Wageningen University - Social Sciences

MSc Thesis Business Economics

Efficiency analysis of food retailers in Europe.

February 2023

MSc program: Management, Economics and Consumer Studies

Specialisation: Management in Life Sciences

Chair group: Business Economics

Student: Javier Playan Garai

Registration number: 1113879

Supervisor: Dr.ir. Frederic Ang

Thesis code: BEC80436

Abstract

Since the 1950s, technical efficiency has been used to evaluate the performance of firms across many industries. The European food retailing sector has undergone significant changes in recent years, such as online channels, motivated by optimization, market competition and consumer demands. Additionally, it plays a crucial role within the food supply chains. The problem addressed in this thesis is the lack of current efficiency measures for the largest European retailers. These measurements and the drivers of efficiency have been studied in the past, but not recently or with an European scope. In this research, firm-specific data from 2020 was gathered from the Orbis database and financial reports, and then used to run a two-stage efficiency analysis. Firstly, a DEA model was used to obtain efficiency scores for each food retailer. Afterwards, these scores were regressed on environmental variables. The standard and double bootstrap approaches were both used and compared. When the standard approach was used, the average score (0.82) was similar to previous studies. However, it decreased with the bootstrap approach (0.73) due to the bias term. According to both methods, firm size (measured as being a top 5 food retailer in terms of revenue and number of outlets) is a significant determinant of efficiency. These companies access better deals with suppliers than smaller firms because they have more buyer power. In combination with economies of scale, this results in an efficiency gap that will likely keep increasing. Debt-to-equity ratio and international activity could have a negative correlation with efficiency, but since both approaches differ on the significance of these variables, it cannot be firmly concluded.

Acknowledgements

Firstly, I would like to express my gratitude to my supervisor Dr.Ir. Frederic Ang. His feedback, supervision and effort were essential for this master thesis to be successfully completed. It has been a pleasure to work with him during the past months. Additionally, Wageningen University and Research (particularly the Business Economics group) deserves special recognition for providing me the tools and resources required to conduct this research.

On a more personal note, I am also thankful to my family, Eva and friends for their constant support and motivation. It has been a delightful journey, full of new experiences and unexpected challenges, that I have completed thanks to you too.

Disclaimer

In this research, data was collected from the Orbis database owned by Bureau van Dijk. The student gained access to it through Wageningen University and Research. The information contained in the database is confidential, and therefore the experimental data set is not included in this report. If desired, it is possible to gain access to the database through their website (https://www.bvdinfo.com/en-gb/our-products/data/international/orbis).

Abbreviations

BCC- DEA model developed by Banker, Charnes and Cooper (1984)

B2C- Business to consumer

CCR- DEA model developed by Charnes, Cooper and Rhodes (1978)

CRS- Constant returns to scale

DEA- Data Envelopment Analysis

DMU- Decision Making Unit

DOS- Diseconomies of scale

EOS- Economies of scale

FMCG- Fast Moving Consumer Goods

LOP- Law of one price

OLS- Ordinary Least Squares

PPPs- Purchasing Power Parities

RTS- Returns to scale

TE- Technical efficiency

VRS- Variable returns to scale

Table of contents

1. Introduction

1.1. Problem statement

Identifying how financially efficient companies are is useful for managers and investors, as it could help companies to improve their performance. Financial efficiency is key for firms operating in competitive and dynamic environments, such as food retailers (Konuk, 2018; Perrigot and Barros, 2008). Inefficient firms are not instantaneously driven out from a competitive market, it may even take years, and these firms normally have enough time to become efficient before this happens (Cooper et al., 2011). Corporate Social Responsibility can be defined as the voluntary services of a firm towards society and the environment where it grows and operates (Babatunde and Ahmed, 2015). Singh and Misra (2021) concluded that there is a positive correlation between firm performance and corporate social responsibility by analysing European corporate firms, and Sial et al. (2018) reached the same conclusion analysing Chinese firms. Therefore, by enhancing firm's performance and efficiency could result in more involvement of the corporate firms with the society and environment.

Additionally, the food retailing sector plays a significant in society, since it connects food processors and consumers (Hirsch et al., 2021). The sector is formed by firms of different size that operate in a competitive environment, with price competition and low margins (Hamilton et al., 2020; Hirsch et al., 2021). In fact, the sector is dominated worldwide by few retailers, which is leading to even higher levels of market concentration (Hamilton et al., 2020). Therefore, improving the performance and efficiency of these few firms would have a relevant impact worldwide, especially because on average the largest retailers operate in eleven different countries (Deloitte, 2022). Increasing the efficiency of the firm benefits the firm itself and also other firms that are part of the same supply chain, for example food processors. Currently, the most important retailers are acting as coordinators of the food supply chain, because of their impact on production, processing and consumption (Hirsch et al., 2021). Due to the relevance of supermarkets in the supply chain, they also play a role in strategies such as energy use (Pinckaers and Phillips, 2019).

Increasing the efficiency of the main food retailers would be positive for the food supply chain since they act as coordinators and the general performance of the chain could potentially improve. If promoting the financial efficiency of these corporate firms turns out to be successful, in principle the same process can be done with social and environmental efficiency for example. Additionally, improving the efficiency of the most important food retailers can increase the resilience of the food supply chain, facilitating the absorption and recovery from shocks such as COVID-19 (Macfadyen et al., 2015). Taking the previous information into consideration, the problem addressed in this thesis is the lack of current technical efficiency measures of European food retailers, and therefore the drivers of efficiency remain largely unknown.

1.2. Research objective

The general goal of this research is to gain useful insights regarding the technical efficiency and performance of food retailers. The aim of this research is not to identify which companies are failing and which ones are successful, but to contribute to the financial performance of corporate food retailers. The conclusions obtained may result useful for firm managers, policy makers and other researchers. To address the problem statement described in the previous section, the research has the following main objective:

"To analyse which characteristics condition the efficiency of European food retailers by benchmarking their economic performance using Data Envelopment Analysis."

To achieve this, three specific research objectives are stated:

- To derive the individual technical efficiency measures for multiple European food retailers.
- To determine the drivers of efficiency in European food retailers.
- To compare the standard and the bootstrap approach of the efficiency analysis.

1.3. Contributions

From a scientific point of view, determining which firms are efficient leads to hypotheses regarding the drivers of efficiency. Identifying these drivers would allow researchers to give useful advice to retail firms and policy makers to improve business performance through measures and policies (Kalirajan and Shand, 1999). It is essential for economic policy makers and economic planning to know what the current efficiency of companies is, and how far they can increase (Farrell, 1957). This research contributes to the study of drivers of financial efficiency, a field were so far many findings contradict each other (Neves et al., 2018). Regarding the analysis and methodology, the research compares standard Data Envelopment Analysis (DEA) and bootstrapping DEA. This is a chance to evaluate the relevance of correlation between DEA scores (and how important is it to consider it in the methodology) with a practical example.

1.4. Research outline/roadmap

In section 1, the problem addresed in this research and its importance are explained, which leads to the research objectives and scientific contribution. Afterwards, section 2 explains (food) retailing and the most important retailers nowadays. A brief description of the history of retailing is also available. In section 3, the core concepts and methods needed to understand this study are detailed. Additionally, the hypotheses, assumptions and theories that lead to them are explained. The methodology, which is divided in data collection and data analysis, is discussed in section 4. In section 5 the results are presented, which are followed by the discussion in section 6. The limitations of the research that may condition its reliability and validity are explained in section 7. Finally, in section 8 the conclusions gained with this research are presented, as well as follow-up studies. Note that there is an appendix available after the reference list.

2. Literature review

The Cambridge Dictionary defines retail as the activity of selling goods to the public, usually in shops (Cambridge Dictionary, 2022). The definition implies that retailers are mainly Business-2-Consumers (B2C) businesses, however they may sell to other companies as a secondary activity. Notice that the definition indicates that a shop is not necessary, and therefore it includes online retailing. The most important retailers have included an online channel in their business model, where consumers can get the products delivered (Deloitte, 2022). In fact, digital-only businesses is a potential future trends, that firms in other sectors (such as ING bank) are moving towards (Verhoef, 2021). From a supply chain perspective, retailers have many suppliers (can be thousands) to manage its inventory, and they sell products to end consumers (Ge et al., 2019). As it can be seen in Figure 1, retailing can be classified as the fifth stage of a standard food supply chain (Kamilaris et al., 2019). Traditionally, it is considered that processors and producers are beyond the scope of retailers. However, nowadays retailers are dealing with these agents in order to introduce new brands in their stores (Ge et al., 2019).

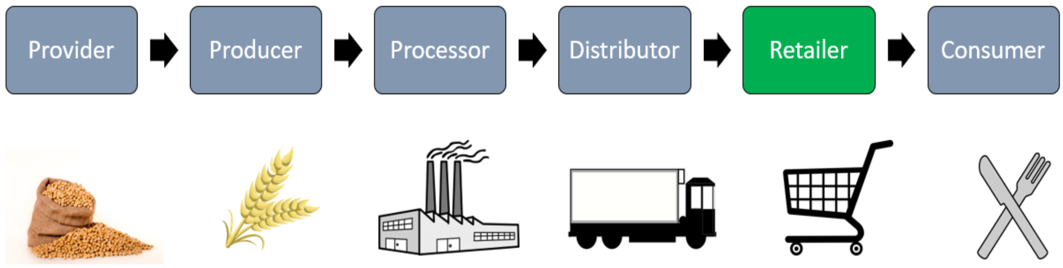

Figure 1: Example of a food supply chain, showing the six relevant stages (Kamilaris et al., 2019).

In Table 1, the top ten largest retailers in terms of revenue in 2020 are indicated. The three largest retailers worldwide are from the United States (US). Seven out of the top ten retailers are from the US, two from Germany and one from China. However, if the top 250 is considered, 33% of the companies are Europeans and 48% from the US (Deloitte, 2022). This information shows that the US dominates the retailing sector, and that there is room for improvement for European retailers. Also, the fact that the ten largest retailers are all headquartered in three countries is worth highlighting. Additionally, all of the ten firms have reported a growth in revenue, the lowest being Walgreens Boots Alliance Inc (1.5%) and the highest Amazon Inc (34.8%). All companies except Walgreens Boots Alliance have reported a higher growth on 2020 than in 2019. On average, the top ten companies operate on 12.6 companies, while the top 250 on 10.8. To sum up, the data from Deloitte (2022) shows that the largest retailers are headquartered mostly outside Europe, that they keep increasing in terms of revenue and that they are very relevant in the international market.

Table 1: Top ten retailers worldwide in terms of revenue in 2020 (Deloitte, 2022).

Rank	Company	Country	Revenue (US$M)	Main format
1	Wal-Mart Stores, Inc	US	559,151	Hypermarket
2	Amazon.com Inc	US	213,573	Non-store
3	Costco Wholesale Corporation	US	166,761	Warehouse club
4	Schwarz Group	Germany	144,254	Discount store
5	The Home Depot, Inc	US	132,110	Home improvement
6	The Kroger Co.	US	131,620	Supermarket
7	Walgreens Boots Alliance, Inc	US	117,705	Drug store
8	Aldi	Germany	117,047	Discount store
9	JD.com	China	94,423	Non-store
10	Target corporation	US	92,400	Discount store

Retailing can be classified into different categories: Apparel and accessories, fast-moving consumer goods (FMCG) and hardlines and consumer goods (Deloitte, 2022). The food industry is included in the FMCG category, since generally food products are sold quickly and at a relatively low price. 56% of the top 250 retailers in 2020 were included in the FMCG category, which indicates that it is the most important category (Deloitte, 2022). Additionally, six out of ten companies in the top ten are mainly related to the food industry (Table 1). This information shows how important the food industry is within retailing. In this study, food retailer is defined as an agent in the food supply chain that is supplied food products by distributors in order to sell them to the end consumer. Food retailers may sell products to companies, but it is not their main activity. Additionally, they may also sell non-food products, which is a very common practice amongst large retailers.

Up to the mid-1800s, corner stores were the main food retailing format in Europe and the US. These stores were mostly independent, without any common organizational structure (Stanton, 2018). The first chain grocery retailer was the Great Atlantic and Pacific Tea company, which was established in 1859 in America (Kaynak and Cavusgil, 1982). The most impactful change in the history of food retailing occurred in the early 1900s, when Clarence Saunders' Piggly Wiggly introduced the self-service store (Kaynak and Cavusgil, 1982). This new format was successful and revolutionary, since consumers now were able to walk the stores and buy products they did not initially plan to (Stanton, 2018). Additionally, since consumers were now choosing the product they want, food processors began branding their products so that the consumer would identify them amongst the competition. By the 1930s, King Kullen stores opened the first supermarket, which was on the outskirts of New York city and it was about 4 or five times larger than the grocery stores at the time (Stanton, 2018). The supermarket was a result of the accumulation of the improvements that had been achieved so far, and it lead to multiple new developments for example related to transportation and refrigeration (Kaynak and Cavusgil, 1982). Such was their success that they also started selling non-food products. In fact, the first supermarket in Europe was Jysk Supermarket, which opened in 1960 to offer food and textiles in the same store (Stanton, 2018).

