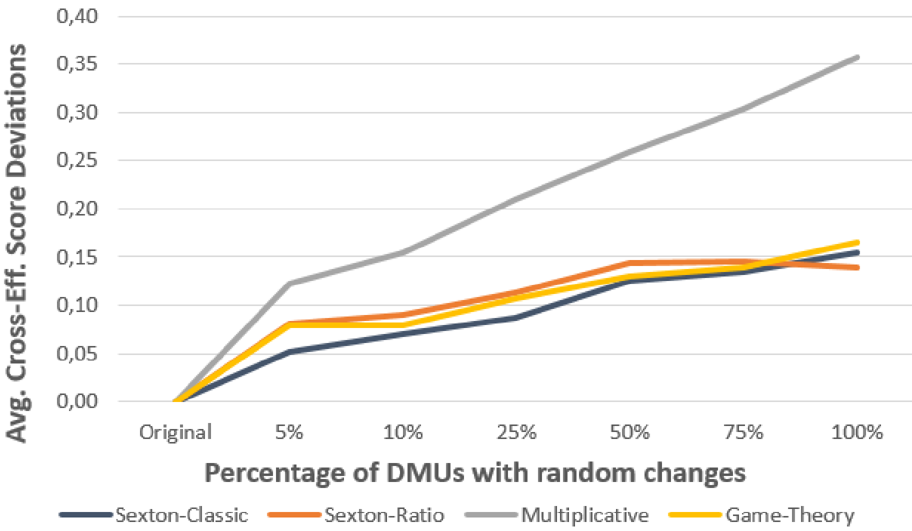


Figure 6.2: Average relative score deviations per method, based on adjusting a given percentage of inputs between 75% - 125%.

Sexton_Classic: **5** Lowest score deviations for almost the entire range of adjusted DMU percentages.

Sexton_Ratio: **5** Slightly higher score deviations than the classic approach, but almost stable between 50% and 100% of adjusted DMUs and lowest scores at the end of the range.

Multiplicative: **3** Consistently score deviations twice as high as the other approaches, yet the deviations increase linearly with the share of inputs that changes.

Game Theory: **5** Sensitivity between the two Sexton approaches with a slight increase at 100% of erroneous entries.

- **Elimination of efficient DMUs** For the efficient DMU elimination sensitivity test, 1-6 efficient DMUs were "eliminated" by setting their outputs to 0 (or close to 0 in the multiplicative case). This tests for how robust the methods are, e.g. when a DMU is reclassified (e.g. in this thesis from one cluster to another) and what impact it has on the other DMUs. Especially when observing maverick DMUs in a particular data set, assessing this sensitivity is important, if one aimed to exclude these non-representative DMUs from analysis. Because of the nature of DEA, eliminating CRS efficient DMUs has an impact on all DMUs that used the eliminated DMU as peer, hence the cross-efficiency score differences are expected to be larger than for the other tests. The average, relative score deviations were then compared (numeric results in appendix N):

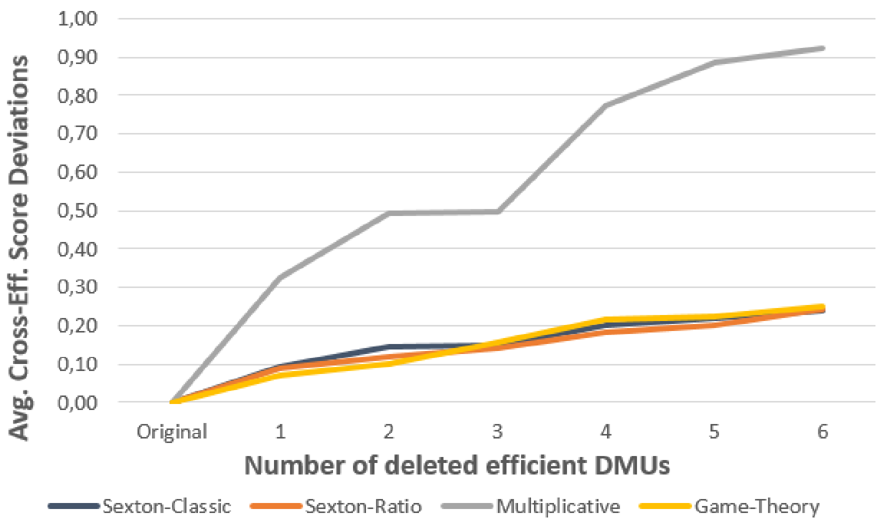


Figure 6.3: Average relative score deviations per method, based on eliminating a given number of simple efficient DMUs.

Sexton_Classic: **4** Lowest score deviations of all methods, but 10% score changes already at one eliminated DMU.

Sexton_Ratio: **4** A very similar score development to Sexton_Classic.

Multiplicative: **1** Already with one eliminated DMUs scores change by 30% and the deviations reach almost 100% for six eliminated DMUs.

Game Theory: **3** Similar magnitude of deviations than Sexton's classic approach, but a less predictable development.

Table 6.4 summarizes the above mentioned information, containing the rating for each method for each dimension:

Metric	Sexton_ Classic	Sexton_ Ratio	Multiplicative	Game Theory
Methodological Proximity to DEA	3	4	2	4
Implementational Ease	5	4	4	2
Extendability	4	4	2	3
Discriminatory Properties	5	5	2	4
Sensitivity to Changes of Scale	5	5	2	5
Sensitivity to Erroneous Data	5	5	3	5
Sensitivity to Dominant DMU Elimination	4	4	1	3

Table 6.4: Ratings of cross-efficiency methods across comparison metrics from 1(lowest) - 5 (highest).

A discussion of which method is recommendable in which situation can be found in section 7.2.

7. Conclusions and discussion

7.1 Warehouse cross-efficiency findings

Based on the 102 warehouses of this thesis' data set, it was found that all four input factors, automation (-0.66), floor space (-0.43), assortment size (-0.78) and workforce (-0.37) correlate negatively with cross-efficiency in 2017. The same orders of magnitude and signs are found for the 2012 data as well as the change in cross-efficiency scores compared to the change in inputs. These results are in line with previous research in the warehouse sector that found smaller warehouses to be more efficient, yet contradicts the main hypothesis of the thesis that automation and cross-efficiency correlate positively. One possible explanation for this is that the automation score did not include common technologies found in warehouses such as reach trucks, racking systems etc. Hence, warehouses relying heavily on classic warehouse setups, with limited or no automation technology asked for in the survey (among others AGVs, automated stackers, RFID technology or conveyor belts) can be very efficient, but receive an automation score of zero. At the same time, when looking at the automation scores at most efficient scale sizes, low-moderate automation levels are prevalent. Also, as efficiency is calculated as outputs divided by inputs, an increase in input, ceteris paribus, reduces efficiency.

As for the other input factors, it has to be noted that those are often times impacted by exogenous strategic decisions made at a corporate level. From a managerial perspective, the size recommendations found in section 5.1.3 are of relevance, as it was shown in this thesis that 1/3 of inefficiencies in warehousing stems from scale-inefficiency and 2/3 are attributable to operational inefficiency. In conjunction with the input and output mix comparison of the industries, the key learning for practitioners is to consider size as well as the choice of input and output, without solely focusing on the overall cross-efficiency scores.

Additionally, this thesis established empiric evidence that the warehouse industry exhibits decreasing returns to scale, making it more difficult for larger warehouses to reach the highest cross-efficiency ranks, independent of operational performance.

Findings for specific warehouse types include that the cold-storage percentage of floor space regresses positively on cross-efficiency in 2017, while value chain position and ownership type do not impact a warehouse's cross-efficiency significantly.

Among the three industries for which individual analyses were performed, grocery warehouses were most efficient (average rank of 45), followed by consumer goods facilities who's average ranking of 48 is above the sample's mean. The construction and engineering sector exhibits significantly lower cross-efficiency scores and consequently an average rank of 65, possibly caused by its higher share of spare-part warehouses.

When the cross-efficiency methods were run only using a sub-category's DMUs, the resulting scores were distributed across similar ranges as in the original model, despite higher homogeneity within one industry. Hence, a similar benchmark with more observations from one particular category of warehouses could prove to be insightful to understand intra-cluster differences.

7.2 Cross-efficiency method discussion

When contrasting the four cross-efficiency methods, the two Sexton approaches and the game theory method beat the multiplicative method across almost all metrics. The game theory results correlate the highest with DEA results, while the two Sexton approaches are methodologically closer to DEA, easier to implement and were 500 times quicker to solve for this data set. All three methods show low sensitivity to input scale changes, DMU elimination or erroneous data, which only holds partially true for the other two methods. Choosing between the two Sexton methods depends on the preference of the user and the context of the application. While the original method is slightly quicker solvable due to linearity, the original ratio framework suggested in this thesis approximates DEA methodologically more closely.

The multiplicative model is quickly implementable, but makes interpretation and comparison of the resulting cross-efficiency scores cumbersome, given its large dispersion of scores. Moreover, it was the least stable implementation in the sensitivity analysis.

One drawback of the game theory method is its property of scoring two DMUs to be mutually efficient (cross-efficiency scores of 1), if they are both CRS efficient and employ mutually exclusive inputs in their optimum. While not the case for the original warehouse data set, the random data transformations during the sensitivity analysis in 10% of runs, resulted in the game theory method finding two DMUs with cross-efficiency scores of 1. Depending on the required application, this may eliminate the method from being used. Moreover, its computational requirements grow faster than quadratic when doubling DMUs, as the number of DMU pairs increases quadratically, which are then run over a larger number of iterations until all scores converge, which reduces its applicability in practice.

For all cross-efficiency models, it was found that running them only using the CRS DEA-efficient DMUs has desirable distribution properties, when focusing on breaking the tie among unity-efficient entities. Especially the two Sexton model show a wide range of score distributions and lower kurtosis than when running them with the entire data-set. Depending of the goal of the cross-efficiency calculation, incorporating only the CRS DEA-efficient DMUs may therefore be appropriate.

In general, the ratio implementation of Sexton's approach is the recommendable model, as it shares all merits of Sexton's standard, surrogate implementation, but does not rely on constraint approximations. If time is scarce, or in a setting of very frequent/automatic application of cross-efficiency benchmarking, the linear Sexton implementation is the method of choice.

7.3 Limitations of study

The limitations of this study arise from three main areas:

First, while a data-set of over 100 warehouses fulfills all DEA minimum DMU requirements, in a estimated market of ≥10,000 facilities in the Netherlands and Belgium, a larger data-set would be preferable. Especially industry clustering and tailored comparisons are limited by the sample size. Next to that, the temporal analysis was done using recollection from respondents about the warehouse's state in the past. A true panel survey, preferably performed over a longer time span than five years, could help in more accurately grasping trends in warehousing performance.

Second, the model was based on a pre-tested warehouse questionnaire, which was modified to capture warehouse automation more granularly. A combination of existing questions, experience and personal judgment was used to develop the automation section of the questionnaire, because a holistic framework to capture mutually exclusive, collectively exhaustive automation components is not available, which limited the automation scoring precision.

Third, the current research on cross-efficiency with variable returns to scale is at a too early stage to circumvent all problems that arise during application, to reliably use VRS cross-efficiency yet. As the DMUs in the sample exhibit variable returns to scale, which were not considered for the cross-efficiency computations, the obtained scores in this thesis are biased by ignoring the scale (in)efficiencies of DMUs.

7.4 Further research

Analogous to the limitations of the study, further research is needed on the constituting components of warehouse efficiency, such as specific product categories, value chain positions, and relevant inputs and outputs. Especially as the industry is evolving at a quick pace, a bi-annually efficiency survey would be of interest for researchers and professionals alike. Alternatively, a web-hosted solution could be programmed to provide warehouse managers with an interface to provide operational data and have their warehouse ranked through the cross-efficiency code that is made publicly available by the author. This way, warehouse data could continuously be gathered, while participants receive results instantly.

As mentioned above, a framework to entirely capture and score warehouse automation is not available. Advances in that area in times of AGVs, virtual reality, deep learning and autonomous logistics should prove to be fruitful. Another aspect that finds itself underrepresented in current publications is the effect of general management practices on warehouse efficiency. There are frameworks for specific areas such as lean-management implementations, but no framework to score a top-level management approach on all dimensions in respect to operational performance of warehouses.

From a methodological perspective, a future focus on VRS cross-efficiency implementations would be highly welcome to increase cross-efficiency acceptance, by removing a central limitation of current models. Further topics with development potential are the implementation of the temporal dimension into cross-efficiency as well as a framework for choosing the right type of inputs and outputs to achieve DEA models with the highest explanatory power.

8. Bibliography

Aczél, János and Fred S. Roberts (1989). “On the possible merging functions”. In: *Mathematical Social Sciences* 17.3, pp. 205–243. ISSN: 01654896. DOI: 10.1016/0165-4896(89)90054-1.

Álvarez, Inmaculada C., Javier Barbero, and José L. Zofío (2016). “A Data Envelopment Analysis Toolbox for MATLAB”. In: *Economic Analysis Working Paper Series*.

Andrejić, Milan, Nebojša Bojović, and Milorad Kilibarda (2013). “Benchmarking distribution centres using Principal Component Analysis and Data Envelopment Analysis: A case study of Serbia”. In: *Expert Systems with Applications* 40.10, pp. 3926–3933. ISSN: 09574174. DOI: 10.1016/j.eswa.2012.12.085.

Arndt, Stephan, Carolyn Turvey, and Nancy C. Andreasen (1999). “Correlating and predicting psychiatric symptom ratings: Spearman’s r versus Kendall’s tau correlation”. In: *Journal of Psychiatric Research* 33.2, pp. 97–104. ISSN: 0022-3956. DOI: http://doi.org/10.1016/S0022-3956(98)90046-2.

Avkiran, Necmi (2006). “Productivity Analysis in the Service Sector with Data Envelopment Analysis”. In: *SSRN Electronic Journal*. DOI: 10.2139/ssrn.2627576.

Balk, Bert M. (2017). “What is Cross-Efficiency?”

Banaszewska, Agata, Frans Cruijssen, Wout Dullaert, and Johanna C. Gerdessen (2012). “A framework for measuring efficiency levels - The case of express depots”. In: *International Journal of Production Economics* 139.2. Compassionate Operations, pp. 484–495. ISSN: 0925-5273. DOI: http://dx.doi.org/10.1016/j.ijpe.2012.05.003.

Banker, Rajiv D., Hsihui Chang, and William W. Cooper (1996). “Equivalence and implementation of alternative methods for determining returns to scale in data envelopment analysis”. In: *European Journal of Operational Research* 89.3, pp. 473–481. ISSN: 0377-2217. DOI: http://dx.doi.org/10.1016/0377-2217(95)00044-5.