Supermarkets easily consolidated in the US and Europe, and they became the most important food retail format (Stanton, 2018). Since supermarket stores were becoming bigger (which leaded to hypermarkets), the convenience stores started to appear in city centres (Kaynak and Cavusgil, 1982). These small shops became successful, for example SPAR in Europe (Stanton, 2018). Costco, and later on Walmart, implemented the warehouse club concept, where an initial fee is paid by the consumer in order to access to lower prices. Aldi was funded in Germany in 1946, and it is considered as the first discounter (Stanton, 2018). It offered lower prices than traditional supermarkets, which resulted in a quick expansion in Europe and the US. Another main distinction of discounters is that the products offered are mainly own and secondary brands, very rarely national products are available. Other discounters appeared, such as Lidl, but not all of them were successful. The food retailing sector is constantly seeking efficiency and effectiveness improvement through different aspects such as inventory management, employees and consumer interaction (Burt, 2010; Ge et al., 2019). These approaches are continuously changing and evolving, due to more knowledge being available and consumer's demand.

The current food supply chain is heavily conditioned by competition in the farm and consumer market levels. As a result, the highly concentrated sectors of food processing and retailing act as bottlenecks (Hamilton et al., 2020). Nowadays, retailers can adopt mainly two pricing strategies: exert market power on the consumer market or on the farms (Hamilton et al., 2020). Gauri et al. (2021) use a different approach to distinguish retail formats, based on the channel used: Physical-store retailing (includes department stores/hypermarkets, discount stores, supermarkets, warehouse club stores and limited assortment/convenience stores) and non-store retailing (such as catalogues, phone & TV, websites and e-commerce). They also highlight the current trend of offering, within the company, different possibilities for the consumer, mainly online shopping and physical stores. This can be done by two different approaches: multichannel (coordinating multiple independent channels) and omnichannel (integrating activities of multiple channels into one). The main advantage of online channels is the accessibility and niche items, while offline channels provide sensory information and higher delivery speed (Brynjolfsson et al., 2009). Since both channels have advantages, companies are implementing both in order to enhance consumer's satisfaction (Gauri et al., 2021).

As a consequence of COVID-19 and the forced closure of restaurants and other food services, the revenue growth of supermarkets between 2019 and 2021 was of 10.2% (Eurocommerce and McKinsey & Company, 2022). The largest European food retailer in 2020 was the Schwarz Group, the German company owning Lidl and Kaufland, which reported a revenue of US$ 144,254 million (Deloitte, 2022). The same source stated that the second largest European retailer that year was the discount store Aldi, with a revenue of US$ 117,047 million in 2020. Other large European retailers are for example Tesco PLC (from the United Kingdom), Carrefour S.A. (from France), Edeka-Verbund and Rewe Group (both German supermarkets).

3. Theoretical background

3.1. Efficiency analysis

3.1.1. Economic, technical and allocative efficiency

Determining the efficiency of a firm is relevant for the firm itself, economic theorists and policy makers (Farrell, 1957). Full efficiency of a Decision Making Unit (DMU) is attained when none of its inputs or outputs can be improved without worsening some of the other inputs or outputs (Cooper et al., 2011). The main drawback of this definition is the fact that, in most situations, the possible levels of efficiency are unknown, and therefore it Is not possible to determine to what extent a firm is efficient (Cooper et al., 2011). Due to this disadvantage, this term is generally avoided in efficiency analysis.

Economic (also known as overall) efficiency is a concept with two different components: technical and allocative efficiency (Cooper et al., 2011; Farrell, 1957; Kalirajan & Shand, 1999; Murillo-Zamorano, 2004). The term technical efficiency is preferred over relative efficiency in economic studies, and therefore in this thesis. Technical efficiency measures the extent to which a firm produces the maximum output for a given set of inputs (Farrell, 1957). Cooper et al. (2011) state that a DMU attains full technical efficiency if the performances of other DMUs do not show that its inputs or outputs can be improved without worsening other inputs or outputs. This definition suggests that benchmarking multiple DMUs is required in order to determine technical efficiency. In this case, the definition implies that the possible levels of efficiency are known, thanks to the fact that multiple DMUs are being compared. The term DMU can refer to companies, hospitals, schools and farms amongst others (Cooper et al., 2011). Basically, any entity that converts inputs into outputs can be considered as a DMU. In this research, the DMUs are firms due to its scope.

Simar and Wilson (2008) presented the output-oriented non-parametric efficiency measurement, heavily relying on Farrell's study. The production process is limited to a production set Ψ, which represents the feasible input and output combinations:

$$\Psi = \{(x, y) \in \mathbb{R}_{+}^{N+M} | x \; can \; produce \; y\},$$

where $x \in \mathbb{R}_{+}^{N}$ stands for the input vector and $y \in \mathbb{R}_{+}^{M}$ the output vector formed by real positive numbers. The upper boundary (or frontier) of this set represents the optimal solutions (i.e., if output-oriented it represents the maximum output level for a specific level of inputs). This frontier is sometimes referred as production or technology frontier ($\partial\Psi$). Inefficient DMUs are located in the interior of Ψ, while efficient ones operate somewhere along $\partial\Psi$. The Farrell-Debreu (named after the main contributors) output technical efficiency score can be expressed as:

$$\lambda(\mathrm{x}, \mathrm{y}) = \sup\{\lambda | (\mathrm{x}, \lambda\mathrm{y}) \in \Psi\},$$

where $\lambda(\mathrm{x}, \mathrm{y})$ stands for the proportionate feasible increase in outputs for a DMU located at (x,y). The right-hand side of the equation stands for the largest possible value of λ for a DMU projected into $(\mathrm{x}, \lambda\mathrm{y})$ that belongs to the productive set Ψ. If $\lambda(\mathrm{x}, \mathrm{y}) = 1$, the unit is technically efficient and therefore is in the frontier $\partial\Psi$, however if the value is higher than 1 then it is inefficient (there is a feasible increase in outputs).

The technical efficiency of a firm is relative to the set of firms used in the analysis (Farrell, 1957). Particularly, if additional firms are added to the set, the technical efficiency score of a firm may reduce, but never increase. Additionally, the technical efficiency of a firm depends on the factors used to measure it (Farrell, 1957). Variability between firms in a specific factor is positive, resulting in technical efficiency scores that actually represent efficiency and quality. To sum up, technical efficiency is the most appropriate term to evaluate firm's efficiencies in this research, and it depends on the set of DMUs as well as on the factors used to study it.

Allocative or price efficiency was introduced by Farrell (1957), and defined it as a firm's success in choosing an optimal set of inputs given input prices. Therefore, the unit prices and unit costs need to be known in order to evaluate allocative efficiency (Cooper et al., 2011). Since this research collects data from one year (2020) and the Law of One Price is assumed (see section 3.3.1.), common price indices are used. Thus, input prices are assumed to be the same for all DMUs and allocative efficiency is constant for all DMUs. Therefore, taking into account the following Farrell's (1957) equation, economic and technical efficiency coincide:

$$Economic \; efficiency = Technical \; efficiency \; x \; Allocative \; efficiency$$

3.1.2. Parametric and non-parametric methods

As discussed in the previous section, evaluating the efficiency of a DMU heavily relies on benchmarking its performance to an efficient level, which in DEA it is represented by a frontier. In the last 30 years, there has been an increase in the amount of published articles per year related to these two concepts (Castro & Frazzon, 2017). The literature distinguishes between parametric and non-parametric methods. There has been a debate amongst researchers about what methods should be used, where some authors prefer parametric and other non-parametric methods depending on their interests (Murillo-Zamorano, 2004).

A parametric model makes use of a particular analytical function with a *a priori* fixed number of parameters (Simar, 1992). In other words, econometrics are used to assume a production function that is used as a benchmark frontier (Arshinova, 2007). This type of method is subdivided into deterministic and stochastic analyses (Murillo-Zamorano, 2004; Simar, 1992). In deterministic parametric models, all observations lie below of the frontier, and the technical inefficiency is measured as the distance to the frontier. The Corrected Ordinary Least Square (COLS) methodology estimates the model's parameters (except the intercept) by OLS (Murillo-Zamorano, 2004). Stochastic parametric models, however, capture the effect of exogeneous shocks since they allow for random noise around the frontier (Simar, 1992). Stochastic Frontier Analysis (SFA) is a stochastic parametric model that decomposes the total deviation of the observations into statistical noise and technical inefficiency (Kumbhakar et al., 2021). SFA can also be used to evaluate the sources of inefficiency. The main drawback of parametric techniques is that the frontier function has to be specified by the analyst, which is an assumption that may not always be valid (Murillo-Zamorano, 2004). Therefore, parametric methods require the observance of the assumptions and restrictions imposed on the model (Arshinova, 2007).

On the other hand, non-parametric methods construct the efficiency frontier based on the investigated DMUs (Arshinova, 2007). Free Disposal Hull (FDH) and DEA are some examples of non-parametric methods (Simar, 1992). This type of methods is often conditioned by their deterministic nature (Murillo-Zamorano, 2004). For instance, it is not possible to distinguish between technical inefficiency and statistical noise. Additionally, they are very sensitive to super-efficient outliers, which can heavily decrease the technical efficiency of other DMUs (Simar, 1992). However, these methods offer high flexibility in terms of the variables used, and researchers have developed modifications to improve them, for example bootstrapped DEA, Network DEA, Stochastic DEA and Fuzzy (Emrouznejad et al., 2022; Wei, 2001).The main advantage of DEA is that it is based on fundamental axioms of production theory, and no assumptions regarding the frontier are made (Olesen and Petersen, 2016). On the other hand, the main limitation of DEA is the noise effect, i.e. attribute deviations to inefficiency.

Nevertheless, new methods have been developed that present parametric and non-parametric characteristics, resulting in cross-over methods. Stochastic Non-smooth Envelopment of Data (StoNED) is an efficiency analysis that integrates a DEA-type non-parametric frontier with a SFA-style composite error term. However, it also shares many of their limitations (Kuosmanen and Kortelainen, 2012). Another example is the stochastic DEA method, which introduces statistical axioms and distribution assumptions in a way that allows inference. This method can be used to handle 1) deviations from the frontier as random deviations and 2) random noise as measurement for example (Olesen and Petersen, 2016). Semi-parametric SFA and Chance Constrained DEA are two more cross-over methods, that allow noise and stochastic inefficiency. The main drawback of these methods is the limited application so far, higher complexity and the fact that results are harder to explain to decision makers (Olesen and Petersen, 2016).

3.1.3. Data Envelopment Analysis

Farrell (1957) stated that, back then, the techniques used to evaluate efficiency could not combine in a valid way multiple inputs. Data Envelopment Analysis (DEA) was first introduced by Charnes et al (1978), solving the issue identified by Farrell, in order to measure the efficiency of resource utilization. They developed this method due to a thesis under the supervision of one of the authors, that evaluated educational programs for disadvantaged students (Cooper et al., 2011). The DEA model they introduced (named CCR model following the authors initials) extended the definitions of efficiency given by Farrell (1957), since now multiple inputs and outputs could be considered to measure efficiency (Wei, 2001). It is important to note that the CCR model assumed constant return to scale (CRS). Researchers noticed the advantages of this method to evaluate efficiency of DMUs, for example excellent and easy methodology and few *a priori* assumptions (Cooper et al., 2011). Consequently, many different authors contributed to DEA over the next years, resulting in a more flexible and precise analysis method (Wei, 2001). Banker et al. (1984) developed a model that assumed variable return to scale (VRS), which was named the BCC model. Additionally, extensions of DEA have been developed in the last years, for example to deal with nondiscretionary and categorical variables (Cooper et al., 2011).

Charnes et al. (1978) defined DEA as “a nonlinear programming model that provides a scalar measure of efficiency of each participating unit by reference to the observational data for the multiple outputs and inputs”. Additionally, equivalences to ordinary linear programming models were explained. The authors considered that the intended use of the model was evaluating activities of non-profit organization. Many authors have provided more recent and simple definitions of the method. For example, Lampe and Hilgers (2015) state that DEA is a non-parametric method that works with multiple inputs and outputs by using an algorithm based on linear programming. Basically, an efficient frontier (also called Pareto Front) is created by the most efficient DMUs in the sample (Cooper et al., 2011). Inefficiency of other DMUs is measured as the distance to the frontier, and therefore efficient DMUs are used as benchmarks (Lampe & Hilgers, 2015). As a result of the analysis, technical efficiency scores are provided for each DMU, with values ranging from 0 (not efficient) to 1 (technically efficient).

As it has been hinted already, DEA can be performed by assuming constant (CCR model) or variable (BCC model) returns to scale and it can also be input or output oriented, depending on the specific research goals (Konuk, 2018). In section 3.3.2. the returns of scale of DEA are discussed into detail. Input oriented models estimate the potential input decrease for a certain level of outputs. On the other hand, output-oriented models determine the potential output increase for a certain level of inputs (Konuk, 2018). Both input and output-oriented models estimate a frontier that identify the same set of DMUs as efficient (Coelli et al., 2005). However, the efficiency measures of inefficient DMUs can differ between the two models.

In this research, the output-oriented BCC model is used to analyse food retailers. Previous studies have used these model characteristics to evaluate the performance of food retailers (Mostafa, 2009; Neves et al., 2018; Perrigot and Barros, 2008). By using an output-oriented model, the technical efficiency score obtained as a result of DEA reports by how much can the output be increased with the current input. In other words, how much can the firm increase its turnover with the current input level of the company. This information is valuable for firm managers, since their goal is to generally maximize profit. The model is composed by the following formulas (Banker et al., 1984; Coelli et al., 2005; Cooper et al., 2011):

$$\max_{\varphi,\lambda} \varphi$$

Subject to:

$$-\varphi y_j + Y\lambda \geq 0 \; for \; all \; r$$

$$x_j - X\lambda \geq 0 \; for \; all \; i$$

$$\sum_{j=1}^{n} \lambda_j = 1$$

$$\lambda_j \geq 0$$

Where:

i=1, 2, …, m

r=1, 2, …, s

j=1, 2, …, n

In this model, DMU_j consumes x_{ij} of input *i*, and generates y_{rj} of output *r*. *X* is the *mxn* input matrix, and *Y* is the *sxn* output matrix. It is assumed that $x_{ij} \geq 0$ and $y_{rj} \geq 0$. φ represents the optimal solution of the linear problem, and it yields an efficiency score for DMU_j. If $\varphi = 1$, the corresponding DMU is located in the efficient frontier, and therefore it is considered as technically efficient. If $\varphi < 1$, the corresponding DMU is not located in the efficient frontier, and therefore it is inefficient. λ represents a *nx1* vector of parameters, also known as firm weights. Each inefficient DMU is projected into the efficient frontier thanks to a peer group. A peer group is a set of efficient DMUs that have a positive value of λ. Efficient DMUs do not require a peer group because they are already part of the efficient frontier, and therefore they do not need to be projected. Identification of the peer units can be useful for example to highlight weaknesses and to set target values of input or output (Martić et al., 2009).

As it can be noticed, the model follows the structure of linear programming, where there is an objective function subjected to multiple constrains. The above model attempts to achieve, via φ, as much expansion of y_{rj} as it is allowed by the constrains (Charnes et al., 1994). It is important to keep in mind that this model is used on all the DMUs, in order to obtain their technical efficiency scores. The first constrain makes sure that, for output *j*, its projected value ($Y\lambda$) is larger or equal than its actual value times the efficiency score (φy_j). Because of how φ is used in the first constrain, it can also be defined as the simultaneous expansion of all outputs, till the constrains allow it. The second constrain ensures that, for every input *i*, its projected value ($X\lambda$) is smaller or equal than its actual value (x_j). The third constrain (which states that the sum of λ_j =1) is particular of BCC models, and it is named the convexity constrain. Without this constrain in the model, a CCR model is obtained.

In order to help understanding the method, a graphical example of an output-oriented BCC model is provided in Figure 2. In this example, there are four DMUs (A,B,C,D) represented. Three of them (A,B,C) compose the Pareto Front or efficient frontier, which implies that they are technically efficient. In other words, this analysis suggests that these DMUs cannot produce more outputs with their current level of inputs. However, this is not the case for DMU D. As it can be seen, DMUs B and C generate more outputs with lower and higher levels of inputs, respectively. For DMU D, the projected value of its output is O_2 and its actual value is O_1. The technical efficiency of a DMU decreases as the distance to the efficient frontier increases. For DMU D, the technical efficiency score can be calculated as follows:

$$TE_D = \frac{O_1}{O_2}$$

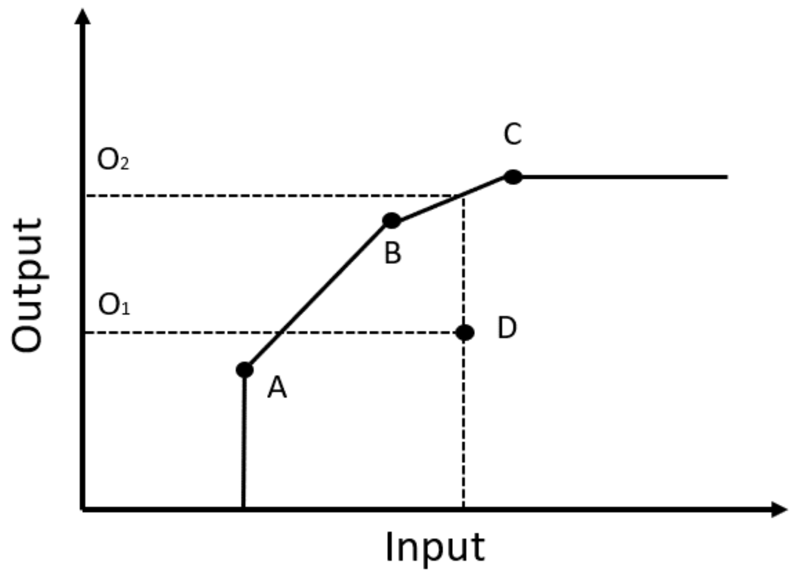

Figure 2: Example of an output-oriented BCC model.

3.1.4. Two-step efficiency analysis

DEA is usually followed by a second step, where a regression model is used to evaluate potential drivers of efficiency. When DEA and regression analysis are combined, it is called a two-step efficiency analysis. Basically, the efficiency scores obtained in the first stage are regressed on environmental variables, which are different from the variables used in the first stage (Simar and Wilson, 2007). The goal is to determine if the environmental variables are drivers (also known as determinants) of efficiency for the analysed DMUs.

Two different methods have been extensively used in studies to run a two-step analysis: 1) Standard DEA followed by Ordinary Least Squares (OLS), and 2) Double bootstrap procedure that uses a truncated regression (Banker & Natarajan, 2008; Neves et al., 2018; Simar and Wilson, 2007, 2011). Banker and Natarajan (2008) described the conditions under which a standard DEA followed by OLS provides consistent estimators of the impact of contextual variables. Additionally, they demonstrated that high correlation between the first-stage variables and environmental variables worsen the two-stage DEA method. This finding is considered in this research, and therefore the environmental variables are not included in the DEA model.

The double bootstrap procedure is detailed in the next section. A truncated regression, which is used in the procedure, excludes observations if the value of the dependent variable is above or below a particular threshold. The main difference with a censored regression (which has also been used in two-step analyses) is that the latter concentrates observations at a single value if the values of the dependent variable are above or below a certain threshold (Maddala, 1983). In the bootstrap approach, truncated regressions are used to exclude DMUs that have a technical efficiency score greater than 1 (Cooper et al., 2011).

Regarding the two-step analysis proposed by Banker and Natarajan (2008), it has been shown that it is efficient under specific and unusual assumptions, which limits its applicability (Simar and Wilson, 2011). However, Banker and Natarajan (2008) also discussed limitations regarding the double bootstrap developed by Simar and Wilson (2007), such as it being even more restrictive and imposing unlikely assumptions. Since both methodologies are still relevant nowadays in two-step efficiency analyses, the two approaches are used in this research. In section 4.2. their use is explained.

3.1.5. Double bootstrapped DEA

Cooper et al. (2011) stated that DEA is often characterized as deterministic, where efficiency is computed and not estimated. However, this is not always the case. Simar and Wilson (2007) noticed that the efficiency scores obtained in the standard DEA are heavily correlated, and therefore the second stage of the analysis is invalid. Additionally, they also stated that, up to that moment, the correlation between efficiency scores had not been taken into account in most of the two-step analyses. Consequently, they developed a methodology to implement bootstrap in the two-step analysis. It has been proven that bootstrap methods are needed for inference in the second stage (Simar and Wilson, 2011).

Simar and Wilson (2007, p. 42) developed two different algorithms, to run single and double bootstrap procedures. The latter is preferred since it improves the statistical efficiency in the second-stage regression. This method provides bias-corrected efficiency scores and robust regression coefficients, both with their corresponding confidence intervals. Note that this approach is used in this research, and that therefore this study heavily relies on these authors' contribution. The double bootstrap algorithm works in the following way:

- <u>Step 1</u>: Compute the efficiency scores (φ_i) for all DMU_i using a DEA model.
- <u>Step 2</u>: Use the maximum likelihood method to obtain an estimate $\hat{\beta}$ of β (regression coefficient) and an estimate $\hat{\sigma}$ of σ (standard deviation) in the truncated regression of φ (dependent variable) on z_i (independent variables).
- <u>Step 3</u>: Loop over the next four steps L_1times to obtain *n* sets of bootstrap estimates.
    - 3.1. For each DMU, draw ε_i (error term of the regression) from the (N, $\hat{\sigma}_\varepsilon^2$) distribution with left truncation at (1-$z_i\hat{\beta}$)
    - 3.2. For each DMU, again compute $\varphi_i^* = z_i\hat{\beta} + \varepsilon_i$
    - 3.3. Set $x_i^* = x_i, y_i^* = y_i \frac{\widehat{\varphi_i}}{\varphi_i^*}$ for all DMU_i

- 3.4. Compute $\hat{\varphi}_i^* = \varphi(x_i, y_i | Y^*, X^*)$, where $Y^* = [y_1^* \dots . y_n^*]$, $X^* = [x_1^* \dots . x_n^*]$

- Step 4: For each DMU_i compute the bias-corrected estimator $\hat{\hat{\varphi}}_i$ by $\hat{\hat{\varphi}}_i = \hat{\varphi}_i - \widehat{BIAS}(\hat{\varphi}_i)$ using the bootstrap estimates obtained in step 3.4 and the original estimate $\hat{\varphi}_i$. $\widehat{BIAS}(\hat{\varphi}_i)$ represents the bootstrap bias estimate.
- Step 5: Use the maximum likelihood method to estimate the truncated regression of $\hat{\hat{\varphi}}_i$ on z_i, yielding estimates $(\hat{\hat{\beta}}, \hat{\hat{\sigma}})$.
- Step 6: Loop over the next three steps L_2times to obtain *n* sets of bootstrap estimates:
    - 6.1. For each DMU_i, draw ε_i from the (N, $\hat{\sigma}_\varepsilon^2$) distribution with left truncation at (1-$z_i\hat{\hat{\beta}}$)
    - 6.2. For each DMU, again compute $\varphi_i^{**} = z_i\hat{\hat{\beta}} + \varepsilon_i$
    - 6.3. Use the maximum likelihood method to estimate the truncated regression of φ_i^{**} on z_i, yielding estimates $(\widehat{\widehat{\beta^*}}, \widehat{\widehat{\sigma^*}})$
- Step 7: Use the bootstrap values $\widehat{\widehat{\beta^*}}$, $\widehat{\widehat{\sigma^*}}$ and the original estimates $\hat{\hat{\beta}}$, $\hat{\hat{\sigma}}$ to construct estimated confidence intervals for each element of β and σ.

The first step estimates efficiency scores (φ_i) for all DMU_i using the chosen DEA model, and the second step uses a truncated regression to estimate the regression coefficients ($\hat{\beta}$) of the environmental variables (z). The first bootstrap is detailed in step 3, where error terms ε_i are sampled from a normal truncated distribution to compute new efficiency scores L_1 times. Afterwards, in step 4, the final bias-corrected efficiency score $\hat{\hat{\varphi}}_i$ is computed. In step 5, the regression coefficients are estimated again using the new bias-corrected efficiency score $\hat{\hat{\varphi}}_i$. The second bootstrap is explained in step 6, where error terms ε_i are sampled from a normal truncated distribution to compute new regression coefficients L_2times. Afterwards, the obtained and original estimates can be used to determine confidence intervals of the robust regression coefficients. The recommended L_1 and L_2 values are 100 and 2000, respectively (Simar and Wilson, 2007).

The double bootstrap procedure developed by Simar and Wilson (2007) has been extensively used for hypothesis testing. Gandhi and Shankar (2014) used this approach on Indian retailers, and they found that the number of outlets and mergers and acquisitions can be considered as drivers of efficiency. Yu and Ramanathan (2009) used the same method to study the efficiency of retailers in China, concluding that retail sector is a significant driver of efficiency. Neves et al. (2018) evaluated the efficiency of Brazilian supermarket chains using a two-step analysis with bootstrap, highlighting that, according to their findings, most of the food retailing studies did not make use of the double bootstrap approach.

3.2. Relevant theories

These theories are part of the conceptual framework that leads to the hypotheses stated in section 3.4. (Table 2). Therefore, it is essential to explain them in order to understand how the hypotheses were reached, as well as the implications of the results and their discussion.

3.2.1. Economies and diseconomies of scale

The origin of economies of scale (EOS) is uncertain, as it derives from a list of sources. Adam Smith described the division of labour as a way to gain efficiency, and while doing so he introduced the concept of returns of scale (Lindsay and Maloney, 1996). This is considered as the first contribution to the EOS theory. In 1956, Robinson published *The Accumulation of Capital*, a book where other aspects such as economies of mass reserves (inventory), of management and of large machines were described. These effects are also part of EOS, since the principle is the same.