Banker, Rajiv D., Abraham Charnes, and William W. Cooper (1984). “Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis”. In: *Management Science* 30.9, pp. 1078–1092.

Banker, Rajiv D. and Robert M. Thrall (1992). “Estimation of returns to scale using data envelopment analysis”. In: *European Journal of Operational Research* 62.1, pp. 74–84. ISSN: 0377-2217. DOI: http://dx.doi.org/10.1016/0377-2217(92)90178-C.

Bogetoft, Peter (2000). “DEA and Activity Planning under Asymmetric Information”. In: *Journal of Productivity Analysis* 13.1, pp. 7–48. ISSN: 1573-0441. DOI: 10.1023/A:1007812822633.

Bogetoft, Peter and Lars Otto (2011). *Benchmarking with DEA, SFA, and R*. Ed. by Martina Bihn. Vol. 2011. International Series in Operations Research & Management Science 157. Springer New York. Chap. Efficiency Measures, pp. 23–55. ISBN: 978-1-4419-7960-5 978-1-4419-7961-2. DOI: 10.1007/978-1-4419-7961-2\-2.

Bogue, Robert (2016). “Growth in e-commerce boosts innovation in the warehouse robot market”. In: *Industrial Robot: An International Journal* 43.6, pp. 583–587. DOI: 10.1108/IR-07-2016-0194.

Bowersox, Donald J., David J. Closs, and M. Bixby Cooper (2013). *Supply Chain Logistics Management*. McGraw-Hill. ISBN: 9780071326216.

Chambers, Robert G., Yangho Chung, and Rolf Färe (1996). “Benefit and Distance Functions”. In: *Journal of Economic Theory* 70.2, pp. 407–419. ISSN: 0022-0531. DOI: 10.1006/jeth.1996.0096.

Charnes, Abraham, William W. Cooper, and E. Rhodes (1978). “Measuring the efficiency of decision making units”. In: *European Journal of Operational Research* 2.6, pp. 429–444. ISSN: 03772217. DOI: 10.1016/0377-2217(78)90138-8.

Charnes, Abraham, William W. Cooper, Lawrence M. Seiford, and J. Stutz (1982). “A multiplicative model for efficiency analysis”. In: *Socio-Economic Planning Sciences* 16.5, pp. 223–224. ISSN: 00380121. DOI: 10.1016/0038-0121(82)90029-5.

Charnes, Abraham, William W. Cooper, Lawrence M. Seiford, and J. Stutz (1983). “Invariant multiplicative efficiency and piecewise cobb-douglas envelopments”. In: *Operations Research Letters* 2.3, pp. 101–103. ISSN: 01676377. DOI: 10.1016/0167-6377(83)90014-7.

Cohen, Morris, Yu-Sheng Zhen, and Vipul Agrawal (1997). “Service parts logistics: a benchmark analysis”. In: *IIE Transactions* 29.8, pp. 627–639. ISSN: 1573-9724. DOI: 10.1023/A:1018513900264.

Cook, Wade D. and Joe Zhu (2014). “DEA Cobb-Douglas Frontier and Cross-Efficiency”. In: *Journal of the Operational Research Society* 65.2, pp. 265–268. ISSN: 0160-5682. DOI: 10.1057/jors.2013.13.

Cooper, William W., Lawrence M. Seiford, and Kaoru Tone (2007). *Data Envelopment Analysis*. Ed. by Rocio Torregrosa. Springer US. ISBN: 978-0-387-45281-4.

Cooper, William W., Lawrence M. Seiford, and Joe Zhu (2000). “A unified additive model approach for evaluating inefficiency and congestion with associated measures in DEA”. In: *Socio-Economic Planning Sciences* 34.1, pp. 1–25. ISSN: 0038-0121. DOI: http://dx.doi.org/10.1016/S0038-0121(99)00010-5.

Cooper, William W., Lawrence M. Seiford, and Joe Zhu (2011). *Handbook on Data Envelopment Analysis*. Vol. 2. International Series in Operations Research & Management Science. Boston, MA: Springer US, p. 497. ISBN: 144196150X. DOI: 10.1007/978-1-4419-6151-8-1.

Daraio, Cinzia and Léopold Simar (2006). *Advanced Robust and Nonparametric Methods in Efficiency Analysis*. en. Vol. 4. Studies in Productivity and Efficiency. Boston, MA: Springer US. ISBN: 9780387351551.

de Koster, M.B.M. and Bert M. Balk (2008). “Benchmarking and Monitoring International Warehouse Operations in Europe”. In: *Production and Operations Management* 17.2, pp. 175–183. ISSN: 1937-5956. DOI: 10.3401/poms.1080.0013.

de Koster, M.B.M., Bert M. Balk, and W.T.I. Van Nus (2009). “On using DEA for benchmarking container terminals”. In: *International Journal of Operations & Production Management* 29.11, pp. 1140–1155. ISSN: 0144-3577. DOI: 10.1108/01443570911000168.

de Koster, M.B.M., Tho Le-Duc, and Nima Zaerpour (2012). “Determining the number of zones in a pick-and-sort order picking system”. In: *International Journal of Production Research* 50.3, p. 28. DOI: ReportERS-2005-029-LIS.

de Koster, M.B.M., Daan Stam, and Bert M. Balk (2011). “Accidents happen: The influence of safety-specific transformational leadership, safety consciousness, and hazard reducing systems on warehouse accidents”. In: *Journal of Operations Management* 29.7-8, pp. 753–765. ISSN: 02726963. DOI: 10.1016/j.jom.2011.06.005.

de Koster, M.B.M. and P.M.J. Warffemius (2005). “American, Asian and third-party international warehouse operations in Europe: A performance comparison”. In: *International Journal of Operations & Production Management* 25.8, pp. 762–780. ISSN: 0144-3577. DOI: 10.1108/01443570510608592.

de Vries, Jelle, M.B.M. de Koster, and Daan Stam (2016). “Exploring the role of picker personality in predicting picking performance with pick by voice, pick to light and RF-terminal picking”. In: *International Journal of Production Research* 7543.July, pp. 1–15. ISSN: 0020-7543. DOI: 10.1080/00207543.2015.1064184.

Dinno, Alexis (2009). “Exploring the Sensitivity of Horn’s Parallel Analysis to the Distributional Form of Random Data”. In: *Multivariate Behavioral Research* 44 (3), pp. 362–88.

Doyle, John and Rodney H. Green (1994). “Efficiency and Cross-Efficiency in DEA: Derivations, Meanings and Uses.” In: *Journal of the Operational Research Society* 45.5, pp. 567–578. ISSN: 01605682. DOI: 10.1057/jors.1994.84.

Doyle, John and Rodney H. Green (1995). "Cross-Evaluation in DEA - Improving Discrimiation Among DMUs". In: *INFOR: Information Systems and Operational Research* 33.3, pp. 205–222.

Dyson, R.G., R. Allen, A.S. Camanho, V.V. Podinovski, C.S. Sarrico, and E.A. Shale (2001). "Pitfalls and protocols in DEA". In: *European Journal of Operational Research* 132.2. Data Envelopment Analysis, pp. 245–259. ISSN: 0377-2217. DOI: http://doi.org/10.1016/S0377-2217(00)00149-1.

Emmett, Stuart (2005). *Excellence in Warehouse Management*. 1st. John Wiley & Sons, Ltd, p. 276. ISBN: 9780470015315.

Eurostat (2017). *Annual detailed enterprise statistics for services (NACE Rev. 2 H-N and S95)*. URL: http://appsso.eurostat.ec.europa.eu/nui/show.do.

Faber, Nynke, M.B.M. de Koster, and Ale Smidts (2013). "Organizing warehouse management." In: *International Journal of Operations & Production Management* 33.9, pp. 1230–1256. ISSN: 01443577. DOI: 10.1108/IJOPM-12-2011-0471.

Färe, Rolf, Dimitris Margaritis, Paul Rouse, and Israfil Roshdi (2015). "Estimating the hyperbolic distance function: A directional distance function approach". In: *European Journal of Operational Research* 254.1, pp. 312–319. ISSN: 03772217. DOI: 10.1016/j.ejor.2016.03.045.

Farrell, Michael J. (1957). "The Measurement of Productive Efficiency". In: *Journal of the Royal Statistical Society* 120.3, pp. 253–290.

Fried, Harold O., C. A. Knox Lovell, Shelton S. Schmidt, and William H. Greene (2008). *The Measurement of Productive Efficiency and Productivity Growth*. Oxford University Press, p. 656. ISBN: 9780195183528.

Golany, Boaz and Yaakov Roll (1989). "An application procedure for DEA". In: *Omega* 17.3, pp. 237–250. ISSN: 0305-0483. DOI: http://dx.doi.org/10.1016/0305-0483(89)90029-7.

Green, Rodney H. and Wade D. Cook (2004). "A Free Coordination Hull Approach to Efficiency Measurement". In: *The Journal of the Operational Research Society* 55.10, pp. 1059–1063. ISSN: 01605682, 14769360.

Gu, Jinxiang, Marc Goetschalckx, and Leon F. McGinnis (2007). "Research on warehouse operation: A comprehensive review". In: *European Journal of Operational Research* 177.1, pp. 1–21. ISSN: 03772217. DOI: 10.1016/j.ejor.2006.02.025.

Hackman, Steven T., Edward H. Frazelle, Paul M. Griffin, Susan O. Griffin, and Dimitra A. Vlasta (2001). "Benchmarking Warehousing and Distribution Operations: An Input-Output Approach". In: *Journal of Productivity Analysis* 16.1, pp. 79–100. ISSN: 0895562X. DOI: 10.1023/A:1011155320454.

Hamberg, Roelof and Verriet Jacques (2012). *Automation in Warehouse Development*. Ed. by Roelof Hamberg and Verriet Jacques. Springer. 256 pp. ISBN: 1447159748.

Hoff, Ayoe (2007). "Second stage DEA: Comparison of approaches for modelling the DEA score". In: *European Journal of Operational Research* 181.1, pp. 425–435. ISSN: 0377-2217. DOI: https://doi.org/10.1016/j.ejor.2006.05.019.

Johnson, Andrew and Leon F. McGinnis (2011). "Performance measurement in the warehousing industry". In: *IIE Transactions* 43.3, pp. 220–230. ISSN: 0740-817X. DOI: 10.1080/0740817X.2010.491497.

Kumbhakar, Subal C. and C. A. Knox Lovell (2003). *Stochastic Frontier Analysis*. Cambridge University Press. Ed. by William H. Greene. New York: Cambridge University Press.

Larco, José Antonio, M.B.M. de Koster, Kees Jan Roodbergen, and Jan Dul (2016). "Managing warehouse efficiency and worker discomfort through enhanced storage assignment decisions". In: *International Journal of Production Research* 7543.1, pp. 1–16. ISSN: 0020-7543. DOI: 10.1080/00207543.2016.1165880.

Liang, Liang, Wu Jie, Wade D. Cook, and Joe Zhu (2008). "The DEA Game Cross-Efficiency Model and Its Nash Equilibrium". In: *Operations Research* 56.5, pp. 1278–1288. DOI: 10.1287/opre.1070.0487.

Lim, Sungmook and Joe Zhu (2015). "DEA Cross Efficiency Under Variable Returns to Scale". In: *Data Envelopment Analysis: A Handbook of Models and Methods*. Ed. by Joe Zhu. Boston, MA: Springer US, pp. 45–66. ISBN: 978-1-4899-7553-9. DOI: 10.1007/978-1-4899-7553-9-3.

Márquez, Fausto Pedro García and Benjamin Lev (2015). *Advanced Business Analytics*. Ed. by Alberto Pliego Marugán. Springer Nature. DOI: 10.1007/978-3-319-11415-6.

Muilerman, Gert-Jan (2001). "Time-based logistics". PhD thesis. Delft University of Technology.

Onstein, A.T.C., J.G.S.N. Visser, J.C. van Ham, and L.A. Tavasszy (2016). *Trends in Distribution Centres and Their Locations: Sprawl and Polarization*. Tech. rep. Amsterdam: Hogeschool van Amsterdam, p. 12.

Sauermann, Henry and Michael Roach (2013). "Increasing web survey response rates in innovation research: An experimental study of static and dynamic contact design features". In: *Research Policy* 42.1, pp. 273–286. ISSN: 0048-7333. DOI: http://doi.org/10.1016/j.respol.2012.05.003.

Schroeder, Larry D. and David L. Sjoquist (2011). "Understanding Regression Analysis". In: ed. by Paula E. Stephan. SAGE Publications. Chap. Multiple Linear Regression, pp. 30–36.

Sexton, Thomas R. (1986). "The methodology of data envelopment analysis". In: *New Directions for Program Evaluation* 1986.32, pp. 7–29. ISSN: 1551-2371. DOI: 10.1002/ev.1438.

Sexton, Thomas R., Richard H. Silkman, and Andrew J. Hogan (1986). "Data envelopment analysis: Critique and extension". In: *New Directions for Program Evaluation* 1986.32, pp. 73–105.

Thompson, Russell G., F. D. Singleton Jr., Robert M. Thrall, Barton A. Smith, and Meg Wilson (1986). "Comparative Site Evaluations for Locating a High-Energy Physics Lab in Texas". In: *Interfaces* 16.6, pp. 35–49. ISSN: 00922102. DOI: 10.2307/25060889.

van der Gaast, Jelmer Pier (2015). *Stochastic models for order picking systems*. ERIM PhD Series in Research in Management, p. 246. ISBN: 9789058924551.

Waldrop, B. Y. M. Mitchell (2016). "More Than More". In: *Nature* 530, p. 145.

Wang, Qian, Richard McIntosh, and Martin Brain (2010). "A new-generation automated warehousing capability". In: *International Journal of Computer Integrated Manufacturing* 23.6, pp. 565–573. ISSN: 0951-192X. DOI: 10.1080/09511921003706215.

Yang, Hongliang and Michael Pollitt (2009). "Incorporating Both Undesirable Outputs and Uncontrollable Variables into DEA: The Performance of Chinese Coal-Fired Power Plants". In: *European Journal of Operational Research* 197.3, pp. 1095–1105. ISSN: 0377-2217. DOI: http://dx.doi.org/10.1016/j.ejor.2007.12.052.

Yu, Ming-Min, Shih-Chan Ting, and Mu-Chen Chen (2010). "Evaluating the cross-efficiency of information sharing in supply chains". In: *Expert Systems with Applications* 37.4, pp. 2891–2897. ISSN: 09574174. DOI: 10.1016/j.eswa.2009.09.048.

Zeydan, Mithat and Cüneyt Colpan (2009). "A new decision support system for performance measurement using combined fuzzy TOPSIS/DEA approach". In: *International Journal of Production Research* 47.15, pp. 4327–4349. DOI: 10.1080/00207540802662870.

Appendices

A Abbreviations

Abbreviation	Definition
AHP	Analytical Hierarchy Process
ANOVA	Analysis of Variance
BCC	Banker, Charnes and Cooper
C+E	Construction and Engineering
CCR	Charnes, Cooper and Rhodes
CRS	Constant Returns to Scale
DEA	Data Envelopment Analysis
DMU	Decision Making Unit
DRS	Decreasing Returns to Scale
EU	European Union
ERP	Enterprise Resource Planning
FDH	Free Disposal Hull
FTE	Full Time Employee
GDP	Gross Domestic Product
IO	Input Orientation
IRS	Increasing Returns to Scale
KS	Kolmogorov-Smirnov
KW	Kruskal-Wallis
MATLAB	Matrix Laboratory
MLR	Multiple Linear Regression
OLS	Ordinary Least Squares
PA	Parallel Analysis
PCA	Principal Component Analysis
PPS	Production Possibility Set
RTS	Returns to Scale
SFA	Stochastic Frontier Analysis
SKU	Stock Keeping Unit
TLN	Transport en Logistiek Nederland
TOPSIS	Technique of Order Preference by Similarity to the Ideal Solution
VRS	Variable Returns to Scale
WMS	Warehouse Management Software
WSR	Wilcoxon Signed Rank Test
3PL	Third Party Logistics Provider

Table 2: List of abbreviations.

B Survey questionnaire

Warehouse Efficiency Survey 2017 NL & BE - Bosch

Q1.1

Q1.2 In cooperation with TLN and the Erasmus University Rotterdam (Christian Kaps and Prof. René de Koster), evofenedex has started research to evaluate the effect of technological developments on warehouse efficiency. We ask you to participate in this research by filling out this survey. What is in it for you? When you participate, you will receive a personalized result of our study, in which we benchmark & rank your warehouse's operational efficiency against the entire set of warehouses as well as an analysis how it competes directly against facilities in your industry. The report will include your operation's overall efficiency as computed through the data you provide & the analysis of individual inputs/outputs (e.g. number of people, order lines, etcetera). For a sample benchmark click here. How can you participate? To help with the study and receive the report, please fill out this survey. Filling it out takes 10-15 minutes and focuses on high level operations data of your warehouse. The survey should be completed by the (assistant) logistics manager of your warehouse facility. The questionnaire focuses on operational changes in your facility during the last 5 years. Each warehouse location of your organization can participate separately. If you have any questions or feedback, please contact Christian.Kaps@student.eur.nl. Thank you for your participation René de Koster Christian Kaps On behalf of EUR, evofenedex & TLN

Q2.1 Questions to be answered by the (assistant) logistics/warehouse manager

Q2.2 Company / warehouse information

- Company name (1)
- Name of location (2)
- Address of warehouse (3)
- Zipcode and city (4)

Q2.3 Personal information

- Name (1)
- Function (2)
- Telephone number (3)
- E-mail address (report will be sent to this address) (4)

Q2.4 Which product category categorizes your warehouse best? (max 2 selections)
- ❑ Automotive (1)
- ❑ Chemicals/Oil/Gas (2)
- ❑ Construction (3)
- ❑ Consumer goods (4)
- ❑ Electronics (5)
- ❑ Engineering (6)
- ❑ Groceries/Food (7)
- ❑ Household appliances (8)
- ❑ Logistics (9)
- ❑ Military/Defense (10)
- ❑ Pharmacy (11)
- ❑ Retail (non-food) (12)
- ❑ Textiles (13)
- ❑ Other, please specify: (14) ____________________

Q2.5 Which of the following options describes the position of your warehouse in your company's value chain best?
- ❍ Production warehouse (1)
- ❍ Wholesale warehouse (2)
- ❍ Retail warehouse (3)

Q2.6 Is your warehouse operated by a logistics service provider?
- ❍ Yes - this warehouse is public, offering services to multiple customers (1)
- ❍ Yes - this warehouse is dedicated/contracted mainly for one specific customer (2)
- ❍ No (3)

Q3.1 How many employees including temporary workers are active in your warehouse at the moment and how many were active 5 years ago (indicated as FTE - full time equivalent; direct - active on the warehouse floor; indirect - not operational on the floor)?

	2017 (1)	5 years ago (2)
Direct FTEs (1)		
Indirect FTEs (4)		

Q3.2 What is the floor space of the warehouse at the moment including mezzanines (in m²)? And what was it 5 years ago?

	2017 (1)	5 years ago (2)
Floor Space (in m²) (1)		

Q3.3 What floor space percentage of the warehouse at the moment consists of cold storage and what was that percentage 5 years ago?

	2017 (1)	5 years ago (2)
% Cold storage (1)		

Q4.1 Approximately how many unique article numbers (Stock Keeping Units) are simultaneously stored in your warehouse at the moment on average and how many were stored 5 years ago?

	2017 (1)	5 years ago (2)
SKUs (1)		

Q4.2 What is the average number of shipping order lines (outgoing order lines) per day at the moment and how high was this number 5 years ago?

	2017 (1)	5 years ago (2)
Order lines per day (1)		

Q4.3 What percentage of shipping order lines was error-free on average over the last year and how many 5 years ago? (examples of errors are: faulty quantities, deliveries not on time (too early or too late), packing mistakes, product errors, incorrect component/modules. Select only one field per period)

	Error-free order lines	Error-free order lines
	2017 (1)	5 years ago (1)
No information available / not tracked (1)	❑	❑
Below 90% error-free order lines (2)	❑	❑
90-95% error-free order lines (3)	❑	❑
95-97% error-free order lines (4)	❑	❑
97-98% error-free order lines (5)	❑	❑
98-99% error-free order lines (6)	❑	❑
99-99.5% error-free order lines (7)	❑	❑
99.5-99.9% error-free order lines (8)	❑	❑
Over 99.9% error-free order lines (9)	❑	❑

Q5.1 Besides the standard processes like handling incoming shipments, visual inspection, storage and order picking, which occur in all warehouses, additional special processes may be performed as well. Which of the following special processes are performed at the moment by your warehouse? Which were also performed 5 years ago? (check all that apply)

	Special processes	Special processes
	2017 (1)	5 years ago (1)
Cross-docking / Transshipment (1)	❑	❑
Transport planning (2)	❑	❑
Internal product movements (relocating stock) for optimization (3)	❑	❑
Re-packing / sealing of products (4)	❑	❑
(Re)coding of products (5)	❑	❑
Quality control of received products (6)	❑	❑
Adding (promotional) material (7)	❑	❑
Receiving and processing customer returns / return logistics (8)	❑	❑
Pricing (9)	❑	❑
Assembly (10)	❑	❑
Other special processes not mentioned yet, namely (11)	❑	❑

Q5.2 Which of the following automated systems were already used to support the processes in the warehouse 5 years ago? (Indicate the number of units of the selected systems where applicable)

❑ Automated Storage/Retrieval system (# cranes) (1) ____________________
❑ Mini-load system (# cranes) (2) ____________________
❑ Flow racks (3)
❑ Conveyor belts (# meters) (4) ____________________
❑ Automated sorter (# output lanes / chutes) (5) ____________________
❑ Automated packing/labelling machine (# machines) (6) ____________________
❑ Automated box stacker (# stackers) (7) ____________________
❑ Automated guided vehicles (# vehicles) (8) ____________________
❑ Robots (# robots) (9) ____________________
❑ Horizontal/vertical carousel or elevator storage module (# units) (10) ____________________
❑ Multi-deep storage of pallets/totes with gantry crane or conveyors (11)
❑ Radio Frequency ID technology (12)
❑ Augmented reality technology (13)
❑ Automated process gamification (15)
❑ Others, for example: (14) ____________________

Q5.3 Which new automated systems were added to support the processes in the warehouse during the last 5 years ? (Indicate the number of units of the selected systems where applicable)

- ❑ Automated Storage/Retrieval system (# cranes) (1) ____________________
- ❑ Mini-load system (# cranes) (2) ____________________
- ❑ Flow racks (3)
- ❑ Conveyor belts (# meters) (4) ____________________
- ❑ Automated sorter (# output lanes / chutes) (5) ____________________
- ❑ Automated packing/labelling machine (# machines) (6) ____________________
- ❑ Automated box stacker (# stackers) (7) ____________________
- ❑ Automated guided vehicles (# vehicles) (8) ____________________
- ❑ Robots (# robots) (9) ____________________
- ❑ Horizontal/vertical carousel or elevator storage module (# units) (10) ____________________
- ❑ Multi-deep storage of pallets/totes with gantry crane or conveyors (11)
- ❑ Radio Frequency technology (12)
- ❑ Augmented reality technology (13)
- ❑ Automated process gamification (15)
- ❑ Others, for example: (14) ____________________

Q5.4 What type of information system does the warehouse make use of at the moment and which type was used 5 years ago? (a standard system is comprised of less than 20% tailor-made software. Select only one field per period)

	Information system	Information system
	2017 (1)	5 years ago (1)
No information system used (e.g. Excel or paper-based tracking) (1)	❑	❑
A standard ERP warehouse module (2)	❑	❑
A standard ERP warehouse module with more than 20% customization (3)	❑	❑
A standard WMS (Warehouse Management System) package (4)	❑	❑
A standard WMS package with more than 20% customization (5)	❑	❑
A tailor made/customized system (6)	❑	❑

Q5.5 Please indicate how you perceive your warehouse's performance at the moment compared to that of competitors, when facing the situations mentioned below.

	Much worse than competition (1)	Worse than competition (2)	Equal to competition (3)	Better than competition (4)	Much better than competition (5)	Not Applicable (6)
Handling fluctuations in order quantities (1)	❍	❍	❍	❍	❍	❍
Handling alterations in customer orders (number/type of SKU) (2)	❍	❍	❍	❍	❍	❍
Expediting specific orders at customer's request (3)	❍	❍	❍	❍	❍	❍
Adding new SKUs to the assortment (4)	❍	❍	❍	❍	❍	❍
Implementation of IT changes (5)	❍	❍	❍	❍	❍	❍
Handling labor requirement fluctuations (6)	❍	❍	❍	❍	❍	❍

Q5.6 Please indicate how you perceived your warehouse's performance 5 years ago compared to that of competitors, when having faced the situations mentioned below.

	Much worse than competition (1)	Worse than competition (2)	Equal to competition (3)	Better than competition (4)	Much better than competition (5)	Not Applicable (6)
Handling fluctuations in order quantities (1)	❍	❍	❍	❍	❍	❍
Handling alterations in customer orders (number/type of SKU) (2)	❍	❍	❍	❍	❍	❍
Expediting specific orders at customer's request (3)	❍	❍	❍	❍	❍	❍
Adding new SKUs to the assortment (4)	❍	❍	❍	❍	❍	❍
Implementation of IT changes (5)	❍	❍	❍	❍	❍	❍
Handling labor requirement fluctuations (6)	❍	❍	❍	❍	❍	❍

Q5.7 As the final question, please indicate at which frequency the following activities are currently performed in your warehouse at the moment and how often they were performed 5 years ago.

	2017					5 years ago				
	Never (1)	Rarely (2)	Sometimes (3)	Frequently (4)	Very Frequently (5)	Never (1)	Rarely (2)	Sometimes (3)	Frequently (4)	Very Frequently (5)
Pre-shift meetings with shop-floor workers (1)	❍	❍	❍	❍	❍	❍	❍	❍	❍	❍
Individualized training plans (2)	❍	❍	❍	❍	❍	❍	❍	❍	❍	❍
Incentive programs for employees (3)	❍	❍	❍	❍	❍	❍	❍	❍	❍	❍
Strategies to empower teams/individuals and hold them accountable (4)	❍	❍	❍	❍	❍	❍	❍	❍	❍	❍
Worker rotation through different jobs / areas (5)	❍	❍	❍	❍	❍	❍	❍	❍	❍	❍
Internal information visibility / productivity benchmarks (6)	❍	❍	❍	❍	❍	❍	❍	❍	❍	❍
Scheduled shop floor walks by management (7)	❍	❍	❍	❍	❍	❍	❍	❍	❍	❍
5S / Continuous improvement measures (8)	❍	❍	❍	❍	❍	❍	❍	❍	❍	❍
Employee involvement / suggestion program (9)	❍	❍	❍	❍	❍	❍	❍	❍	❍	❍
Customer involvement / suggestion program (10)	❍	❍	❍	❍	❍	❍	❍	❍	❍	❍

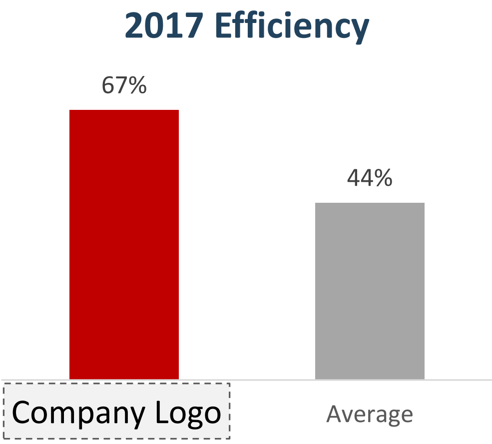
Q5.8 You have reached the end of the survey. Would you like to receive a personalized efficiency benchmark of your warehouse, based on the results of this study? (Reports will be sent in June/July to the provided e-mail address)

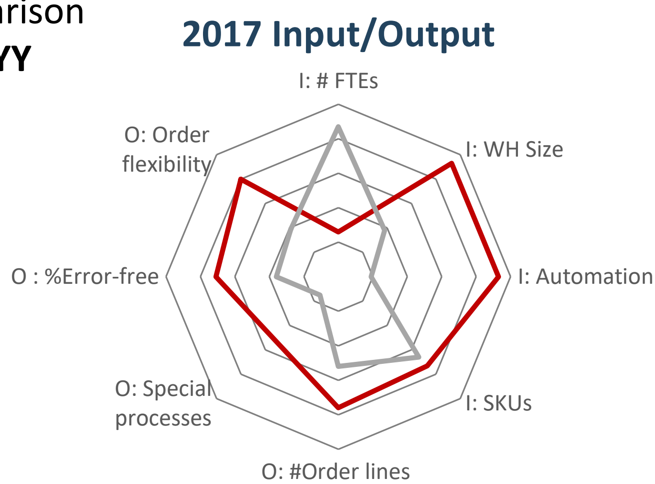
- ❍ Yes (1)
- ❍ No (2)