EOS has been classified based on the level of aggregation. Junius (1997) considered that internal EOS refers to individual firms, while external EOS refers to industries or regions. Due to the scope of this research, internal EOS is the main focus. Additionally, this author distinguishes between dynamic and static internal EOS. Dynamic EOS considers learning effects (i.e., higher output leads to higher productivity because there is a learning process within the firm), but static doesn't. This learning effect was described by Adam Smith, and it may have been more common back then compared to today (Lindsay and Maloney, 1996).

Regardless of the learning effect, a firm experiences EOS if the long-run average unit costs decline as the firm expands its output (Rittenberg and Tregarthen, 2012). A firm can experience it due to specialization, which means that as the firm's operations increase, the use of its factors is more specialized and productivity increases. Additionally, increasing the output makes it possible to use large-scale processes and systems that reduce the cost per unit (Rittenberg and Tregarthen, 2012).

On the other hand, it is also possible for a firm to experience diseconomies of scale (DOS). It is the opposite of EOS, since long-run average unit costs increase as the firm expands its output (Rittenberg and Tregarthen, 2012). Due to DOS, there is a limit in firm size and growth (Canback et al., 2006). The neoclassical relationship between unit cost and output (Figure 3) shows that when EOS are exhausted, DOS increase the unit cost (Canback et al., 2006). In this approach there is one desired level of output, where unit cost is minimized (S). However, this is not the only approach. Stigler (1958) proposed a different relation between unit cost and output, where there is a wide range of outputs (S_1 and S_2) for which the unit cost is basically constant (Figure 3). This model explains why large and small firms coexist in competitive markets, while still stating that there are limits to firm size (Canback et al., 2006).

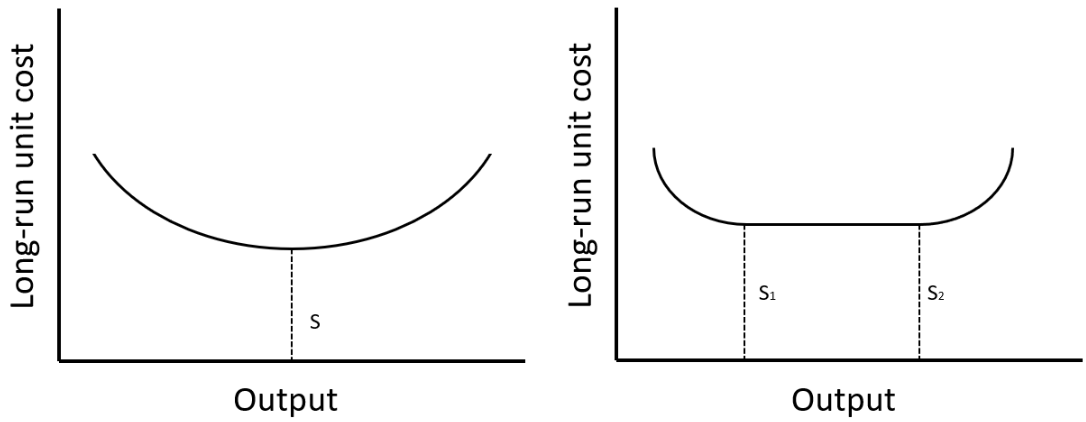

Figure 3: The neoclassical approach of economies and diseconomies of scale is represented on the left (Canback et al., 2006). On the right, the approach suggested by (Stigler, 1958).

3.2.2. Trade-off theory

Financial leverage can be defined as the proportion of a firm's assets that is financed with debt instead of equity, which is represented by the debt-to-equity ratio (Çerkezi, 2013). Firms manage their capital structure with the goal of optimizing the value of the company. Particularly, managers aim to reduce the Weighted Average Cost of Capital (WACC) of the company in order to maximize the shareholders return (Arnold, 2005).

Modigliani and Miller (1958) introduced the *Irrelevance of Capital Structure* theorem, stating that in a perfect market (where there are no taxes) financial leverage does not condition the value of a firm. However, in the actual world this is not the case, and capital structure conditions the value of the firm. In their second proposition, the effect of leverage on the value of the firm is studied taking into account taxes. It is said that leverage in this case increases the value of the firm, since the tax shield also increases and it benefits the company. In their study extremely high values of leverage are not considered. This proposition is closer to the real world but it still does not represent reality, since companies don't finance their assets exclusively with debt. Nevertheless, these theorems lead to many new studies related to capital structure.

The Trade-off theory builds on their theorems, considering the effects of taxes and bankruptcy costs (Çerkezi, 2013). Myers (1984) stated that firms maximize their value with a debt-to-equity ratio that trades off the advantages and disadvantages of debt. This concept is represented in Figure 4. In other words, the value of the firm is not optimized by maximizing or minimizing the financial leverage, but by balancing different costs. The Trade-off theory is divided into two types: static and dynamic. The static theory suggests that companies balance bankruptcy costs and the marginal present value of interest tax shield (Çerkezi, 2013). The dynamic variant suggests that, since there is a cost associated to issuing and repurchasing debt, companies adjust the capital structure when the benefits of doing so outweigh the costs (Çerkezi, 2013). As a result, the debt-to-equity ratio varies within an optimal range.

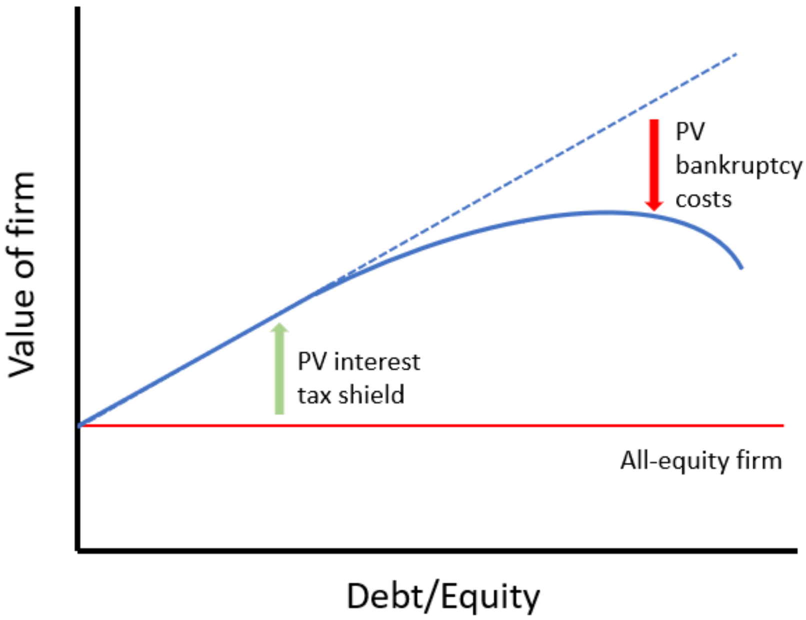

Figure 4: Representation of the trade-off theory (Corelli, 2018).

Regarding the advantages of debt, it is less expensive than equity finance because the costs of raising funds are lower (e.g., arrangement fees). Additionally, debt holders have priority over equity holders for annual pay-outs and liquidation processes, resulting in lower rate of returns to attract investors (Arnold, 2005). As discussed before, financial leverage provides a tax shield to companies because it reduces the income taxes and therefore saves cash flows (Modigliani & Miller, 1958). Other benefits of debt are explained by the Free Cash Flow theory, which are discussed in section 3.2.3.

On the other hand, relying heavily on debt has a negative impact on the firm. Increasing debt results in higher agency costs of debt, such as bankruptcy costs (Jensen, 1986). As leverage increases, the debt holders will demand higher interest rates and shareholders higher returns from investments (Çerkezi, 2013). Additionally, equity can act as a shock absorber if the company performs poorly over a few years. The firm is not forced to pay dividends or to repay capital to the shareholders, but in the case of debt the firm is obliged to pay under the agreed conditions (Arnold, 2005).

3.2.3. Free cash flow theory

Agency costs originate from the "principal-agent theory": In large corporate firms, the shareholders (principals) generally don't manage the firm, and a manager (agent) is employed to do so (Arnold, 2005). However, the agents don't always act for the best interest of the principals. These conflicts of interests can be considered as an expense, which is called agency cost. One of the decisions to be made by corporate managers is the payout policy, through dividends or repurchase of stock. Shareholders are interested on payouts, however payouts reduce the manager's resources and power. Additionally, growing the company over optimal size has advantages for managers, such as more power, resources and economic compensation (Jensen, 1986), but not for the firm due to for example DOS.

Free cash flow is defined as any cash flow in excess of that required to fund all projects that have positive net present values when discounted at the relevant cost of capital (Jensen, 1986). The Free cash flow theory states that debt reduces the agency costs of free cash flow by reducing the cash flow available for spending at the discretion of managers. Managers are forced to pay future cash flows if debt is issued, which may not be the case if stock is issued. This results in a decrease of the agency costs. It is also stated that high leverage motivates the organizations to be efficient, due to the risk of failure to make debt payments (Jensen, 1986).

However, the theory also considers that increased leverage results in increased agency costs of debt, as stated in the trade-off theory. Therefore, the optimal debt-to-equity ratio is the point where the marginal costs of debt offset the marginal benefits.

Interestingly, the Free Cash Flow theory predicts that some mergers and acquisitions (M&A) likely result in less value, rather than creating (Jensen, 1999). M&A can result in free cash flow being spent, and therefore it is not paid out to shareholders which generates agency costs of free cash flow. Corporate managers that have high free cash flows available are more likely to complete low-benefit or value-destroying M&A (Jensen, 1999). The impact of M&A depends on the industry, internal organization, waste of resources in the firm and how the M&A is financed.

3.3. Assumptions

3.3.1. The law of one price

The Law of One Price (LOP) states that, when prices are converted to a common currency, the same good is sold for the same price in different countries (Miljkovic, 1999). Because of this, firms face the same price for their inputs and outputs if LOP is assumed (Kuosmanen et al., 2006). Assuming LOP is essential for many studies that have an international scope, since if it is violated the conclusions would be invalid (Baffes, 1991). The theory states that commodities (domestic and foreign) are characterised by a high degree of substitutability, and therefore arbitrage is perfect and instantaneous (Milone, 1986). Arbitrage stands for the purchase or sell of a good or commodity in order to make a profit. Due to arbitrage, the prices, which can be different in the beginning, end up converging.

The empirical evidence related to LOP is extensive, and some of it supports the theory but there are other studies that do not (Miljkovic, 1999). Baffes (1991) highlighted the importance of also adjusting the price in respect of transportation costs, since it may be a reason for the LOP to fail. Additionally, Miljkovic (1999) stated that export demand variation, exchange rate risk and again transportation costs could be causes why LOP does not hold. Cox and Wohlgenant (1986) analysed the variation of cross-sectional prices of multiple commodities, and they also identified quality effects as responsible for this variation. However, they evaluated these variances and concluded that the quality effects are quite small and can be considered as non-relevant. These findings are relevant for efficiency analysis studies, since DEA assumes that the inputs and outputs of the DMUs have the same quality, and they only differ in quantity (Kuosmanen et al., 2006). Cox and Wohlgenant (1986) also showed that quality differences are expressed as quantity differences when using a common price index, like this research (since the data collected refers only to 2020, the price index is common).

This research heavily relies on this law being held. The goal is to benchmark firms from different European countries, which sell very similar products. In order for the comparison to be valid, LOP must be assumed. Kuosmanen et al. (2006) explain that, even if DEA is based on inefficiencies and LOP on perfect competition, LOP is a requirement for DEA.

3.3.2. Variable returns to scale

In section 5.1.2. two different types of returns to scale in DEA were introduced. Constant returns to scale (CRS) is considered in the CCR model, and it suggests that a proportional increase in the input level will result in an equally proportional increase on the output level (Konuk, 2018). CRS assumes that all DMUs under the analysis perform at an optimal scale, which is most likely not the case (Murillo-Zamorano, 2004). If this assumption is not met, the model may provide misleading measures of technical efficiency, due to a biased scale of efficiency.

On the other hand, variable returns to scale (VRS) suggests that an increase in the input level can lead to an unequally proportional (increasing-decreasing return) increase on the output level. In other words, it considers the possibility that the average productivity at the most productive scale size may not be possible in different scale sizes (Banker et al., 1984). VRS is assumed in the BCC model (Banker et al., 1984; Konuk, 2018). As stated in section 5.1.2., this assumption is modelled by adding the convexity constrain ($\sum_{j=1}^{n} \lambda_j = 1$). VRS avoids biased scale of efficiency, and therefore the technical efficiency scores are not damaged (Murillo-Zamorano, 2004).

The example introduced in section 3.1.3. is used again (Figure 2), now to explain graphically the differences between VRS and CRS (Figure 5):

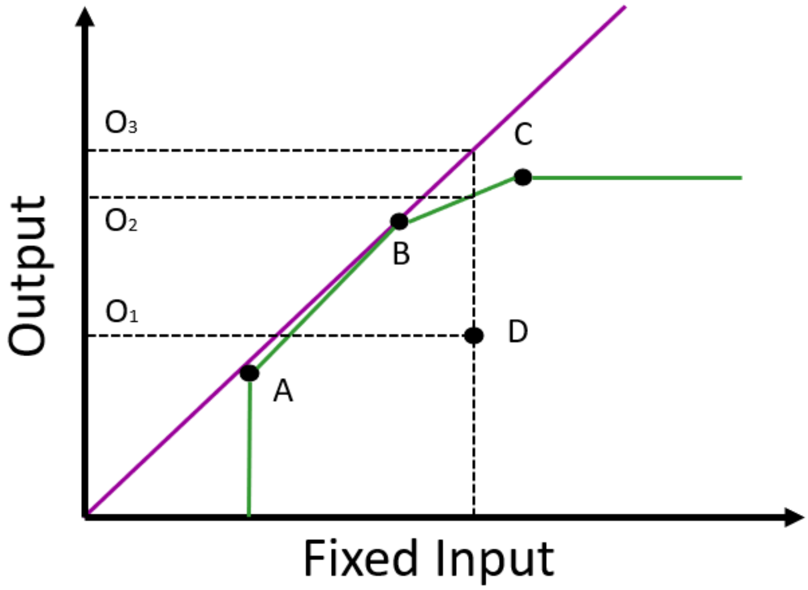

Figure 5: Example of an output-oriented model, where the VRS frontier (green) and CRS frontier (purple) are represented.

In the example provided above, the VRS frontier is formed by the efficient DMUs (A,B and C). However, the CRS frontier is determined by the DMU with the highest $\frac{Output}{Input}$ ratio (B). In both models, DMU D is inefficient. The technical efficiency of D in the different models is determined using the following formulas:

VRS: $TE_D = \frac{O_1}{O_2}$ CRS: $TE_D = \frac{O_1}{O_3}$

As it can be concluded from Figure 5, TE_D under CRS is smaller than under VRS in this example. In this research, VRS is assumed, and therefore the BCC model is the one used. Other researchers have assumed VRS to evaluate the efficiency of food retailers (Mostafa, 2009; Neves et al., 2018; Perrigot and Barros, 2008). Additionally, due to economies and diseconomies of scale, it is not expected that average productivity in the most productive input level is possible in all input levels. Finally, it is unlikely that all of the firms evaluated perform at an optimal scale, as stated by Murillo-Zamorano (2004). Therefore, assuming VRS rather than CRS makes the model more realistic in this sense.

3.3.3. Disposability

The conventional constant and variable returns to scale models assume strong or free disposability of all inputs and outputs. Considering a DMU that is part of a production technology (Ψ), strong disposability assumes that increasing any of its inputs or reducing any of its outputs keeps the DMU in the technology (Mehdiloo and Podinovski, 2019). For example, if the input level increases from x to x', the output level y that was obtained with x can still be feasibly obtained with x' inputs. Additionally, given a level of input x and of output y, y' can also be produced as long as $y' \leq y$.

There are other types of assumptions such as weak disposability, which considers the relative proportions between the outputs or inputs. Nevertheless, free disposability is assumed because it represents better the interaction between parameters of firms. For example, if a food retailer increases the number of employees (input of the DEA model) it is still feasible to obtain the same turnover (output). Additionally, firms can modify independently parameters such as number of employees, inventory and fixed assets. Assuming a different type of disposability would not capture the behaviour of firms properly.

3.4. Hypotheses

The potential drivers of efficiency studied in this research are firm-specific factors, rather than environmental factors (e.g., unemployment rate in the corresponding country). The aim is to obtain insights and conclusions which are part of the scope of the companies. Therefore, if the company desires, the characteristics of the organization can be changed in order to seek efficiency. In table 2 an overview of the potential drivers of efficiency studied and their expected correlation with efficiency is provided.

International activity has multiple advantages for companies. In theory, it results in a competitive advantage, increased turn value due to economies of scale and larger scope, growth opportunities, diversification benefits and new potential resources, facilities and knowledge (Attig et al., 2016). These authors also stated that there are some potential drawbacks from international expansion, such as increased pressure from a bigger list of stakeholders and a potentially hostile international environment.

> H_1: The international activity of a firm is positively correlated with technical efficiency.

It is frequently hypothesized that, in general, the size of the firm and technical efficiency are negatively correlated (Hanousek et al., 2015). Even if larger firms can profit more from economies of scale, they may suffer from bureaucratic frictions, lack of motivation and difficulty in monitoring (Diaz and Sanchez, 2008). Hirsch et al. (2021) described the differences between small and large firms within the European food retailing sector. Large firms have the advantage in terms of bargain position, scope, economies of scale, distribution and access to the capital market. This results in less competition and an increased persistence for profit in large dominant firms. In this research the hypothesis is that, due to economies of scale and the competitiveness of the food retail industry, size and efficiency are positively correlated. Size is captured by the number of outlets and being a top five European food retailer in terms of revenue (Hirsch et al., 2021; Neves et al., 2018).

> H_2: The number of outlets of a firm and being a top five European food retailer in terms of revenue are positively correlated with technical efficiency.

Intrafirm transactions (within one company) may be more efficient than interfirm transactions (between two different companies), as deals become more idiosyncratic. In other words, particular transactions are more efficient when they are done within the organization. Teece (1982) used this argument to justify that multiproduct firms may be more efficient than specialized firms. Multiproduct firms have an internal organization that enables the transfer of human and physical resources. This flexibility can lead to higher revenue and efficiency. Due to standard balance sheets and the size of the companies studied, it is not possible to obtain a breakdown of the revenue and input purchases per product/market, which makes it hard to capture the effect. Therefore, in this research the flexibility of multiproduct firms is represented by the number of firms in the corporate group (i.e., more firms within the corporate group imply more flexibility for the organization and more efficiency). Additionally, there are several reasons why corporate groups exist: 1) reduced transaction costs (as already mentioned), 2) coordinated investments across firms resulting in a "big push", 3) providing mutual insurance and reducing risk and 4) higher monopoly power (Yafeh, 2003). It is hypothesised that these advantages become stronger as the number of firms in the corporate group increases, resulting in higher efficiency.

> H_3: The number of firms in the corporate group is positively correlated with technical efficiency.

In section 3.2. the Trade-off theory and the Free cash flow theory are described. It is hypothesised that these theories also apply to firm efficiency.

> H_4: The relationship between the debt-to-equity ratio and technical efficiency has an inverted "U" shape.

Teece (1982) theorised that mergers and acquisitions may contribute positively to economic efficiency of the acquiring and acquired firms. The acquiring firm acts as a way to use discipline for the acquired firm, reducing managerial discretion. Arnold (2005) indicated that in mergers and acquisitions there is a synergy between both firms: the two firms together are worth more than the value of the firms apart.

> H_5: Merge and acquisition strategy is positively correlated with technical efficiency.

Table 2: Overview of the potential determinants of efficiency studied and their expected correlation with technical efficiency.

Potential determinant of efficiency	Hypothesis number	Expected correlation	Literature/Theory
International activity	1	+	Attig et al., 2016
Size	2	+	Hirsch et al., 2021 / Economies of scale
Number of companies in corporate group	3	+	Teece, 1982; Yafeh, 2003
Debt-to-equity ratio	4	-/+	Trade-off theory Free cash flow theory(Jensen, 1986)
Merge & Acquisition	5	+	Arnold, 2005; Teece, 1982

4. Methodology

4.1. Data collection

A summary of the data collection process is given in Table 3. The Orbis database, owned by Bureau van Dijk, was used in the first steps of the data collection and it was accessed through Wageningen University & Research. Using the filters and Boolean search that appear in Table 3, it was possible to gather financial information of 25 European firms related to retailing. The searched data refers to 2020. However, companies may publish their results on any date up to 31 March 2021. Orbis provided information on the eight following variables: number of employees, fixed assets, inventory, operating revenue (turnover), number of companies in the corporate group, non-current liabilities, current liabilities and shareholder’s funds.

Once the 25 firms were obtained, each of them was individually checked in order to remove those that are not related to food retailing. As a result, 7 of the 25 firms were deleted from the data set. Afterwards, it was also checked that all companies were independent from each other. In the database, there is a distinction between subsidiaries and ultimate owners of the group, for which usually there is consolidated financial statements. The 19 firms were checked, in order to remove the subsidiaries and only keep the ultimate owners. Consequently, 4 companies were removed. In the end, all the financial information in the data set belongs to consolidated financial statements of the corporate firms. Lastly, it was not possible to find consolidated, complete and reliable financial information of Aldi, which is one of the most important food retailers. Therefore, it was removed from the dataset.

In order to increase the size of the dataset, the Global Powers of Retailing Report (Deloitte, 2022) was used to identify relevant food retailers that were not included yet. European food retailers with more than 10,000 million euros of revenue in 2020 were listed. Afterwards, if their consolidated financial statements were available on their annual reports, they were included in the experimental data set. In the end, the following 5 companies were included: Rewe combine, Edeka Zentrale, Auchan retail, ICA Gruppen and The Co-operative Group. The final experimental data set contains information of 18 companies, all of which reported a revenue higher than 10,000 million euros in 2020.

Table 3: Overview of the data collection process, classified in four main stages.

1. Orbis search steps		**Number of firms**
1) Status	Active companies, unknown situation	323,353,864
2) World region	Western Europe, Eastern Europe	130,774,504
3) BvD sector	27 – Retail	38,800,084
4)Operating revenue (Turnover)	Top 25, 2020, exclusion of companies with no recent financial data and Public authorities/States/Governments	25
Boolean search	4 (1 and 2 and 3)	
2. Individual check		**Number of firms**
1) Industry	Food retailing as the main activity	18
2) Ownership	Subsidiaries are removed	14
3) Information	Lack of reliable information	13
3. Missing food retailers		**Number of firms**
1) Addition	Relevant European food retailers with information available	18
4. Research of missing variables		**Number of firms**
1) Research	Addition of missing variables	18

Not all the variables of this research are available in the Orbis database. To gather this information, annual reports of 2020 and the official websites of each company were used. Particularly, the information gathered referred to: international activity (the firm is active in multiple countries, those outside of Europe are also considered), number of outlets and merge and acquisition strategy (in the last 20 years the company has acquired another food retailing company or has been acquired by another food retailing company). It is frequently hard to classify mergers and acquisitions independently in real cases (Arnold, 2005), which is why they are both considered in one variable. The debt-to-equity ratio is calculated by adding current and non-current liabilities, and then dividing by the shareholders funds. In order to study the fourth hypothesis (H_4), the variable debt-to-equity squared was created. The Top 5 variable represents if a company is amongst the top five European food retailers in terms of revenue or not before taking into account price level differences. In the final experimental data set, there are three dummy variables (International activity, Top 5 and Merge & Acquisitions) for which 0= No and 1=Yes.

4.2. Data analysis

Firstly, in order to take into account the differences amongst countries, the Purchasing power parities (PPPs) index is used. In this research, the PPPs of the specific countries are gathered from the Organisation for Economic Co-operation and Development (OECD) and refer to 2020 (OECD, 2023). This intergovernmental organisation defines PPPs as the rates of currency conversion, that eliminate the difference in price levels amongst countries in order to equalise the purchasing power of different currencies (OECD, 2023). The country-specific PPPs are presented in Table 4. In order to measure PPPs, a basket of comparable goods and services representative of consumption patterns is defined, and the price levels are compared amongst multiple countries (Eurostat, 2018).

Table 4: Purchasing power parities (PPPs) of the countries included in the experimental data set in 2020 (OECD, 2023).

Country	Purchasing power parities
Netherlands	0.8
France	0.7
United Kingdom	0.7
Germany	0.7
Spain	0.6
Russia	24.5
Sweden	8.7
Portugal	0.6

In order to correct for the different purchasing powers, the financial observations of each firm were divided by the corresponding PPPs index. Particularly, the corrected financial variables were fixed assets, inventory, turnover, current and non-current liabilities and equity. For example, in the case of Tesco PLC, all of these variables were divided by 0.7 in order to obtain the corrected values. Note that the debt-to-equity ratio, and therefore the squared ratio, do not change when PPPs is considered.

To carry out the analysis, R Studio software was used. The "rDEA" package offers the possibility to estimate robust and standard DEA scores with and without environmental variables (Simm and Besstremyannaya, 2020). As explained in section 3.1.4., two different methods are frequently used to determine the technical efficiency scores and the significance of the potential determinants of efficiency. Both methods have been used on the same experimental data set, and the corresponding R code is available in Appendix 1.