C Sample report

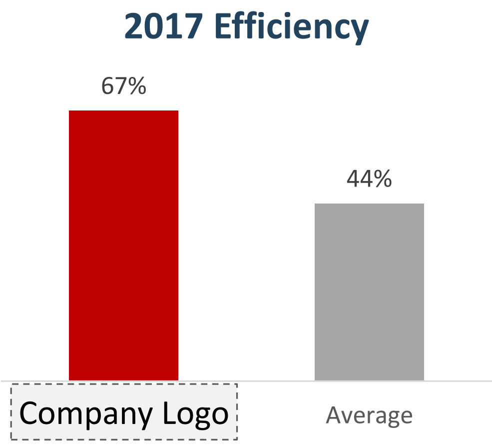
Warehouse Efficiency Survey 2017 NL & BE

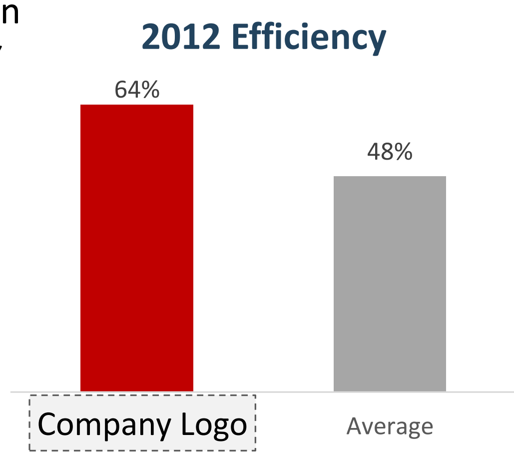
Overall comparison
Rank: XX/YY



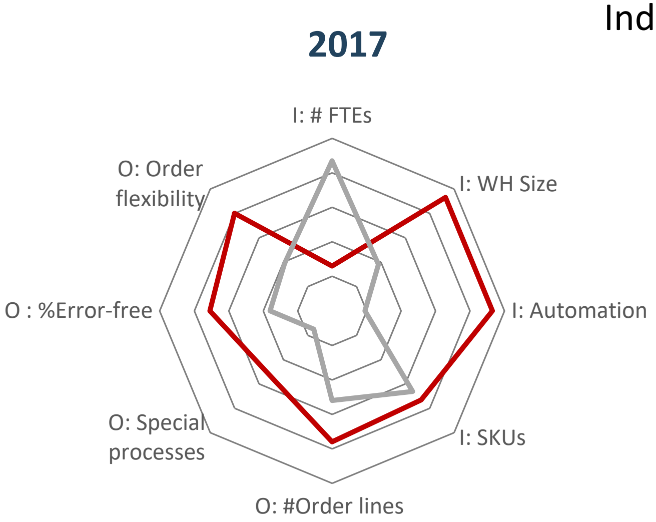


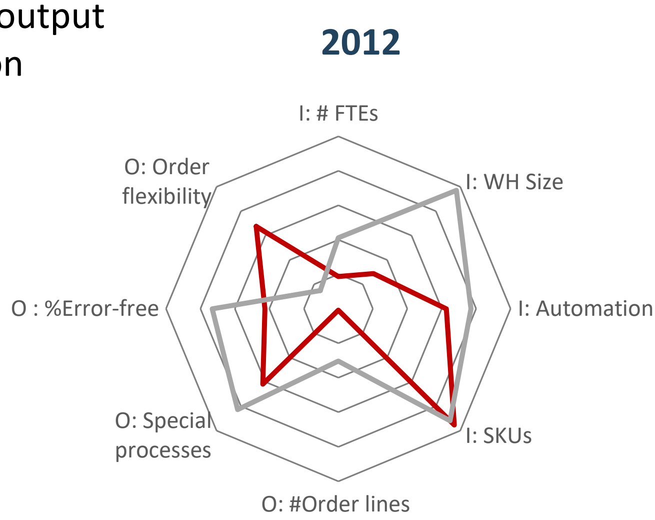
Industry comparison
Rank 2017: XX/YY
Rank 2012: xx/yy





Industry input/output comparison





D Input and output score distribution of sample

Figure 1: Automation scores across the warehouse sample in 2012 and 2017.

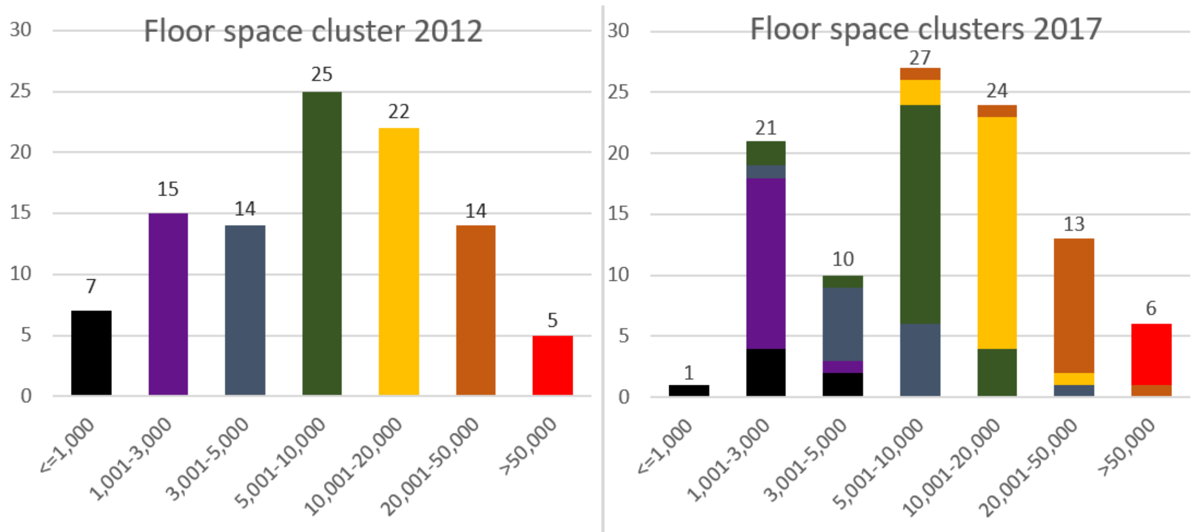


Figure 2: Floor space clusters across the warehouse sample in 2012 and how warehouses have changed between clusters until 2017.

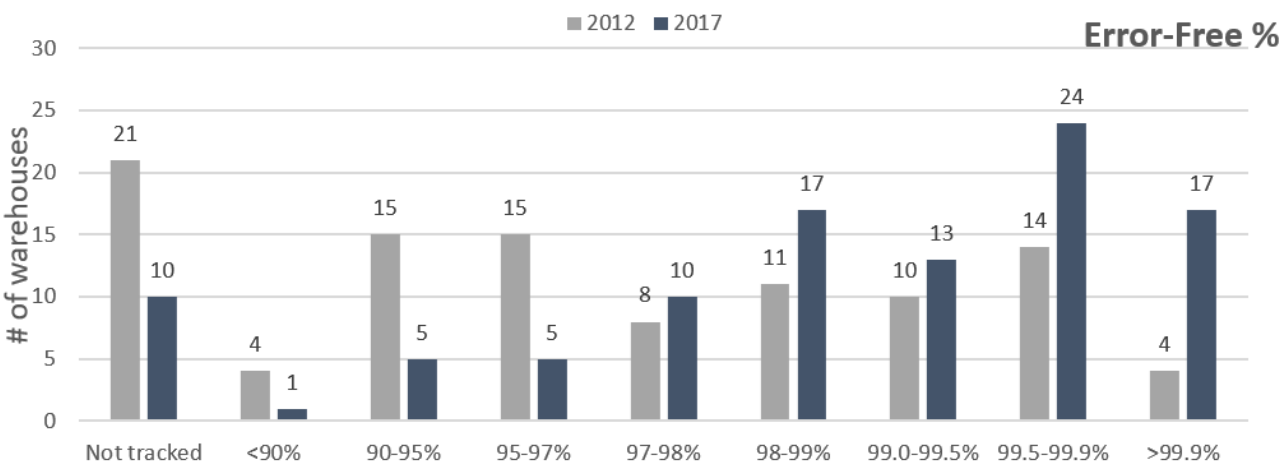


Figure 3: Error free % across the warehouse sample in 2012 and 2017.

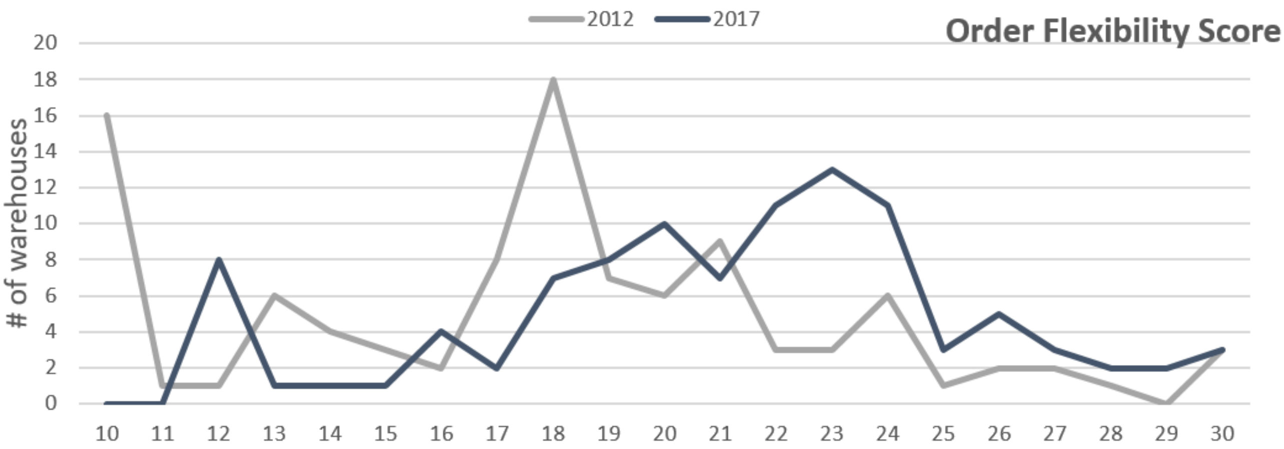


Figure 4: Order flexibility scores across the warehouse sample in 2012 and 2017.

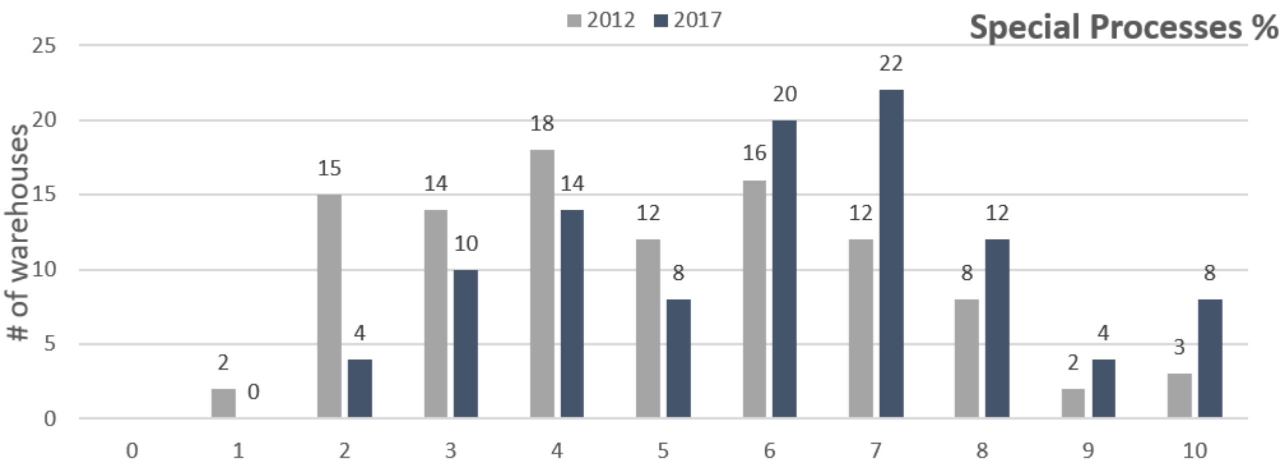


Figure 5: Special process scores across the warehouse sample in 2012 and 2017.