4.2.1. Standard two-step analysis

The DEA model used in the first step is output-oriented, and it assumes VRS (section 3.1.3.) (Mostafa, 2009; Neves et al., 2018; Perrigot and Barros, 2008). Employees, Fixed assets and Inventory are used as input variables, and turnover as output (Mostafa, 2009; Neves et al., 2018). Once the technical efficiency scores are obtained, a standard OLS regression model is used (Banker and Natarajan, 2008). The reciprocal of the scores (i.e., $\frac{1}{\varphi_i}$) are used as the dependent variable, while the independent variables are the different potential determinants of efficiency (Group, Debt to equity, $(D/E)^2$, International activity, Outlets, Top 5 and Merge & Acquisition). Therefore, the regression coefficients (β) measure the effect of the environmental variable on the technical inefficiency. The regression model used to test the hypotheses is the following:

$$\frac{1}{\varphi_i} = \beta_0 + \beta_1 Group_i + \beta_2 DE_i + \beta_3 DESQ_i + \beta_4 IACT_i + \beta_5 Outlets_i + \beta_6 TOP_i + \beta_7 M\&A_i + \varepsilon_i$$

4.2.2. Double bootstrap procedure

The "rDEA" package is used to run a bias-corrected DEA with environmental variables. The package implements Simar and Wilson's (2007) second algorithm (Section 3.1.5.), and returns the bias-corrected technical efficiency scores and the robust regression coefficients, both with their corresponding confidence intervals. Just as in the previous approach, the DEA model is also output-oriented, and it assumes VRS. Additionally, it uses the same inputs, outputs and potential drivers of efficiency (environmental variables) as in the standard two-step analysis. The values of L_1 and L_2 are the ones recommended by Simar and Wilson (2007). Regarding the robust regression coefficients, they belong to a truncated regression of the reciprocal of the DEA score on the environmental variables.

4.2.3. Robustness check

A robustness check is used to evaluate the same hypotheses (stated in section 3.4.) with the same methodology (standard and bootstrap approach). However, in this case the assumptions are slightly different, particularly that all of the European countries have the same purchasing power (i.e., same PPPs value). Therefore, the financial variables are not corrected using the PPPs stated in Table 4. For this scenario, the descriptive statistics are listed in Table 5.

Table 5: Descriptive statistics of the variables used in the robustness check (i.e., PPPs are not considered).

Variable	Mean	Standard deviation	Min	Max
Employees	190,915.67	119,731.92	23,196.00	414,000.00
Fixed assets	16,970,447,290.84	11,204,470,665.15	4,605,300,000.00	40,101,121,153.30
Inventory	2,092,528,784.86	1,469,892,439.77	460,000,000.00	5,327,000,000.00
Turnover	36,914,055,584.76	22,386,923,552.79	11,472,000,000.00	74,955,000,000.00
Group	364.56	422.98	15.00	1,305.00
Debt to equity	3.22	2.22	0.57	11.37
$(D/E)^2$	15.05	28.90	0.33	129.28
International activity	0.61	0.50	0.00	1.00
Outlets	7,210.89	6,589.90	492.00	21,564.00
Top 5	0.28	0.46	0.00	1.00
Merge & Acquisition	0.83	0.38	0.00	1.00

5. Results

5.1. Data set

The descriptive statistics of the experimental data set are presented in Table 6. The standard deviation of the quantitative (non-dummy) variables is quite large compared to their means, which indicates that there is high heterogeneity in the data set and therefore in the food retail sector. Additionally, this characteristic is also supported by the large differences between their minimum and maximum values. The means of the dummy variables indicate that, on average, food retailers have international activities (they operate in more than one country) and most of them have also been involved in merges and acquisitions in the last 20 years.

Table 6: Descriptive statistics of the variables used in the research, after correcting for different purchasing powers. The first four variables are used in the first stage of the analyses, and the remaining seven variables are evaluated as potential drivers of efficiency in the second steps.

Variable	Mean	St. deviation	Min	Max
Employees	190,916	119,732	23,196	414,000
Fixed assets	22,106,992,902	17,559,363,683	305,476,356	57,287,315,933
Inventory	2,646,634,904	2,295,488,864	54,193,772	7,610,000,000
Turnover	48,612,144,208	35,157,686,764	708,166,968	103,071,428,571
Group	365	423	15	1,305
Debt to equity	3.22	2.22	0.57	11.37
$(D/E)^2$	15.05	28.90	0.33	129.28
International activity	0.61	0.50	0.00	1.00
Outlets	7,211	6,590	492	21,564
Top 5	0.28	0.46	0.00	1.00
Merge & Acquisition	0.83	0.38	0.00	1.00

In Figure 6, the count of countries in the data set is available. The United Kingdom (UK), a country formed by constituent countries such as England, is the most common country in the data set, followed by Germany, France and Russia. These four countries are, with Turkey, the most populated countries in Europe as of 2020 (Eurostat, 2022; Statista Research Department, 2022). Even if Russia's population is substantially larger than the other countries, Russian retailers are not as prevalent. Additionally, they are not within the top ten food retailers. Interestingly, the only Dutch food retailer (Koninklijke Ahold Delhaize N.V.) has the highest revenue in the dataset. However, the Schwarz Group (owner of Lidl and Kaufland) and Aldi have higher revenues (Deloitte, 2022). Four countries (Netherlands, UK, France and Germany) make up the top ten European food retailers. Out of the 44 European countries, only eight are represented in the dataset.

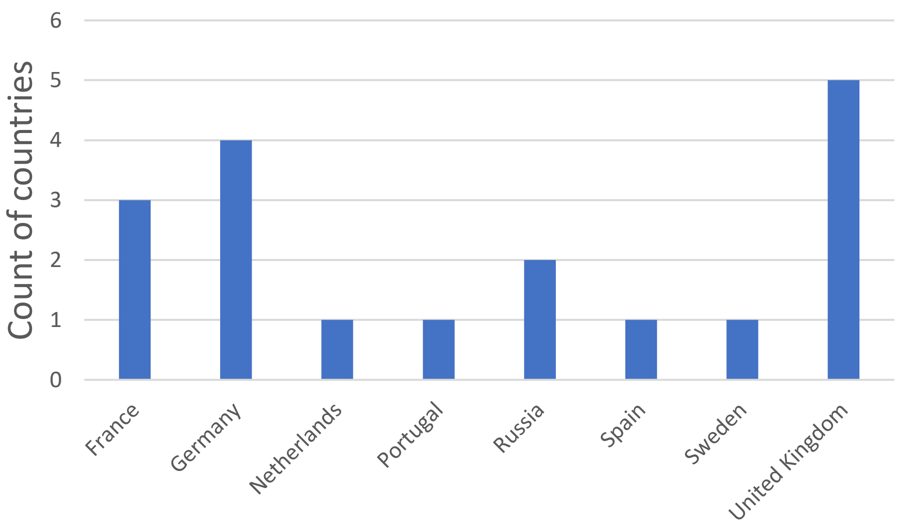

Figure 6: Count of countries in the experimental data set (Appendix 1).

5.2. Technical efficiency scores

5.2.1. Standard technical efficiency scores

The results of the first stage of the standard two-step analysis are available in Table 7. Eight out of 18 firms are efficient since their technical efficiency (and its reciprocal) scores are equal to one. The Portuguese retailer Jeronimo Martins (firm 15) reports the lowest efficiency score, 0.54. This score indicates that, using the same level of inputs, the firm is currently producing 54% of the technical output (turnover) that could be achieved. The firms that are most used as peers are Edeka and Tesco (8 times), which combined act as peers of all inefficient firms. It highlights their performance compared to other efficient firms, since they are used to project all the inefficient DMUs to the efficient frontier.

Table 7: Technical efficiency (TE) scores and its reciprocal of each DMU analysed (food retailers are ordered in terms of revenue, from highest to lowest). The DEA model is output-oriented and assumes VRS. The peer count column indicates how many times the corresponding company acts as a peer to another DMU.

Firm	Name	Reciprocal	TE	Peer count
1	Koninklijke Ahold Delhaize N.V.	1	1	2
2	Carrefour	1	1	2
3	Tesco PLC	1	1	8
4	Rewe combine	1	1	1
5	Lidl Stiftung & CO. KG	1.05	0.95	-
6	Edeka Zentrale Stiftung & CO.KG	1	1	8
7	Casino Guichard-Perrachon	1.69	0.59	-
8	Sainsbury (J) PLC	1.29	0.78	-
9	Auchan Retail	1.59	0.63	-
10	ASDA Stores Limited	1.61	0.62	-
11	Mercadona SA	1	1	3
12	Kaufland Stiftung & CO. KG	1.59	0.63	-
13	X5 Retail Group N.V.	1.59	0.63	-
14	WM Morrison Supermarkets Limited	1.61	0.62	-
15	Jeronimo Martins SGPS S.A.	1.85	0.54	-
16	Public Joint Stock Company Magnit	1	1	-
17	ICA Gruppen AB	1	1	2
18	The Co-operative Group Limited	1.22	0.82	-

The technical efficiency scores are represented in Figure 7. The first six firms (which are also the top six food retailers in term of revenue) are mostly technically efficient, meanwhile the remaining 12 firms have in general lower values, around 0.65. However, firms 11, 16 and 17 are exceptions since they are also technically efficient. The mean technical efficiency score is 0.82, and the standard deviation is 0.19.

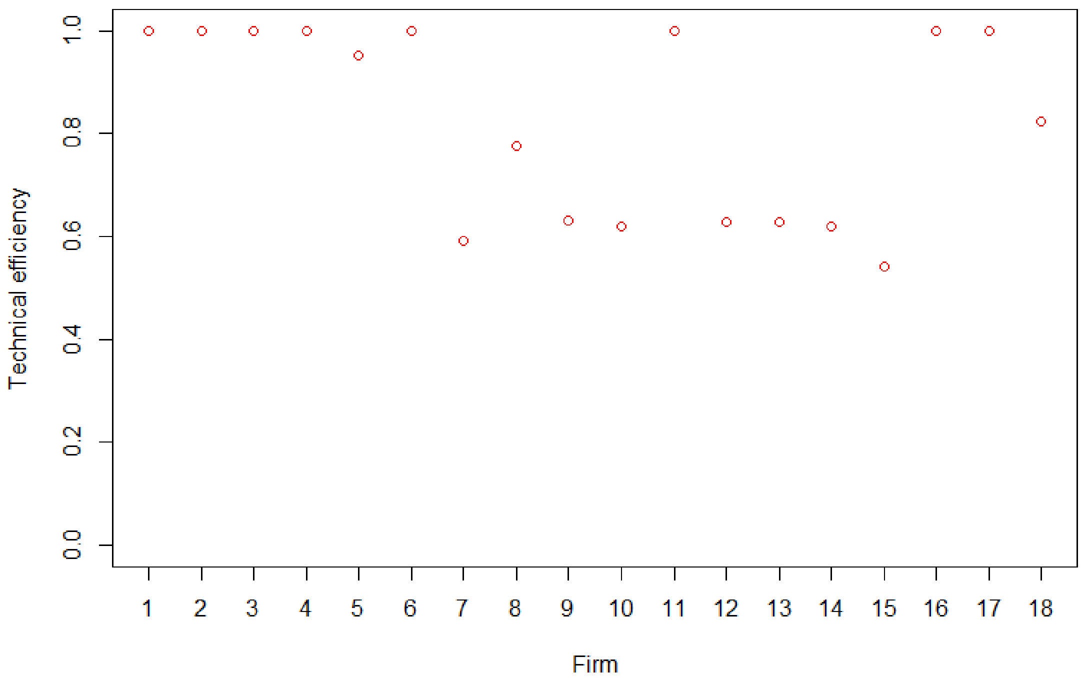

Figure 7: Representation of the technical efficiency scores per firm. The firm's numbers are the same as in Table 7.

5.2.2. Bias-corrected technical efficiency scores

The results of the first step using the double bootstrap procedure are available in Table 8. A general trend can be observed, where the scores are lower compared to the standard approach. However, firm 18 is an exception since the score reported by this approach is slightly higher (the bias term is negative). The firms that have the largest decrease in TE score are DMUs 3 and 17. In fact, both of them were stated as efficient in the standard approach. In this case, none of the firms in the data set have a TE equal to one. Only five firms (4, 5, 11, 16 and 18) have the value of one within their confidence interval. Just like in the previous approach, Jeronimo Martins obtained the lowest score, in this case 0.47. Since this procedure starts off with a standard DEA (section 3.1.5), the peer count available in Table 7 also applies to this approach.

Table 8: Bias-corrected technical efficiency scores, with lower (LB) and upper (UB) bound of the confidence interval, for each of the DMUs. Bias represents the difference between the standard technical efficiency score (Table 7) and the bias-corrected one.

Firm	Name	Robust-TE	LB robust TE	UB robust TE	Bias
1	Koninklijke Ahold Delhaize N.V.	0.85	0.76	0.90	0.15
2	Carrefour	0.91	0.84	0.98	0.09
3	Tesco PLC	0.78	0.71	0.83	0.22
4	Rewe combine	0.92	0.85	1.03	0.08
5	Lidl Stiftung & CO. KG	0.91	0.88	1.00	0.04
6	Edeka Zentrale Stiftung & CO.KG	0.90	0.82	1.00	0.10
7	Casino Guichard-Perrachon	0.54	0.52	0.57	0.05

8	Sainsbury (J) PLC	0.71	0.66	0.77	0.07
9	Auchan Retail	0.57	0.54	0.61	0.06
10	ASDA Stores Limited	0.55	0.52	0.59	0.07
11	Mercadona SA	0.91	0.83	1.05	0.09
12	Kaufland Stiftung & CO. KG	0.56	0.52	0.63	0.06
13	X5 Retail Group N.V.	0.56	0.52	0.63	0.06
14	WM Morrison Supermarkets Limited	0.58	0.52	0.68	0.04
15	Jeronimo Martins SGPS S.A.	0.47	0.44	0.52	0.07
16	Public Joint Stock Company Magnit	0.96	0.93	1.05	0.04
17	ICA Gruppen AB	0.68	0.64	0.71	0.32
18	The Co-operative Group Limited	0.85	0.73	1.17	-0.03

Figure 8 represents the robust TE scores (with their confidence interval) of all firms. In this approach, the average bias-corrected technical efficiency score is 0.73 and the standard deviation 0.17.