E DEA input and output tables

DMU Inputs 2017	FTEs	Floor Space	SKUs	Automation
DMU001	64	14.499	4.600	7
DMU002	16	2.400	2.500	7
DMU003	22	3.000	5.000	14
DMU004	53	1.500	5.600	7
DMU005	135	20.000	1.200	5
DMU006	40	1.050	950	8
DMU007	90	14.000	5.000	14
DMU008	240	45.000	250	14
DMU010	170	4.000	5.300	11
DMU012	22	15.000	2.400	6
DMU014	100	16.000	50.000	9
DMU017	41	59.000	23.760	7
DMU019	16	5.800	37.000	7
DMU020	45	8.000	15.000	8
DMU022	18	8.000	350.000	8

DMU Inputs 2017	FTEs	Floor Space	SKUs	Automation
DMU023	100	20.000	3.500	10
DMU024	31	34.000	500	6
DMU026	10	5.700	1.850	7
DMU027	12	500	200	7
DMU028	6	1.500	2.200	2
DMU030	23	8.140	12.000	11
DMU033	69	7.500	40.000	8
DMU034	19	5.000	6.000	12
DMU037	70	18.000	1.000	6
DMU040	18	5.000	4.000	5
DMU041	14	5.000	712	4
DMU042	45	15.000	3.360	3
DMU043	14	10.000	950	4
DMU044	18	2.500	4.000	5
DMU045	5	2.500	250	4
DMU046	52	275.000	1.600	5
DMU047	36	9.600	1.800	5
DMU049	52	4.000	5.500	6
DMU050	7	6.600	1.253	3
DMU051	50	3.600	22.000	5
DMU052	295	90.000	60.000	5
DMU053	15	2.500	2.200	5
DMU054	17	12.000	5.000	5
DMU055	38	15.000	70.000	8
DMU056	60	8.000	4.500	8
DMU058	25	10.000	2.000	6
DMU059	11	7.000	10.000	8
DMU060	313	52.000	203.932	13
DMU061	40	6.634	850	9
DMU062	60	8.000	3.500	11
DMU063	32	4.000	400.000	4
DMU064	25	20.000	850	3
DMU065	162	10.000	9.000	10
DMU066	6	1.910	636	5
DMU067	14	1.280	500	7
DMU068	17	7.500	9.100	5
DMU069	45	16.500	2.600	7
DMU070	27	15.000	75.500	8
DMU071	5	1.500	700	5
DMU072	90	37.000	5.000	16
DMU073	103	50.000	1.500	8
DMU074	15	12.600	20.000	8
DMU076	64	18.000	5.000	7
DMU077	25	6.000	14.000	10
DMU080	13	10.000	5.000	5
DMU081	11	7.500	6.000	6
DMU082	23	8.000	10.000	5
DMU083	19	10.000	650	9
DMU086	13	3.000	4.300	4

DMU Inputs 2017	FTEs	Floor Space	SKUs	Automation
DMU087	243	55.000	1.520	6
DMU089	260	19.000	115.000	8
DMU090	10	26.000	62.000	5
DMU091	251	17.500	110.000	7
DMU093	120	50.000	2.600	8
DMU094	89	4.515	28.033	7
DMU096	19	2.400	17.600	7
DMU097	95	12.000	7.000	4
DMU098	16	2.000	900	2
DMU099	53	30.000	2.600	2
DMU100	11	2.550	2.780	3
DMU101	17	1.260	18.000	9
DMU102	180	19.000	7.000	6
DMU103	29	9.500	12.000	5
DMU104	9	5.000	1.000	2
DMU105	35	8.000	3.600	5
DMU106	13	1.900	19.000	4
DMU107	6	1.400	1.100	3
DMU108	15	1.885	1.500	3
DMU109	40	7.200	32.500	9
DMU111	14	1.500	25.000	7
DMU113	26	9.000	12.000	5
DMU114	70	13.000	4.600	4
DMU115	5	2.500	6.000	5
DMU116	38	13.500	3.200	14
DMU117	40	7.000	1.673	5
DMU118	6	4.000	10.000	8
DMU119	13	10.000	2.128	5
DMU122	33	19.000	500	5
DMU123	17	14.000	1.500	6
DMU124	40	37.000	20.000	8
DMU125	9	15.000	100	6
DMU126	350	140.000	25.000	15
DMU127	97	50.000	3.000	9
DMU128	270	34.000	6.500	11
DMU129	27	35.000	16.000	6
DMU130	83	45.000	12.000	7
DMU131	48	28.000	2.500	6

Table 3: DEA input table 2017.

DMU Inputs 2012	FTEs	Floor Space	SKUs	Automation
DMU001	94	14.499	4.400	6
DMU002	20	1.600	2.000	8
DMU003	15	2.500	4.000	9
DMU004	32	1.200	4.000	7
DMU005	105	12.000	700	5
DMU006	6	450	200	7

DMU Inputs 2012	FTEs	Floor Space	SKUs	Automation
DMU007	115	14.000	3.500	10
DMU008	165	30.000	175	7
DMU010	70	1.200	2.000	10
DMU012	17	15.000	2.200	6
DMU014	120	12.000	60.000	4
DMU017	47	59.000	33.480	6
DMU019	16	5.800	25.000	7
DMU020	36	8.000	12.000	4
DMU022	12	4.000	250.000	7
DMU023	110	20.000	4.000	7
DMU024	22	24.000	300	3
DMU026	9	5.700	1.590	7
DMU027	7	400	200	6
DMU028	8	3.000	2.500	2
DMU030	30	5.740	14.000	4
DMU033	32	4.000	18.000	2
DMU034	12	5.000	4.500	12
DMU037	60	13.000	1.000	5
DMU040	10	4.750	3.000	4
DMU041	14	500	847	7
DMU042	31	9.600	2.500	3
DMU043	10	5.000	300	4
DMU044	11	1.000	2.500	4
DMU045	5	1.800	200	2
DMU046	54	275.000	1.200	2
DMU047	32	8.200	1.800	6
DMU049	28	8.000	3.500	3
DMU050	5	6.600	1.550	2
DMU051	47	3.600	15.000	6
DMU052	205	90.000	40.000	3
DMU053	37	2.500	1.200	2
DMU054	15	12.000	4.500	4
DMU055	33	15.000	70.000	3
DMU056	72	21.000	4.000	2
DMU058	25	10.000	4.000	4
DMU059	11	7.000	10.000	3
DMU060	312	32.000	134.403	11
DMU061	30	6.634	600	5
DMU062	180	8.000	4.000	8
DMU063	19	1.000	80.000	6
DMU064	31	20.000	900	5
DMU065	118	10.000	7.500	7
DMU066	8	2.400	1.946	4
DMU067	7	920	700	6
DMU068	17	7.500	9.100	6
DMU069	36	10.000	3.800	3
DMU070	51	20.000	120.000	7
DMU071	7	1.500	500	6
DMU072	63	12.000	2.000	7

DMU Inputs 2012	FTEs	Floor Space	SKUs	Automation
DMU073	136	50.000	1.000	6
DMU074	19	12.600	22.000	6
DMU076	21	8.000	1.000	1
DMU077	18	4.000	6.000	4
DMU080	32	10.000	4.150	4
DMU081	14	7.500	7.000	2
DMU082	13	13.000	3.000	1
DMU083	10	5.000	400	4
DMU086	10	6.000	3.500	6
DMU087	200	55.000	980	5
DMU089	257	17.500	100.000	5
DMU090	11	26.000	55.000	6
DMU091	257	17.500	100.000	7
DMU093	145	30.000	2.400	6
DMU094	83	4.104	29.957	7
DMU096	29	3.050	24.000	6
DMU097	115	12.000	6.500	7
DMU098	14	2.000	900	4
DMU099	70	27.000	3.200	4
DMU100	6	1.750	1.800	2
DMU101	14	1.550	13.000	2
DMU102	90	19.000	5.000	3
DMU103	32	9.500	10.500	7
DMU104	8	5.000	1.000	2
DMU105	15	8.000	2.400	5
DMU106	13	1.350	12.000	2
DMU107	7	700	900	3
DMU108	17	1.885	1.500	4
DMU109	44	7.200	19.500	4
DMU111	14	1.500	25.000	5
DMU113	47	10.500	12.000	5
DMU114	108	32.000	5.800	5
DMU115	19	7.000	32.000	3
DMU116	11	17.000	3.700	9
DMU117	17	5.000	500	1
DMU118	7	4.000	8.000	3
DMU119	15	10.000	2.313	4
DMU122	23	18.000	450	6
DMU123	8	6.000	250	5
DMU124	27	37.000	15.000	5
DMU125	9	15.000	100	4
DMU126	210	110.000	20.000	8
DMU127	97	50.000	3.000	2
DMU128	195	34.000	6.500	7
DMU129	21	30.000	11.000	6
DMU130	16	30.000	750	3
DMU131	29	5.000	1.000	2

Table 4: DEA input table 2012.

DMU Outputs 2017	Order Lines	Special Processes	Error Free %	Order Flexibility
DMU001	1.050	7	8	22
DMU002	400	2	7	23
DMU003	500	4	6	16
DMU004	1.000	4	3	13
DMU005	7.000	2	6	19
DMU006	5.250	5	5	26
DMU007	250	3	8	29
DMU008	20.000	6	3	23
DMU010	1.300	5	6	21
DMU012	420	6	8	26
DMU014	4.000	6	6	26
DMU017	1.740	6	4	19
DMU019	1.750	4	1	24
DMU020	3.000	6	6	22
DMU022	350	7	7	23
DMU023	15.000	8	8	19
DMU024	450	9	9	24
DMU026	285	5	7	21
DMU027	200	10	5	23
DMU028	200	4	9	19
DMU030	8.500	7	5	24
DMU033	3.300	7	9	28
DMU034	8.500	8	6	24
DMU037	1.100	3	7	24
DMU040	3.000	10	7	20
DMU041	268	7	7	22
DMU042	400	5	3	18
DMU043	65	7	6	19
DMU044	1.200	6	4	20
DMU045	175	3	5	12
DMU046	8.000	7	1	23
DMU047	3.700	6	8	23
DMU049	32.421	3	6	22
DMU050	1.200	6	9	25
DMU051	2.500	7	2	23
DMU052	45.000	4	6	27
DMU053	980	4	7	22
DMU054	130	8	8	12
DMU055	6.500	7	9	24
DMU056	1.250	7	8	21
DMU058	66	6	8	30
DMU059	60	7	8	30
DMU060	33.856	7	5	23
DMU061	3.400	5	8	20
DMU062	7.000	10	8	12
DMU063	15.000	7	3	20
DMU064	300	4	7	21
DMU065	19.200	6	4	20

DMU Outputs 2017	Order Lines	Special Processes	Error Free %	Order Flexibility
DMU066	350	8	6	18
DMU067	2.850	8	7	23
DMU068	1.200	4	9	18
DMU069	300	6	5	17
DMU070	3.600	7	9	22
DMU071	400	3	8	12
DMU072	3.000	3	7	26
DMU073	4.000	9	6	24
DMU074	250	5	9	20
DMU076	1.500	4	6	19
DMU077	2.100	10	5	21
DMU080	978	4	4	18
DMU081	330	6	6	12
DMU082	2.200	7	4	16
DMU083	250	7	5	21
DMU086	500	6	6	24
DMU087	1.230	6	1	19
DMU089	30.000	7	8	20
DMU090	3.400	6	1	23
DMU091	30.000	6	9	24
DMU093	1.000	9	1	14
DMU094	7.561	7	8	22
DMU096	75	2	7	17
DMU097	3.000	7	8	18
DMU098	200	6	5	27
DMU099	1.200	10	3	12
DMU100	54	7	1	15
DMU101	800	6	9	22
DMU102	300	4	1	20
DMU103	1.000	8	8	12
DMU104	400	5	6	16
DMU105	2.500	6	9	22
DMU106	1.200	8	9	19
DMU107	96	3	5	24
DMU108	1.000	2	8	12
DMU109	220	8	7	25
DMU111	1.339	5	8	23
DMU113	120	4	8	18
DMU114	12.000	3	8	23
DMU115	2.500	3	6	28
DMU116	850	10	8	29
DMU117	650	8	8	16
DMU118	800	8	6	22
DMU119	55	4	9	22
DMU122	1.800	7	1	18
DMU123	250	6	9	30
DMU124	2.600	9	9	24
DMU125	200	4	1	20

DMU Outputs 2017	Order Lines	Special Processes	Error Free %	Order Flexibility
DMU126	35.000	10	9	26
DMU127	3.000	8	7	23
DMU128	55.000	8	1	25
DMU129	3.000	3	8	20
DMU130	3.000	10	9	27
DMU131	500	7	8	21

Table 5: DEA output table 2017.

DMU Outputs 2012	Order Lines	Special Processes	Error Free %	Order Flexibility
DMU001	1.100	7	7	14
DMU002	300	2	6	23
DMU003	450	4	7	18
DMU004	800	3	4	13
DMU005	4.000	2	5	19
DMU006	1.500	3	3	23
DMU007	175	2	3	20
DMU008	15.000	6	3	16
DMU010	800	5	5	13
DMU012	350	4	3	18
DMU014	3.200	3	3	24
DMU017	2.598	5	4	19
DMU019	1.560	4	1	24
DMU020	2.200	4	6	18
DMU022	100	7	1	17
DMU023	13.500	8	8	18
DMU024	300	3	8	19
DMU026	240	5	6	21
DMU027	150	10	5	25
DMU028	272	5	8	21
DMU030	6.500	4	4	19
DMU033	1.900	5	8	24
DMU034	7.250	8	4	26
DMU037	800	3	6	16
DMU040	2.000	6	5	11
DMU041	240	7	4	22
DMU042	250	3	4	10
DMU043	25	6	4	10
DMU044	700	3	1	10
DMU045	325	2	2	10
DMU046	3.000	7	1	20
DMU047	3.200	6	7	17
DMU049	11.421	1	1	13
DMU050	550	3	7	20
DMU051	2.000	5	2	18
DMU052	35.000	2	3	17
DMU053	422	2	4	14

DMU Outputs 2012	Order Lines	Special Processes	Error Free %	Order Flexibility
DMU054	110	8	8	10
DMU055	6.000	2	6	19
DMU056	1.250	6	7	10
DMU058	85	6	6	30
DMU059	50	7	3	30
DMU060	23.860	5	4	22
DMU061	2.650	5	7	18
DMU062	9.000	9	8	10
DMU063	9.000	7	3	13
DMU064	200	4	6	21
DMU065	14.000	6	1	17
DMU066	250	8	5	18
DMU067	1.500	8	6	26
DMU068	950	4	9	18
DMU069	270	6	4	14
DMU070	5.000	3	8	21
DMU071	350	3	7	10
DMU072	1.750	2	2	20
DMU073	2.500	7	3	18
DMU074	300	5	9	10
DMU076	400	1	3	18
DMU077	1.300	2	5	21
DMU080	1.355	5	3	18
DMU081	330	6	4	10
DMU082	250	6	1	10
DMU083	150	2	1	22
DMU086	300	4	3	18
DMU087	880	6	1	21
DMU089	29.000	7	7	21
DMU090	3.200	4	1	15
DMU091	30.000	6	9	24
DMU093	1.100	4	1	13
DMU094	7.200	6	8	24
DMU096	100	2	4	17
DMU097	2.500	5	4	18
DMU098	180	4	3	27
DMU099	1.300	10	2	10
DMU100	35	7	1	10
DMU101	550	2	1	15
DMU102	140	3	1	10
DMU103	950	6	8	10
DMU104	300	4	5	17
DMU105	500	2	1	19
DMU106	1.100	7	7	13
DMU107	70	2	4	20
DMU108	1.200	2	8	10
DMU109	171	9	6	17
DMU111	1.184	5	8	23

DMU Outputs 2012	Order Lines	Special Processes	Error Free %	Order Flexibility
DMU113	160	4	3	18
DMU114	18.000	3	6	14
DMU115	500	4	3	12
DMU116	400	8	8	28
DMU117	300	3	1	15
DMU118	260	8	5	20
DMU119	50	4	8	18
DMU122	1.600	6	1	18
DMU123	50	4	1	30
DMU124	1.800	7	8	21
DMU125	200	4	1	21
DMU126	20.000	10	9	24
DMU127	2.000	8	4	18
DMU128	37.500	7	1	19
DMU129	2.500	3	7	18
DMU130	200	4	6	27
DMU131	250	6	1	17

Table 6: DEA output table 2012.

F Input and output correlation 2012

Automation Method Kendall's τ 12	**Method 1**	**Method 2**	**Method 3**
Method 1	1***		
Method 2	0,63***	1***	
Method 3	0,94***	0,61***	1***

Table 7: Cross-efficiency ranking correlation for different automation score calculations 2012.