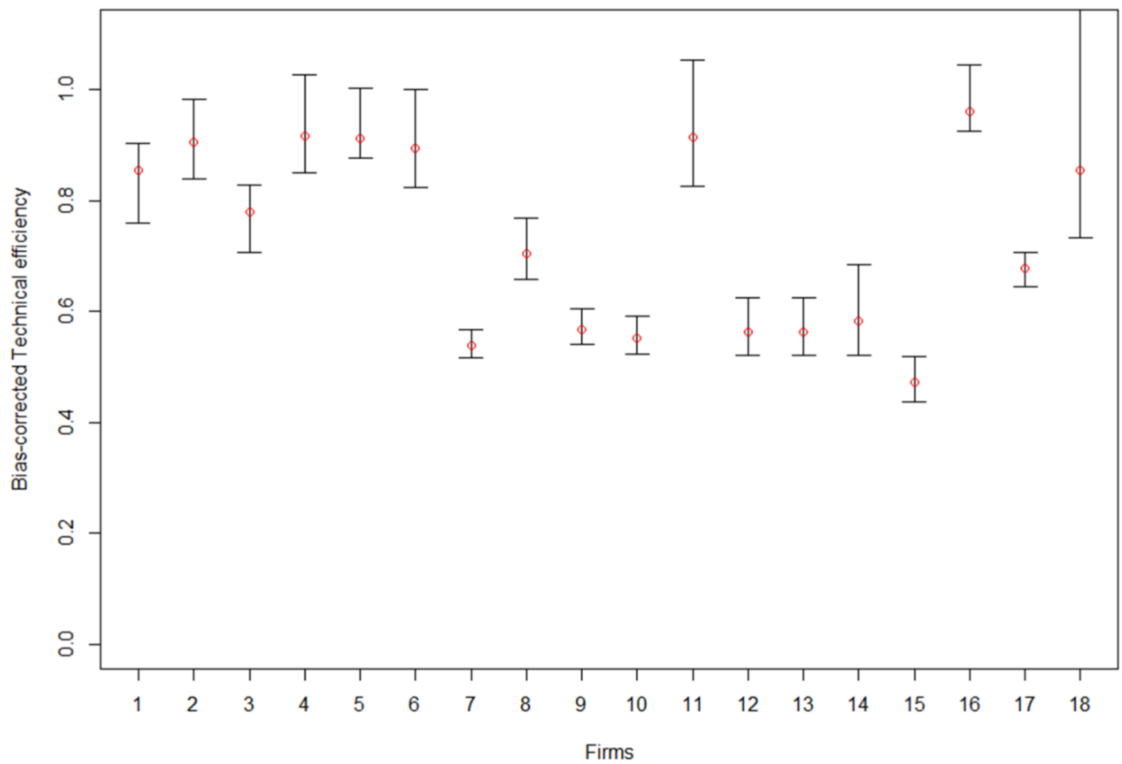

Figure 8: Representation of the robust technical efficiency scores with CI per firm. The firm's numbers are the same as in Table 8.

5.3. Effect of environmental variables on technical efficiency

5.3.1. Standard regression coefficients

In the standard two-step efficiency analysis, the effect of the chosen environmental variables on the technical efficiency was evaluated by using a standard OLS regression. The regression coefficients (β) obtained from the analysis are available in Table 9. Only one of the studied environmental variables is significantly correlated to TE with a level of significance (α) of 0.05. The number of outlets is negatively correlated with technical inefficiency (i.e., positively correlated with technical efficiency). Interestingly, being a top 5 food retailer in terms of revenue (used together with number of outlets to represent firm size) is positively correlated with technical efficiency if $\alpha = 0.1$ is considered. The R^2 value of the model is 0.64, and if adjusted it is 0.39. The F-statistic is 3.22, and its p-value is 0.046. Therefore, the model has significant predictive capabilities if $\alpha = 0.05$ is considered.

Table 9: Regression coefficients (β) of a standard OLS regression model to measure the effects of the environmental variables on the technical inefficiency (reciprocal of the technical efficiency score). (*) indicates that the coefficient is significant with $\alpha = 0.05$, and (**) with $\alpha = 0.1$.

Environmental variable	Regression coefficient	Standard deviation	t value	Pr(>|t|)
Intercept	0.88	0.26	3.40	0.01*
Group	0.0002	0.0002	1.28	0.23
International activity	0.07	0.17	0.44	0.67
D/E	0.22	0.14	1.63	0.13
$(D/E)^2$	-0.010	0.01	-1.08	0.31
M&A	-0.008	0.19	-0.04	0.97
Outlets	-0.00003	0.00001	-2.32	0.04*
Top5	-0.29	0.16	-1.83	0.10**

5.3.2. Robust regression coefficients

The second stage of the bootstrap approach provides the robust regression coefficients of the environmental variables studied, which are available in Table 10. In this approach, there are four environmental variables that have a significant effect on technical efficiency. Both international activity and debt-to-equity ratio are negatively correlated with technical efficiency. On the other hand, number of outlets and being a top 5 food retailer in terms of revenue have a positive effect on technical efficiency. The robust standard deviation of errors in the truncated regression of reciprocal of DEA scores on environmental variables ($\hat{\hat{\sigma}}$) is equal to 0.176 (0.169 – 0.287). Unfortunately, the package does not provide information on the F-statistic or the R^2 value.

Table 10: Robust regression coefficients (β) of truncated regression model to measure the effects of the environmental variables on the bias-corrected technical inefficiency. For each coefficient, the lower and upper bounds of the confidence interval at 95% confidence level are provided. (*) indicates that the coefficient is significant with $\alpha = 0.05$.

Parameter	Robust regression coefficient	Lower bound	Upper bound
Intercept	0.58	0.09	1.09
Group	0.0001	-0.0002	0.0004
International activity	0.39*	0.12	0.68
D/E	0.32*	0.10	0.55
D/E^2	-0.01	-0.03	0.003
M&A	0.22	-0.10	0.56
Outlets	-0.0001*	-0.00008	-0.00003
Top5	-0.45*	-0.78	-0.17

The robust regression coefficients are represented in Figure 9 to facilitate the interpretation of results. Outlets' regression coefficient is very close to zero, as well as its confidence interval. Consequently, it cannot be seen properly in the figure that it has a significant effect. There is a similar issue with the number of companies in the corporate group, but in this case the coefficient is not significant.

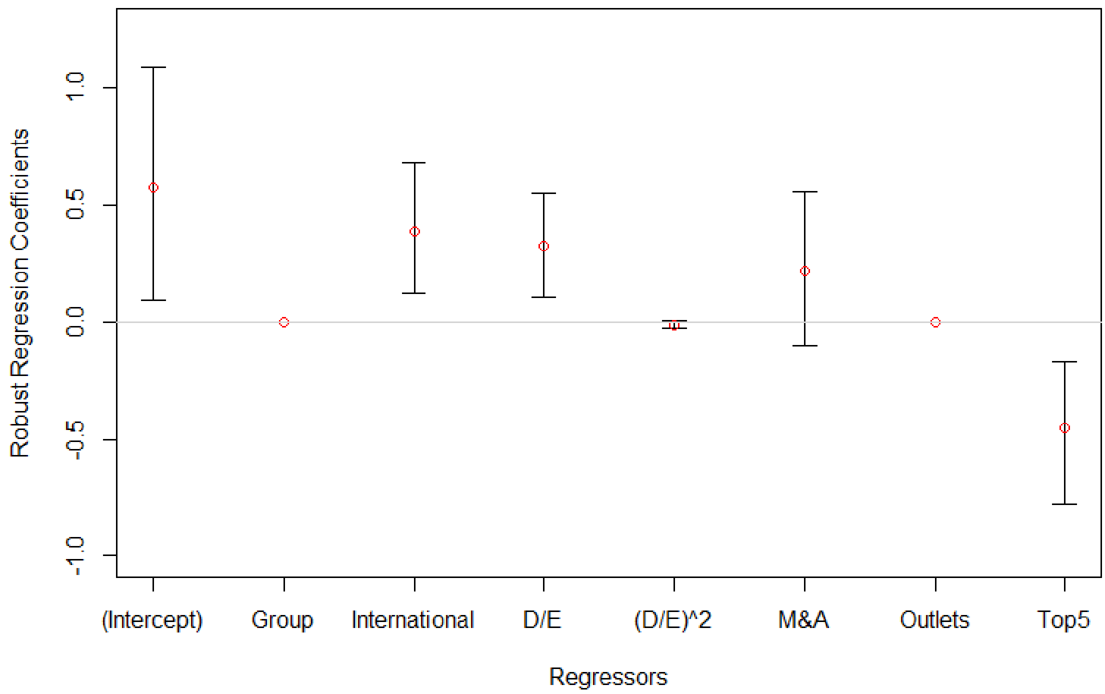

Figure 9: Representation of the robust regression coefficients with CI per environmental variable. The robust regression coefficients measure the effect of the corresponding environmental variable on the bias-corrected technical inefficiency.

5.4. Robustness check

The results of the first step for both the standard and the bootstrap approach are available in Table 11. Regarding the standard approach, the average TE score is 0.79, and its standard deviation 0.22. Comparing these values to the actual results (section 5.2.1.), the robustness check reports a lower average. Interestingly, in this check there are also 8 technically efficient firms, which are basically the same set of firms. In this analysis the bootstrap approach results on an average robust TE score of 0.65, and a standard deviation of 0.23. As in the standard approach, the average score obtained in this check is lower than the one obtained in the actual results (5.2.2.).

Table 11: Results of the first step in the robustness check, given by the standard and the bootstrap approach. The bootstrap approach reports the robust TE, and its confidence interval. Bias represents the difference between the standard and the bootstrap approach.

Firm	Name	Standard TE	Robust-TE	Confidence interval	Bias
1	Koninklijke Ahold Delhaize N.V.	1	0.98	0.96 – 1.03	0.02
2	Carrefour	1	0.97	0.93 – 1.02	0.03
3	Tesco PLC	1	0.98	0.96 – 1.02	0.02
4	Rewe combine	1	0.96	0.93 – 1.05	0.04
5	Lidl Stiftung & CO. KG	0.97	0.95	0.93 – 0.99	0.02
6	Edeka Zentrale Stiftung & CO.KG	1	0.67	0.65 – 0.70	0.33
7	Casino Guichard-Perrachon	0.59	0.52	0.50 – 0.53	0.08
8	Sainsbury (J) PLC	0.77	0.60	0.56 – 0.62	0.17
9	Auchan Retail	0.63	0.51	0.49 – 0.53	0.11
10	ASDA Stores Limited	0.62	0.48	0.44 – 0.51	0.14
11	Mercadona SA	1	0.68	0.65 – 0.70	0.32
12	Kaufland Stiftung & CO. KG	0.62	0.45	0.42 – 0.47	0.17
13	X5 Retail Group N.V.	0.47	0.36	0.35 – 0.37	0.11
14	WM Morrison Supermarkets Limited	0.63	0.48	0.41 – 0.50	0.15
15	Jeronimo Martins SGPS S.A.	0.54	0.38	0.37 – 0.39	0.16
16	Public Joint Stock Company Magnit	0.40	0.29	0.28 – 0.30	0.11
17	ICA Gruppen AB	1	0.70	0.66 – 0.73	0.30
18	The Co-operative Group Limited	1	0.73	0.68 – 0.77	0.27

In Table 12 the results of the second step of the efficiency analysis are available, both for the standard and the bootstrap approach. In this scenario, both the standard and the boostrap approach conclude that the only environmental variable with a significant effect on technical efficiency is being a top 5 food retailer.

Table 12: The regression coefficients and their t-values are obtained from the standard approach. (*) indicates that the coefficient is significant with $\alpha = 0.05$. Additionally, the robust regression coefficients, and their confidence interval are provided by the bootstrap approach.