Kendall's τ 2012	**FTEs**	**Floor Space**	**SKUs**	**Automation**
FTEs	1			
Floor Space	0,46***	1		
SKUs	0,25***	0,13*	1	
Automation	0,2***	0,06	0,12*	1

Table 8: Input factors' Kendall's τ correlation coefficients 2012.

Kendall's τ 2012	**Order Lines**	**Special Processes**	**Error Free %**	**Order Flexibility**
Order Lines	1			
Special Processes	0,07	1		
Error Free %	0,1	0,14*	1	
Order Flexibility	0,03	0,04	0,09	1

Table 9: Output factors' Kendall's τ correlation coefficients 2012.

Kendall's τ 2012	Order Lines	Special Processes	Error Free %	Order Flexibility
FTEs	0,42***	0,05	0,04	-0,03
Floor Space	0,24***	0,09	0,01	0,04
SKUs	0,28***	0,1	0,14*	-0,01
Automation	0,22***	0,14*	0,13*	0,18**

Table 10: Input and output factors' Kendall's τ correlation coefficients 2012.

G CRS and VRS efficiency score distributions

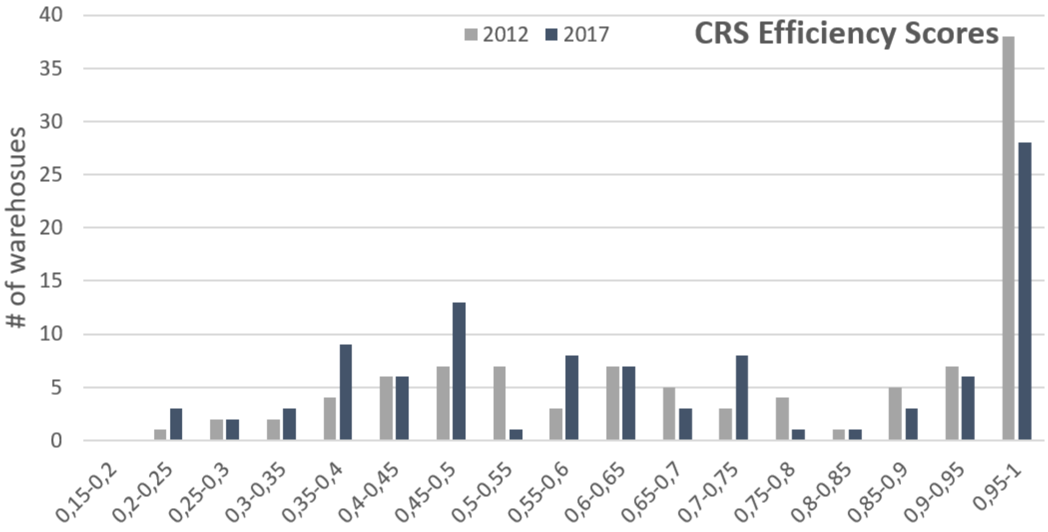


Figure 6: DEA efficiency score distribution under CRS assumptions.

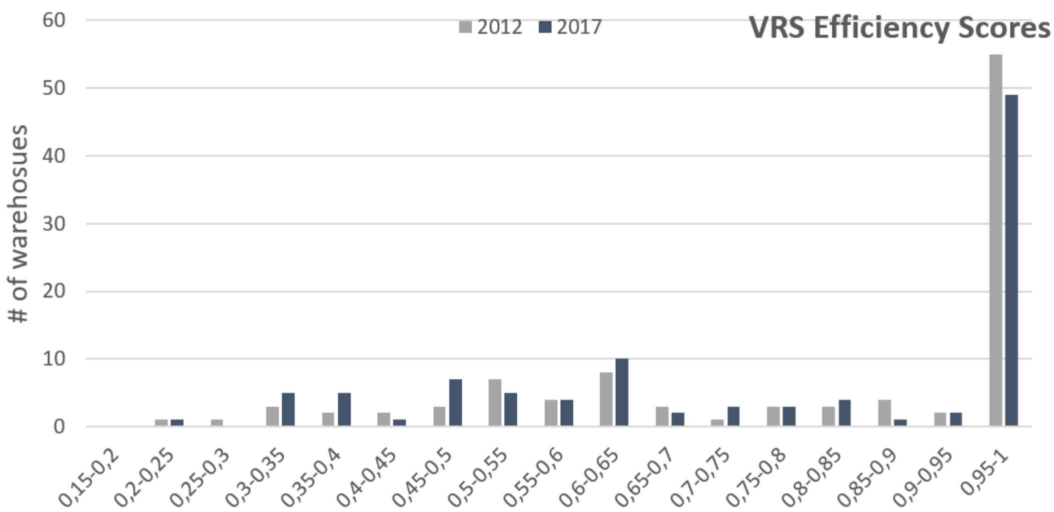


Figure 7: DEA efficiency score distribution under VRS assumptions.

H Efficient DMUs and peer selections

DMU Number	**CRS 2012**	**Peer Selection**
DMU006	1,00	2
DMU008	1,00	6
DMU024	1,00	6
DMU027	1,00	28
DMU028	1,00	34
DMU033	1,00	8
DMU034	1,00	4
DMU045	1,00	1
DMU046	1,00	6
DMU049	1,00	25
DMU050	1,00	20
DMU052	1,00	5
DMU056	1,00	1
DMU061	1,00	6
DMU063	1,00	14
DMU067	1,00	21
DMU071	1,00	4
DMU076	1,00	2
DMU082	1,00	11
DMU089	1,00	4
DMU091	1,00	1
DMU098	1,00	3
DMU100	1,00	13
DMU104	1,00	8
DMU106	1,00	18
DMU107	1,00	16
DMU108	1,00	3
DMU117	1,00	27
DMU118	1,00	5
DMU123	1,00	6
DMU125	1,00	1
DMU127	1,00	3
DMU128	1,00	17
DMU130	1,00	2
DMU131	1,00	3

Table 11: Efficient DMUs and number of selections as peers - CRS 2012.

DMU Number	**VRS 2012**	**Peer Selection**
DMU006	1,00	7
DMU008	1,00	6
DMU023	1,00	3
DMU024	1,00	7
DMU027	1,00	19
DMU028	1,00	25
DMU033	1,00	7
DMU034	1,00	4

DMU Number	VRS 2012	Peer Selection
DMU045	1,00	13
DMU046	1,00	5
DMU049	1,00	18
DMU050	1,00	15
DMU052	1,00	4
DMU054	1,00	5
DMU056	1,00	1
DMU058	1,00	1
DMU059	1,00	4
DMU061	1,00	4
DMU062	1,00	2
DMU063	1,00	10
DMU065	1,00	1
DMU067	1,00	14
DMU068	1,00	2
DMU071	1,00	4
DMU074	1,00	1
DMU076	1,00	4
DMU082	1,00	4
DMU089	1,00	4
DMU091	1,00	3
DMU094	1,00	2
DMU098	1,00	7
DMU099	1,00	1
DMU100	1,00	9
DMU101	1,00	3
DMU104	1,00	5
DMU106	1,00	10
DMU107	1,00	12
DMU108	1,00	5
DMU109	1,00	1
DMU111	1,00	4
DMU114	1,00	1
DMU116	1,00	3
DMU117	1,00	21
DMU118	1,00	2
DMU123	1,00	5
DMU125	1,00	2
DMU126	1,00	2
DMU127	1,00	3
DMU128	1,00	15
DMU130	1,00	2
DMU131	1,00	5

Table 12: Efficient DMUs and number of selections as peers - VRS 2012.

I Scale inefficiencies

DMU Efficiencies 2017	**CRS**	**VRS**	**Scale Efficiency**
DMU001	0,38	0,47	0,82
DMU002	0,55	0,65	0,85
DMU003	0,31	0,37	0,84
DMU004	0,41	0,68	0,60
DMU005	0,95	1,00	0,95
DMU006	1,00	1,00	1,00
DMU007	0,24	0,61	0,40
DMU008	1,00	1,00	1,00
DMU010	0,32	0,33	0,97
DMU012	0,48	0,51	0,94
DMU014	0,29	0,31	0,93
DMU017	0,35	0,36	0,98
DMU019	0,46	0,47	0,98
DMU020	0,35	0,35	1,00
DMU022	0,46	0,48	0,96
DMU023	0,71	1,00	0,71
DMU024	1,00	1,00	1,00
DMU026	0,60	0,61	0,99
DMU027	1,00	1,00	1,00
DMU028	1,00	1,00	1,00
DMU030	0,74	0,80	0,92
DMU033	0,39	1,00	0,39
DMU034	0,94	1,00	0,94
DMU037	0,72	0,72	1,00
DMU040	1,00	1,00	1,00
DMU041	1,00	1,00	1,00
DMU042	0,50	0,67	0,75
DMU043	0,86	0,88	0,99
DMU044	0,64	0,64	0,99
DMU045	1,00	1,00	1,00
DMU046	0,92	0,93	0,98
DMU047	0,74	0,74	0,99
DMU049	1,00	1,00	1,00
DMU050	1,00	1,00	1,00
DMU051	0,62	0,65	0,95
DMU052	1,00	1,00	1,00
DMU053	0,64	0,64	1,00
DMU054	0,71	0,79	0,90
DMU055	0,48	1,00	0,48
DMU056	0,36	0,48	0,75
DMU058	0,57	1,00	0,57
DMU059	0,68	1,00	0,68
DMU060	0,50	0,57	0,88
DMU061	0,80	1,00	0,80
DMU062	0,55	1,00	0,55
DMU063	1,00	1,00	1,00

DMU Efficiencies 2017	CRS	VRS	Scale Efficiency
DMU064	1,00	1,00	1,00
DMU065	0,49	0,54	0,91
DMU066	1,00	1,00	1,00
DMU067	1,00	1,00	1,00
DMU068	0,44	0,60	0,73
DMU069	0,33	0,33	0,98
DMU070	0,46	0,84	0,54
DMU071	1,00	1,00	1,00
DMU072	0,23	0,25	0,95
DMU073	0,69	0,77	0,90
DMU074	0,42	0,43	0,98
DMU076	0,28	0,30	0,90
DMU077	0,56	0,65	0,86
DMU080	0,43	0,51	0,84
DMU081	0,55	0,57	0,97
DMU082	0,63	0,64	0,98
DMU083	0,57	0,57	1,00
DMU086	0,73	0,81	0,91
DMU087	0,48	0,51	0,95
DMU089	0,91	1,00	0,91
DMU090	0,94	0,94	0,99
DMU091	0,90	1,00	0,90
DMU093	0,46	0,50	0,93
DMU094	0,58	0,78	0,75
DMU096	0,45	0,50	0,90
DMU097	0,70	0,83	0,84
DMU098	1,00	1,00	1,00
DMU099	1,00	1,00	1,00
DMU100	1,00	1,00	1,00
DMU101	0,94	1,00	0,94
DMU102	0,25	0,33	0,75
DMU103	0,61	0,74	0,83
DMU104	1,00	1,00	1,00
DMU105	0,57	1,00	0,57
DMU106	1,00	1,00	1,00
DMU107	1,00	1,00	1,00
DMU108	1,00	1,00	1,00
DMU109	0,38	0,62	0,61
DMU111	0,84	1,00	0,84
DMU113	0,37	0,40	0,93
DMU114	0,90	0,97	0,93
DMU115	1,00	1,00	1,00
DMU116	0,38	1,00	0,38
DMU117	0,70	1,00	0,70
DMU118	1,00	1,00	1,00
DMU119	0,61	0,62	0,98
DMU122	0,94	1,00	0,94
DMU123	0,75	1,00	0,75
DMU124	0,48	1,00	0,48

DMU Efficiencies 2017	**CRS**	**VRS**	**Scale Efficiency**
DMU125	1,00	1,00	1,00
DMU126	0,50	1,00	0,50
DMU127	0,43	0,48	0,90
DMU128	1,00	1,00	1,00
DMU129	0,40	0,40	1,00
DMU130	0,48	1,00	0,48
DMU131	0,49	0,54	0,91

Table 13: Scale Efficiency of DMUs 2017.

DMU Efficiencies 2012	**CRS**	**VRS**	**Scale Efficiency**
DMU001	0,50	0,58	0,86
DMU002	0,61	0,65	0,94
DMU003	0,58	0,62	0,94
DMU004	0,50	0,61	0,83
DMU005	0,87	0,88	0,99
DMU006	1,00	1,00	1,00
DMU007	0,23	0,24	0,97
DMU008	1,00	1,00	1,00
DMU010	0,52	0,55	0,95
DMU012	0,41	0,41	0,99
DMU014	0,51	0,53	0,96
DMU017	0,35	0,36	0,98
DMU019	0,56	0,57	0,97
DMU020	0,47	0,50	0,95
DMU022	0,47	0,51	0,92
DMU023	0,87	1,00	0,87
DMU024	1,00	1,00	1,00
DMU026	0,62	0,65	0,96
DMU027	1,00	1,00	1,00
DMU028	1,00	1,00	1,00
DMU030	0,85	0,87	0,98
DMU033	1,00	1,00	1,00
DMU034	1,00	1,00	1,00
DMU037	0,65	0,65	0,99
DMU040	0,91	0,91	0,99
DMU041	0,75	0,84	0,90
DMU042	0,44	0,49	0,89
DMU043	0,91	0,95	0,95
DMU044	0,53	0,89	0,60
DMU045	1,00	1,00	1,00
DMU046	1,00	1,00	1,00
DMU047	0,77	0,83	0,93
DMU049	1,00	1,00	1,00
DMU050	1,00	1,00	1,00
DMU051	0,51	0,52	0,99
DMU052	1,00	1,00	1,00
DMU053	0,88	0,98	0,90

DMU Efficiencies 2012	CRS	VRS	Scale Efficiency
DMU054	0,71	1,00	0,71
DMU055	0,83	0,91	0,92
DMU056	1,00	1,00	1,00
DMU058	0,67	1,00	0,67
DMU059	0,98	1,00	0,98
DMU060	0,48	0,54	0,88
DMU061	1,00	1,00	1,00
DMU062	0,92	1,00	0,92
DMU063	1,00	1,00	1,00
DMU064	0,64	0,65	0,98
DMU065	0,97	1,00	0,97
DMU066	0,93	0,99	0,95
DMU067	1,00	1,00	1,00
DMU068	0,55	1,00	0,55
DMU069	0,61	0,61	1,00
DMU070	0,43	0,79	0,54
DMU071	1,00	1,00	1,00
DMU072	0,40	0,40	1,00
DMU073	0,63	0,65	0,97
DMU074	0,42	1,00	0,42
DMU076	1,00	1,00	1,00
DMU077	0,65	0,67	0,98
DMU080	0,48	0,49	0,98
DMU081	0,78	0,80	0,97
DMU082	1,00	1,00	1,00
DMU083	0,85	0,87	0,98
DMU086	0,49	0,55	0,90
DMU087	0,57	0,63	0,90
DMU089	1,00	1,00	1,00
DMU090	0,69	0,78	0,89
DMU091	1,00	1,00	1,00
DMU093	0,26	0,26	0,99
DMU094	0,95	1,00	0,95
DMU096	0,38	0,40	0,94
DMU097	0,34	0,35	0,99
DMU098	1,00	1,00	1,00
DMU099	0,76	1,00	0,76
DMU100	1,00	1,00	1,00
DMU101	0,90	1,00	0,90
DMU102	0,26	0,33	0,79
DMU103	0,35	0,60	0,59
DMU104	1,00	1,00	1,00
DMU105	0,47	0,48	0,99
DMU106	1,00	1,00	1,00
DMU107	1,00	1,00	1,00
DMU108	1,00	1,00	1,00
DMU109	0,61	1,00	0,61
DMU111	0,92	1,00	0,92
DMU113	0,32	0,33	0,95

DMU Efficiencies 2012	CRS	VRS	Scale Efficiency
DMU114	0,93	1,00	0,93
DMU115	0,43	0,55	0,78
DMU116	0,70	1,00	0,70
DMU117	1,00	1,00	1,00
DMU118	1,00	1,00	1,00
DMU119	0,71	0,81	0,87
DMU122	0,69	0,71	0,97
DMU123	1,00	1,00	1,00
DMU124	0,52	0,99	0,53
DMU125	1,00	1,00	1,00
DMU126	0,66	1,00	0,66
DMU127	1,00	1,00	1,00
DMU128	1,00	1,00	1,00
DMU129	0,45	0,62	0,73
DMU130	1,00	1,00	1,00
DMU131	1,00	1,00	1,00

Table 14: Scale Efficiency of DMUs 2012.

J Cross-efficiency comparison

Cross-efficiency 2012	Sexton_Classic	Sexton_Ratio	Multiplicative	Game Theory
Minimum Score	0,113	0,113	0,000	0,197
Average Score	0,379	0,383	0,028	0,614
Maximum Score	0,887	0,881	0,927	1,000
Std. Dev. Score	0,166	0,172	0,107	0,227

Table 15: Results comparison of cross-efficiency methods 2012.

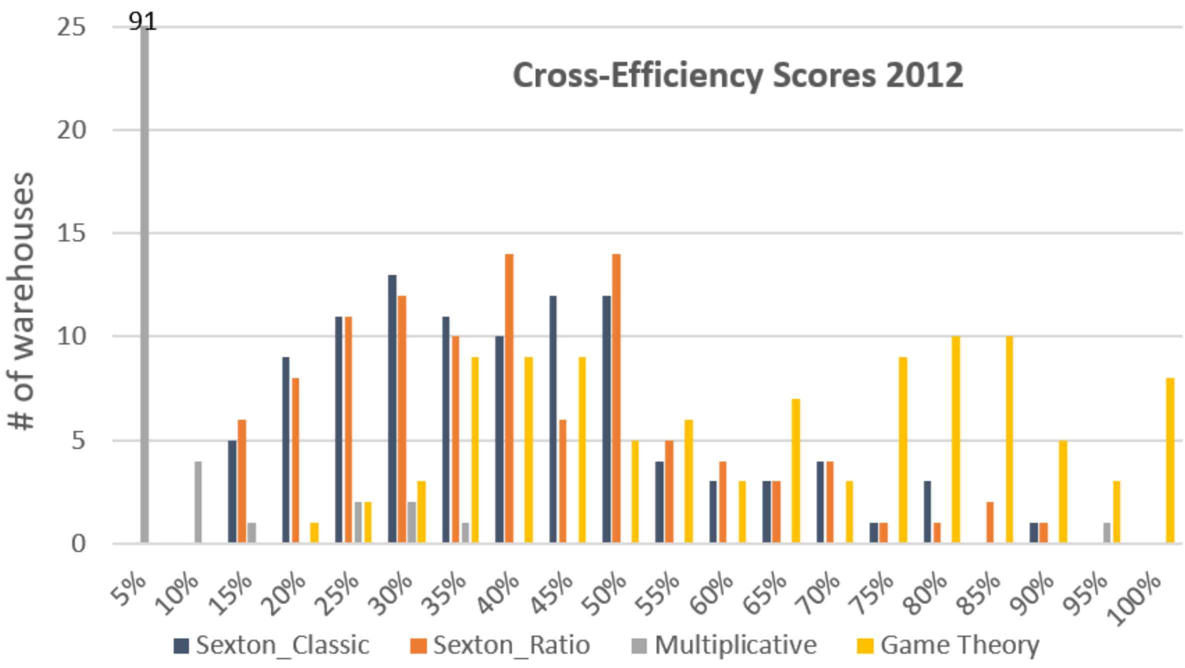


Figure 8: Cross-efficiency score distribution 2012 per method.

WSR Results 2017	**Sexton_Classic**	**Sexton_Ratio**	**Multiplicative**	**Game Theory**
Sexton_Classic	0			
Sexton_Ratio	1	0		
Multiplicative	1	1	0	
Game Theory	1	1	1	0

Table 16: Wilcoxon rank test results for cross-efficiency method comparison 2017.

Cross-efficiency Ranking 2017	**Sexton_Classic**	**Sexton_Ratio**	**Multiplicative**	**Game Theory**
Top 1	**DMU098**	**DMU098**	**DMU067**	DMU049
Top 2	**DMU050**	**DMU050**	**DMU050**	**DMU098**
Top 3	**DMU028**	**DMU028**	**DMU028**	**DMU050**
Top 4	DMU049	DMU049	**DMU027**	**DMU028**
Top 5	**DMU104**	**DMU104**	**DMU098**	**DMU104**
Top 6	**DMU066**	**DMU066**	**DMU066**	**DMU066**
Top 7	**DMU027**	**DMU027**	**DMU104**	**DMU027**
Top 8	**DMU067**	**DMU067**	DMU006	**DMU067**
Top 9	DMU041	DMU041	DMU115	DMU107
Top 10	DMU107	DMU040	DMU040	DMU041
Top 11	DMU040	DMU107	DMU049	DMU128
Top 12	DMU043	DMU043	DMU071	DMU040
Top 13	DMU071	DMU071	DMU106	DMU099
Top 14	DMU045	DMU128	DMU108	DMU100
Top 15	DMU128	DMU024	DMU047	DMU071
Top 16	DMU106	DMU045	DMU041	DMU106
Top 17	DMU024	DMU106	DMU107	DMU045
Top 18	DMU100	DMU047	DMU114	DMU043

Cross-efficiency Ranking 2017	Sexton_Classic	Sexton_Ratio	Multiplicative	Game Theory
Top 19	DMU099	DMU100	DMU045	DMU108
Top 20	DMU047	DMU099	DMU053	DMU024
Top 21	DMU086	DMU108	DMU105	DMU006
Top 22	DMU108	DMU064	DMU086	DMU064
Top 23	DMU064	DMU006	DMU034	DMU114
Top 24	DMU117	DMU117	DMU061	DMU115
Top 25	DMU063	DMU086	DMU044	DMU063
Top 26	DMU006	DMU114	DMU024	DMU046
Top 27	DMU114	DMU123	DMU118	DMU122
Top 28	DMU123	DMU063	DMU099	DMU047
Top 29	DMU115	DMU053	DMU117	DMU086
Top 30	DMU053	DMU115	DMU111	DMU034
Top 31	DMU105	DMU105	DMU064	DMU118
Top 32	DMU058	DMU023	DMU097	DMU125
Top 33	DMU023	DMU122	DMU023	DMU117
Top 34	DMU054	DMU034	DMU062	DMU005
Top 35	DMU119	DMU058	DMU101	DMU008
Top 36	DMU044	DMU054	DMU123	DMU052
Top 37	DMU034	DMU119	DMU068	DMU123
Top 38	DMU122	DMU044	DMU030	DMU054
Top 39	DMU012	DMU061	DMU094	DMU023
Top 40	DMU026	DMU026	DMU026	DMU053
Top 41	DMU061	DMU062	DMU063	DMU061
Top 42	DMU118	DMU118	DMU005	DMU044
Top 43	DMU062	DMU012	DMU082	DMU105
Top 44	DMU083	DMU008	DMU008	DMU091
Top 45	DMU046	DMU083	DMU012	DMU090
Top 46	DMU082	DMU046	DMU043	DMU089
Top 47	DMU103	DMU005	DMU073	DMU119
Top 48	DMU008	DMU073	DMU122	DMU097
Top 49	DMU052	DMU082	DMU002	DMU030
Top 50	DMU131	DMU131	DMU103	DMU062
Top 51	DMU097	DMU103	DMU080	DMU058
Top 52	DMU059	DMU030	DMU077	DMU082
Top 53	DMU030	DMU059	DMU056	DMU073
Top 54	DMU073	DMU097	DMU083	DMU111
Top 55	DMU005	DMU081	DMU125	DMU103
Top 56	DMU081	DMU125	DMU037	DMU026
Top 57	DMU125	DMU065	DMU055	DMU101
Top 58	DMU065	DMU052	DMU130	DMU059
Top 59	DMU068	DMU111	DMU033	DMU037
Top 60	DMU111	DMU068	DMU020	DMU083
Top 61	DMU077	DMU116	DMU001	DMU012
Top 62	DMU042	DMU077	DMU065	DMU081
Top 63	DMU101	DMU101	DMU051	DMU051
Top 64	DMU116	DMU037	DMU116	DMU002
Top 65	DMU090	DMU056	DMU081	DMU077
Top 66	DMU056	DMU042	DMU091	DMU065

Cross-efficiency Ranking 2017	Sexton_Classic	Sexton_Ratio	Multiplicative	Game Theory
Top 67	DMU080	DMU002	DMU131	DMU094
Top 68	DMU002	DMU001	DMU124	DMU131
Top 69	DMU001	DMU080	DMU070	DMU042
Top 70	DMU094	DMU094	DMU119	DMU130
Top 71	DMU037	DMU090	DMU127	DMU126
Top 72	DMU091	DMU127	DMU058	DMU068
Top 73	DMU051	DMU051	DMU089	DMU124
Top 74	DMU089	DMU091	DMU042	DMU080
Top 75	DMU130	DMU130	DMU054	DMU055
Top 76	DMU127	DMU089	DMU129	DMU127
Top 77	DMU124	DMU069	DMU052	DMU001
Top 78	DMU055	DMU055	DMU046	DMU116
Top 79	DMU069	DMU124	DMU100	DMU056
Top 80	DMU126	DMU126	DMU004	DMU070
Top 81	DMU070	DMU070	DMU059	DMU093
Top 82	DMU020	DMU020	DMU076	DMU019
Top 83	DMU113	DMU113	DMU128	DMU033
Top 84	DMU033	DMU033	DMU003	DMU113
Top 85	DMU109	DMU074	DMU010	DMU060
Top 86	DMU074	DMU109	DMU090	DMU022
Top 87	DMU129	DMU076	DMU069	DMU129
Top 88	DMU076	DMU093	DMU074	DMU069
Top 89	DMU022	DMU022	DMU126	DMU020
Top 90	DMU004	DMU129	DMU113	DMU087
Top 91	DMU093	DMU004	DMU014	DMU074
Top 92	DMU019	DMU003	DMU019	DMU109
Top 93	DMU060	DMU019	DMU109	DMU004
Top 94	DMU003	DMU060	DMU072	DMU096
Top 95	DMU017	DMU010	DMU017	DMU017
Top 96	DMU010	DMU096	DMU096	DMU076
Top 97	DMU096	DMU087	DMU022	DMU010
Top 98	DMU014	DMU017	DMU060	DMU003
Top 99	DMU087	DMU014	DMU087	DMU014
Top 100	DMU007	DMU007	DMU007	DMU007
Top 101	DMU072	DMU072	DMU093	DMU072
Top 102	DMU102	DMU102	DMU102	DMU102

Table 17: Cross-efficiency ranking for all four methods 2017 (DMUs in bold, when in Top10 for all four methods).

Cross-efficiency Ranking 2012	Sexton_Classic	Sexton_Ratio	Multiplicative	Game Theory
Top 1	**DMU028**	**DMU028**	**DMU067**	**DMU028**
Top 2	**DMU067**	**DMU067**	**DMU050**	DMU128
Top 3	**DMU050**	**DMU027**	DMU106	**DMU027**
Top 4	**DMU027**	**DMU050**	DMU006	DMU049
Top 5	**DMU104**	**DMU104**	**DMU027**	**DMU067**

Cross-efficiency Ranking 2012	Sexton_Classic	Sexton_Ratio	Multiplicative	Game Theory
Top 6	DMU049	DMU049	**DMU028**	**DMU050**
Top 7	DMU106	DMU117	DMU033	DMU100
Top 8	DMU128	DMU128	DMU045	DMU117
Top 9	DMU117	DMU107	**DMU104**	**DMU104**
Top 10	DMU107	DMU066	DMU061	DMU107
Top 11	DMU066	DMU106	DMU117	DMU076
Top 12	DMU100	DMU100	DMU076	DMU131
Top 13	DMU118	DMU098	DMU040	DMU106
Top 14	DMU098	DMU045	DMU008	DMU098
Top 15	DMU045	DMU131	DMU127	DMU066
Top 16	DMU131	DMU061	DMU108	DMU006
Top 17	DMU033	DMU118	DMU024	DMU118
Top 18	DMU076	DMU076	DMU056	DMU082
Top 19	DMU061	DMU006	DMU047	DMU123
Top 20	DMU023	DMU023	DMU066	DMU033
Top 21	DMU040	DMU033	DMU053	DMU130
Top 22	DMU006	DMU040	DMU023	DMU108
Top 23	DMU114	DMU041	DMU118	DMU045
Top 24	DMU082	DMU108	DMU030	DMU061
Top 25	DMU059	DMU114	DMU082	DMU059
Top 26	DMU108	DMU008	DMU071	DMU024
Top 27	DMU053	DMU053	DMU041	DMU008
Top 28	DMU030	DMU123	DMU107	DMU034
Top 29	DMU041	DMU130	DMU111	DMU114
Top 30	DMU130	DMU024	DMU034	DMU089
Top 31	DMU034	DMU034	DMU130	DMU071
Top 32	DMU008	DMU043	DMU005	DMU023
Top 33	DMU043	DMU059	DMU046	DMU063
Top 34	DMU047	DMU047	DMU098	DMU053
Top 35	DMU058	DMU082	DMU063	DMU062
Top 36	DMU123	DMU030	DMU131	DMU065
Top 37	DMU024	DMU058	DMU081	DMU040
Top 38	DMU065	DMU065	DMU114	DMU091
Top 39	DMU071	DMU071	DMU049	DMU046
Top 40	DMU063	DMU062	DMU077	DMU052
Top 41	DMU119	DMU119	DMU062	DMU043
Top 42	DMU052	DMU063	DMU020	DMU030
Top 43	DMU089	DMU026	DMU094	DMU127
Top 44	DMU062	DMU083	DMU125	DMU125
Top 45	DMU081	DMU111	DMU089	DMU056
Top 46	DMU111	DMU054	DMU055	DMU083
Top 47	DMU054	DMU125	DMU080	DMU047
Top 48	DMU091	DMU081	DMU124	DMU041
Top 49	DMU026	DMU077	DMU068	DMU005
Top 50	DMU077	DMU005	DMU100	DMU111
Top 51	DMU125	DMU089	DMU122	DMU094
Top 52	DMU083	DMU052	DMU101	DMU081
Top 53	DMU056	DMU091	DMU091	DMU101

Cross-efficiency Ranking 2012	Sexton_Classic	Sexton_Ratio	Multiplicative	Game Theory
Top 54	DMU094	DMU127	DMU037	DMU054
Top 55	DMU127	DMU056	DMU026	DMU058
Top 56	DMU005	DMU116	DMU044	DMU119
Top 57	DMU116	DMU094	DMU126	DMU055
Top 58	DMU069	DMU069	DMU069	DMU077
Top 59	DMU080	DMU080	DMU052	DMU099
Top 60	DMU020	DMU064	DMU128	DMU122
Top 61	DMU064	DMU002	DMU073	DMU026
Top 62	DMU002	DMU122	DMU116	DMU069
Top 63	DMU101	DMU101	DMU043	DMU116
Top 64	DMU068	DMU020	DMU058	DMU126
Top 65	DMU122	DMU068	DMU064	DMU064
Top 66	DMU126	DMU044	DMU129	DMU002
Top 67	DMU109	DMU003	DMU003	DMU037
Top 68	DMU044	DMU037	DMU001	DMU109
Top 69	DMU055	DMU126	DMU004	DMU080
Top 70	DMU003	DMU099	DMU054	DMU044
Top 71	DMU099	DMU086	DMU099	DMU020
Top 72	DMU037	DMU109	DMU083	DMU068
Top 73	DMU086	DMU046	DMU002	DMU073
Top 74	DMU124	DMU042	DMU065	DMU003
Top 75	DMU042	DMU001	DMU010	DMU124
Top 76	DMU046	DMU055	DMU042	DMU001
Top 77	DMU001	DMU124	DMU059	DMU086
Top 78	DMU012	DMU004	DMU123	DMU019
Top 79	DMU004	DMU012	DMU119	DMU105
Top 80	DMU051	DMU010	DMU051	DMU051
Top 81	DMU129	DMU105	DMU086	DMU090
Top 82	DMU105	DMU051	DMU097	DMU060
Top 83	DMU103	DMU073	DMU103	DMU042
Top 84	DMU010	DMU129	DMU115	DMU010
Top 85	DMU073	DMU103	DMU109	DMU004
Top 86	DMU097	DMU097	DMU012	DMU087
Top 87	DMU019	DMU072	DMU014	DMU129
Top 88	DMU060	DMU019	DMU072	DMU012
Top 89	DMU115	DMU060	DMU070	DMU014
Top 90	DMU072	DMU115	DMU105	DMU115
Top 91	DMU074	DMU074	DMU017	DMU072
Top 92	DMU113	DMU087	DMU074	DMU103
Top 93	DMU014	DMU113	DMU087	DMU074
Top 94	DMU070	DMU014	DMU019	DMU097
Top 95	DMU087	DMU096	DMU090	DMU022
Top 96	DMU090	DMU070	DMU060	DMU070
Top 97	DMU096	DMU090	DMU113	DMU096
Top 98	DMU017	DMU093	DMU093	DMU113
Top 99	DMU093	DMU017	DMU096	DMU017
Top 100	DMU102	DMU102	DMU102	DMU102
Top 101	DMU022	DMU007	DMU007	DMU093

Cross-efficiency Ranking 2012	Sexton_Classic	Sexton_Ratio	Multiplicative	Game Theory
Top 102	DMU007	DMU022	DMU022	DMU007

Table 18: Cross-efficiency ranking for all four methods 2012 (DMUs in bold, when in Top10 for all four methods).

Kendall's τ 2012	**Sexton_Classic**	**Sexton_Ratio**	**Multiplicative**	**Game Theory**
Sexton_Classic	1***			
Sexton_Ratio	0,94***	1***		
Multiplicative	0,66***	0,66***	1***	
Game Theory	0,83***	0,82***	0,65***	1***

Table 19: Kendall's τ correlations for 2012 cross-efficiency rankings.

Kendall's τ 2012	**DEA**	**Super-efficiency**
Sexton_Classic	0,645***	0,652***
Sexton_Ratio	0,636***	0,644***
Multiplicative	0,591***	0,579***
Game Theory	0,776***	0,791***
Sexton_EfficientOnly	0,64***	0,653***
DEA	1***	0,925***
Superefficiency	0,925***	1***

Table 20: Kendall's τ correlation between cross-efficiency, DEA and super-efficiency scores 2012.

K Cluster analysis

Industry Differences 2017	**Sum Squared Differences**	**Degrees of Freedom**	**Mean Squared Differences**	**Chi-squared**	**Probability ≥ Chi-squared**
Groups	1.803	2	902	6,32	0,0424
Error	14.451	55	263		
Total	16.255	57			

Table 21: Kruskal-Wallis test for different industries 2017.

Industry Differences 2012	**Sum Squared Differences**	**Degrees of Freedom**	**Mean Squared Differences**	**Chi-squared**	**Probability ≥ Chi-squared**
Groups	1.954	2	977	6,85	0,0325
Error	14.301	55	260		
Total	16.255	57			

Table 22: Kruskal-Wallis test for different industries 2012.

Cluster Ranks 2012	**Overall Sample**	**Construction+ Engineering**	**Consumer Goods**	**Food/ Groceries**
Observations	102	20	19	19
Minimum Rank	1	3	1	6
Maximum Rank	102	102	99	101
Average Rank	51,50	66,10	44,58	45,84
Std. Dev. Rank	29,44	27,90	28,67	26,61
WSR vs. Remainder p-value		0,01	0,26	0,36

Table 23: Cluster metrics and Wilcoxon rank test results vs. remaining panelists 2012.

2012 Value Chain's Cross-Efficiency	**Overall Sample**	**Production**	**Wholesale**	**Retail**
Observations	102	27	61	14
Minimum Score	0,11	0,12	0,11	0,29
Maximum Score	0,88	0,80	0,88	0,69
Average Score	0,38	0,36	0,37	0,46
Std. Dev. Score	0,17	0,17	0,18	0,12

Table 24: Kruskal-Wallis test for different value chain positions' effect on cross-efficiency scores 2012.

2012 Ownership's Cross-Efficiency	**Overall Sample**	**In-House**	**3PL-Dedicated**	**3PL-Multiple**
Observations	102	58	10	34
Minimum Score	0,11	0,11	0,15	0,12
Maximum Score	0,88	0,88	0,68	0,68
Average Score	0,38	0,40	0,36	0,36
Std. Dev. Score	0,17	0,20	0,16	0,12
KW p-value			0,7780	

Table 25: Kruskal-Wallis test for different ownership types' effect on cross-efficiency scores 2012.

L Cross-efficiency and input regression coefficients

MLR Coefficients 2012	**All DMUs**	**Construction+ Engineering**	**Consumer Goods**	**Food/ Groceries**
Cold-Storage	0,09	-0,25	0,03	0,12
FTEs	-0,49***	-0,81	-0,58	-0,37*
Floor Space	-0,32***	0,22	0,03	-0,4*
SKUs	-0,34***	-0,35	-0,55**	-0,09
Automation	-0,44***	-0,04	-0,7***	-0,65***
Order Lines	0,7***	0,58	0,97**	0,85***
Special Processes	0,24***	0,71***	0,01	0,26*
Error Free %	0,06	0,38	0,09	0,12
Order Flexibility	0,23***	0,35	0,59**	0,32**
Adjusted R^2	*0,55*	*0,58*	*0,78*	*0,84*
RMSE	*0,59*	*0,65*	*0,47*	*0,41*
F-value	*14,40*	*3,88*	*7,98*	*11,10*

Table 26: Multiple linear regression coefficients of inputs, outputs and cold-storage on cross-efficiency score 2012.

MLR Coefficients Change	**All DMUs**	**Construction+ Engineering**	**Consumer Goods**	**Food/ Groceries**
Cold-Storage	-0,05	-0,75*	0,26	0,34
FTEs	-0,2**	0,07	0,05	-0,43
Floor Space	0,05	0,24	-0,49	0,18
SKUs	-0,54***	-0,31	-0,15	-1,27
Automation	-0,68***	-0,42*	-1,32***	-0,69**
Order Lines	0,59***	0,11	0,33	1,29
Special Processes	0,28***	-0,19	0,33	0,46
Error Free %	0,18**	0,81**	0,61	-0,39
Order Flexibility	0,24***	0,64**	0,12	0,22
Adjusted R^2	*0,52*	*0,75*	*0,62*	*0,36*
RMSE	*0,69*	*0,50*	*0,62*	*0,80*
F-value	*13,10*	*7,22*	*4,30*	*2,13*

Table 27: Multiple linear regression coefficients of change on inputs, outputs and cold-storage on change in cross-efficiency score between 2017 and 2012.

M Cluster-specific cross-efficiency results

Cross-efficiency Scores 2012	Construction+ Engineering	Consumer Goods	Food/ Groceries
Minimum	0,21	0,19	0,25
Maximum	0,95	0,99	0,98
Average	0,50	0,55	0,69
Std. Dev.	0,21	0,22	0,20
Observations	20	19	19
Adjusted R^2	*0,69*	*0,76*	*0,89*
RMSE	*0,56*	*0,49*	*0,34*
F-value	*5,60*	*7,18*	*16,50*

Table 28: Descriptive statistics cross-efficiency score distribution for calculations with individual clusters 2012.

Construction+ Engineering Cluster	2012	2017	Change
Cold-Storage	-0,12	0	-0,79
FTEs	-0,59	-1,33	-0,33
Floor Space	0,35	-0,45	0,34
SKUs	-0,22	-0,22	-0,16
Automation	-0,33	-0,64***	-0,12
Order Lines	0,45	1,82*	0,06
Special Processes	0,55**	0,34	-0,16
Error Free %	0,35	0,48**	1,13**
Order Flexibility	0,48**	0,01	0,31
Adjusted R^2	*0,69*	*0,59*	*0,58*
RMSE	*0,56*	*0,64*	*0,65*
F-value	*5,60*	*4,03*	*3,94*

Table 29: Multiple linear regression coefficients of inputs, outputs and cold-storage on cross-efficiency score - construction+engineering cluster.

Consumer Goods Cluster	2012	2017	Change
Cold-Storage	0,19	0,13	0,13
FTEs	-0,9	-0,48	-0,13
Floor Space	-0,15	-0,44	-0,51*
SKUs	-0,27	-0,03	0,24
Automation	-0,62**	-0,7*	-1,25***
Order Lines	1,48***	1,37*	0,24
Special Processes	0,26	0,31	0,85**
Error Free %	0,21	0,26	0,37
Order Flexibility	0,2	0,17	0,06
Adjusted R^2	*0,76*	*0,47*	*0,75*
RMSE	*0,49*	*0,73*	*0,50*
F-value	*7,18*	*2,76*	*6,99*

Table 30: Multiple linear regression coefficients of inputs, outputs and cold-storage on cross-efficiency score - consumer goods cluster.

Food/Groceries Cluster	2012	2017	Change
Cold-Storage	0,02	0,31***	0,16
FTEs	-0,25	-0,09	0,56
Floor Space	-0,47**	-0,25*	0,2
SKUs	-0,4**	-0,24**	-2,21
Automation	-0,79***	-0,76***	-0,69**
Order Lines	0,51***	0,45***	1,35
Special Processes	0,34**	0,15*	0,14
Error Free %	0,29*	0,26**	-0,17
Order Flexibility	0,3**	0,35***	0,13
Adjusted R^2	*0,89*	*0,92*	*0,38*
RMSE	*0,34*	*0,28*	*0,79*
F-value	*16,50*	*25,40*	*2,24*

Table 31: Multiple linear regression coefficients of inputs, outputs and cold-storage on cross-efficiency score - food/groceries cluster.

N Sensitivity analysis

Scale Changes	0,1x	0,2x	0,5x	1x	2x	5x	10x
Sexton_Classic	0,25	0,22	0,16	0,00	0,12	0,19	0,21
Sexton_Ratio	0,29	0,23	0,12	0,00	0,13	0,20	0,22
Multiplicative	1,26	0,94	0,45	0,00	0,23	0,45	0,63
Game Theory	0,21	0,19	0,14	0,00	0,10	0,14	0,21

Table 32: Average relative score deviations per method, based on re-scaling 10% of inputs by a given factor.

Entry Errors	Original	5%	10%	25%	50%	75%	100%
Sexton_Classic	0,00	0,05	0,07	0,09	0,13	0,13	0,15
Sexton_Ratio	0,00	0,08	0,09	0,11	0,14	0,15	0,14
Multiplicative	0,00	0,12	0,15	0,21	0,26	0,30	0,36
Game Theory	0,00	0,08	0,08	0,11	0,13	0,14	0,17

Table 33: Average relative score deviations per method, re-scaling a different percentage of inputs by 75% - 125%.

Efficient DMU Elimination	Original	1	2	3	4	5	6
Sexton_Classic	0,00	0,10	0,14	0,15	0,20	0,22	0,24
Sexton_Ratio	0,00	0,09	0,12	0,14	0,18	0,20	0,24
Multiplicative	0,00	0,33	0,49	0,50	0,77	0,89	0,92
Game Theory	0,00	0,07	0,10	0,16	0,22	0,22	0,25

Table 34: Average relative score deviations per method, based on elimination of a given number of efficient DMUs.