Environmental variable	Regression coefficient	t value	Robust regression coefficient	Confidence interval
Intercept	0.84	2.04	0.97	-0.18 - 2.15
Group	0.000	0.21	0.000	-0.001 – 0.000
International activity	0.050	1.15	0.21	-0.47 – 0.92
D/E	0.25	-0.92	0.45	-0.09 – 1.01
D/E^2	-0.014	0.19	-0.03	-0.07 – 0.001
Outlets	0.00002	1.02	0.00004	-0.00001 – 0.00009
Top5	-0.62*	-2.48	-6.46*	-11.64 – (-2.15)
M&A	-0.09	-0.32	-0.30	-1.21 – 0.48

6. Discussion

The standard and the double bootstrap approach of the efficiency analysis using DEA were used in this research. As expected, the latter approach reported in general lower TE scores, since it accounts for the bias term. The standard approach does not consider sampling noise, which results in an overestimation of efficiency scores (Emrouznejad et al., 2022). In this study, the average TE of the food retailers decreases from 0.82 to 0.73 due to the bias amongst scores. Mostafa (2009) benchmarked retailers and food consumer stores in the USA and reported an average 0.80 TE score using the standard DEA approach. Cruz Roche et al. (2019) evaluated the efficiency of retail companies in 25 European countries between 2006 and 2015, in order to study country-specific factors. The average TE score was 0.82, and 0.77 when accounting for bias with the bootstrap procedure. Interestingly, Cruz Roche et al. (2019) did not include Russian companies, which are included in this research. Compared to the studies mentioned above, this research obtains a similar average TE score, around 0.8. However, in this study the bootstrap approach decreases the average TE score two times more than in the mentioned study, which could be explained by a high variance of error terms ($\hat{\sigma}_{\varepsilon}^2$). This could potentially be changed by adding more observations (firms) to the data set.

Since DEA is based on benchmarking, the TE scores obtained depend on what DMUs are included. Additionally, this research has a relatively low number of DMUs, so the average is more sensitive to inefficient firms. Therefore, comparing results of studies with different DMUs and/or year has a limited interpretation. The TE scores provide a good overview of the performance of the analysed firms during 2020, but the most interesting part for the relevant stakeholders (i.e., managers, policy makers and researchers) likely is the study of environmental variables.

The standard approach concludes that international activity is not a significant determinant of efficiency. On the other hand, the bootstrap approach identifies it as a significant environmental variable, that is negatively correlated with technical efficiency. Both approaches report a positive regression coefficient, but only the latter concludes that it is significant. In any case, both approaches conclude that the hypothesis (H_1) is rejected. Even if the described advantages may be true, more impactful advantages may arise from limiting the activity to one country. For example, it may be easier to coordinate operations within one country, and the bureaucratic work could be potentially reduced. There are some potential drawbacks from international expansion, such as increased pressure from a bigger list of stakeholders and a potentially hostile international environment (Attig et al., 2016). Additionally, operating in one country allows the firm to me more specialized, which can be key in a competitive market such as food retailing (Hirsch et al., 2021). Expanding into new countries is an investment that is not guaranteed to be successful, and multiple factors need to be taken into account in order to succeed. For example, Carrefour's expansions have sometimes succeeded (e.g. Latin America, Europe) and in other occasions failed (e.g. China, India) due to the complexity linked to internationalization (Zentes et al., 2017).

The top five retailers in terms of revenue can be considered as (almost) technically efficient in both approaches. As a result, the corresponding environmental variable is positively correlated with efficiency ($\alpha = 0.1$ in the standard approach, $\alpha = 0.05$ in the bootstrap approach). One way to explain it is that, for large corporate firms, the benefits provided by firm size such as economies of scale (reduced average unit cost because of large output level) are larger than drawbacks such as bureaucratic frictions and complex monitoring (Diaz and Sanchez, 2008). Additionally, the food supply chain is characterised by high retail concentration in multiple European countries (McCorriston, 2014). As a result, the buyer power of large retailers is increased, specially towards small and medium suppliers. Large food manufacturers that produce "must-have" products (such as large FMCG firms) may offset the retailer's buyer power (McCorriston, 2014). Inderst and Valletti (2011) describe the "Waterbed effect": a theory that suggests that more advantageous terms of trade for large and powerful buyers can result in worse terms for smaller buyers. As a result, there is an oligopoly of 3-6 dominant supermarket chains and a group of specialised fringe retailers that are price takers and adapt to niche markets. This situation was recently proven by Hirsch et al. (2021), since their study shows that the top five retailers have a significantly higher profit persistence, resulting in power imbalances within the food supply chain. Similar results have been found in other industries. For example, Sanchez-Robles et al. (2022) analysed the efficiency of European oil companies, where market concentration is also notable, and the size of the firms was also positively correlated to efficiency (specially for the largest firms in terms of turnover).

The number of outlets was also chosen to represent size of the firm. In this case, both approaches conclude with $\alpha = 0.05$ that the number of outlets has a significant effect on technical efficiency, particularly that there is a positive correlation. Barros (2006) identified this environmental variable as a positive determinant of efficiency in Portuguese food retailers. Gandhi and Shankar (2014) concluded, thanks to a two-step efficiency analysis, that the number of outlets was a positive driver of efficiency of Indian retailers, potentially because of economies of scale. More recently, Neves et al. (2018) concluded that number of outlets was not a significant driver of efficiency in Brazilian supermarkets. This different result could potentially be explained for instance by the different store formats. Interestingly, Neves et al. (2018) did conclude that store size was positively correlated with efficiency. This suggests that large outlets tend to be more efficient, which could also be related to economies of scale.

Additionally, online/non-store retailing is a trend that has recently been implemented by many retailers (Deloitte, 2022; Gauri et al., 2021), which could be a drawback for the use of this variable to measure firm size. Nevertheless, the results show that number of outlets has a positive effect on efficiency, explained by economies of scale. Since it is concluded that both number of outlets and being a top 5 food retailer are positively correlated with technical efficiency, H_2 is accepted.

Regarding the number of companies in the corporate group (H_3), both approaches consider it as non-significant. All food retailers can be considered as multiproduct firms, and therefore it may be difficult to capture a difference between intra and interfirm transactions. The advantages of corporate groups have been studied and proven in previous studies. For example, Perrigot and Barros (2008) concluded that belonging to a corporate group was positively correlated with efficiency in French food retailers. Since the firms that are considered in this analysis are the largest European retailers, it is possible that these advantages are not significant amongst these firms because they are all part of large corporate groups.

There is no significant correlation between financial leverage and technical efficiency according to the standard approach. However, the bootstrap approach indicates that there is a significant negative correlation between D/E and technical efficiency. Since the regression coefficient of the squared D/E ratio is not significantly different from 0 according to both approaches, there is no relevant quadratic trend in the model. In order for the hypothesis (H_3) to be actually observed, D/E should be negatively correlated with inefficiency, and the squared ratio positively correlated (as inefficiency is regressed, the hypothesized effect is actually a U shape). Since this is not concluded, the hypothesis is rejected (H_3). As a reminder, the trade-off theory states that firms maximize their value with a debt-to-equity ratio that trades off the advantages and disadvantages of debt (Myers, 1984). It is possible that the analysed corporate firms have found their corresponding unique optimal financial leverage, especially because of the competitive environment. Studies have obtained mixed results about the correlation between financial leverage and firm performance (Abu-Abbas et al., 2019). However, it has been proven in other industries and countries that the hypothesised inverted U-shape relationship between financial leverage and firm efficiency does exist, evidencing the relevant capital structure theories (for example Guo et al. (2021)). The bootstrap approach suggests that reducing financial leverage is positively correlated to technical efficiency. A possible explanation to this finding are high agency costs of debt (such as bankruptcy costs), that offset the tax shield and the agency costs of free cash flow.

According to both approaches, merge and acquisitions do not have a significant effect on technical efficiency, and therefore the hypothesis (H_5) is not fulfilled. Additionally, the standard approach reports a negative regression coefficient, but the robust regression coefficient reported by the bootstrap approach is positive. A positive correlation between M&A and efficiency could be explained by the advantages of market concentration. In fact, the largest food retailers, which benefit the most from economies of scale and buyer power, are consolidating their position in the market through mergers and acquisitions (McCorriston, 2014). These advantages would also support the synergy suggested by Arnold (2005), where both firms involved in the merger/acquisition benefit from it. Gandhi and Shankar (2014) identified merge and acquisitions as a positive driver of efficiency of Indian retailers, since it allowed new players, technologies and logistics to enter the Indian retail sector. On the other hand, merge and acquisitions are frequently financial failures, that have positive short-term consequences but also negative long-term impacts such as share price underperformance (Marks and Mirvis, 2011). Additionally, the Free Cash Flow theory (see section 3.2.3.) suggests that merge and acquisitions reduce the value of the firm, as the agency costs of free cash flow may increase. Regarding the dummy variable itself, it measures any kind of merge and acquisitions with two potential values, 0 and 1. In other words, all of these financial operations are considered the same, independently of their size and effect. Using a quantitative variable to represent these operations, such as a ratio that compares the sizes of the involved firms, would potentially capture the effect more appropriately.

Regarding the robustness check, where purchasing power amongst countries is assumed to be the same and therefore PPPs are not considered, it provides interesting insights. Focusing on the first step of the analysis (Table 11), in this scenario both approaches report a slightly lower average TE score compared to the actual analysis. Additionally, basically the same set of firms (1,2,3,4,6 and 17) are identified as technically efficient. Interestingly, both of the Russian retailers (firms 13 and 16) report drastically lower TE scores in this case. For example, firm 16 is identified as technically efficient (section 5.2.) but in the robustness check it is the firm with the lowest TE score. This change could be explained by the fact that Russia has the largest PPPs amongst the evaluated countries (Table 4). As a results, if PPPs are not considered, Russian retailers obtain lower TE scores and therefore the average TE score also decreases. Regarding the studied environmental variables, in this scenario both methods conclude that being a top 5 food retailer is a significant driver of efficiency. Therefore, it also shows that the largest food retailers currently have an advantageous position in the market. However, number of outlets (also used to represent firm size) does not have a significant effect on efficiency in this analysis. In the end, not considering PPPs result in less significant correlations being observed.

The comparison of both approaches shows that the double bootstrap procedure report less firms as technically efficient and a lower average TE score. Theoretically, the bias-corrected TE scores reported by the bootstrap approach are more accurate, since they are based on numerous samples (Emrouznejad et al., 2022). Regarding the second stage, both methods report that size (represented by number of outlets and being a top 5 food retailer) has a significant effect on technical efficiency. Despite of this similarity, the bootstrap approach also concludes that debt-to-equity ratio and international activity have significant negative effects on technical efficiency. Therefore, the results obtained from both methods agree on some aspects, but also differ on others.

Banker and Natarajan (2008) show that the standard efficiency analysis has a higher accuracy, and that it is more appropriate and flexible than the bootstrap approach. However, Simar and Wilson (2011) show that their double bootstrap approach is more accurate than the standard approach, contradicting the previous authors. These studies extensively discuss the advantages of their preferred method and the limitations of the other method. However, both are widely used by researchers nowadays. To sum up, the standard is characterised as simple, and its methodology and results are easier to understand. Meanwhile, the bootstrap approach provides confidence intervals and allows for better hypothesis testing thanks to higher complexity and specific assumptions. Therefore, since the methods have different theoretical backgrounds, running both can provide interesting insights and a more complete view to the stakeholders. For example, in this research the methods lead to different conclusions regarding two drivers of efficiency, which is something interesting to consider. Additionally, if both methods lead to the same conclusions, which is partially the case in this research, it proves a good internal validity. Importantly, if a large data set is used the running time of the bootstrap procedure drastically increases, in some cases taking days to complete. This limitation may limit its application on large data sets if results are needed rapidly.

7. Limitations

Firstly, several of the considered corporate firms have other activities that are not related to food retailing, such as gas stations and drugstores. This is especially common in the largest corporate groups. When it was possible, the data gathered was exclusively related to retailing, but in some cases this was be ensured. Additionally, some of these supermarket chains also sell non-food products to consumers, such as cleaning products and clothes. Retailers have offered these products since the 20th century as mentioned in section 2, and it is practically impossible to exclude them when analysing food retailers. Nevertheless, the main activity and source of income of these companies is food retailing.

Secondly, not all the relevant food retailers could be included in the data set. For example, Aldi (one of the largest food retailers in the world) was not included because its financial information could not be gathered from a reliable source. A similar issue came up with Italian supermarket chains, such as Conad. Due to this limitation, the final data set does not represent completely the largest European food retailers. Additionally, because of the complexity of gathering the data and the different sources, the size of the experimental data set is limited to 18 firms, which may not be enough to represent the European food retailing sector.

Additionally, only eight out of the 44 European countries are considered in the dataset. Even if the firms included operate in the remaining countries, this means that most countries are not represented in the data set. Due to this, the representability and external validity of the data set is limited. Besides this limitation, two of the eight countries are not part of the EU (United Kingdom and Russia). As a result, countries with different economic framework are being compared, and this may have an impact in the results obtained.

Regarding the data analysis, the standard two-step procedure can be exactly replicated if the same experimental data set is used. However, this is not the case for the double bootstrap procedure. Since both loops included in the algorithm sample random error terms from normal distributions, the exact results cannot be obtained. However, this limitation does not have a high impact as the confidence intervals of bias-corrected TE scores and robust regression coefficients slightly change.

8. Conclusions

Firstly, the individual technical efficiency scores were obtained for the studied European food retailers, evaluating their performance during 2020 and considering the corresponding PPPs. The average standard TE score was 0.82, and it decreased to 0.73 when bias was considered. The latter implies that, on average, the largest European food retailers generated during 2020 73% of their maximum turnover for the corresponding level of input. Other studies regarding retailers have obtained similar average TE scores. These efficiency scores constitute a valuable snapshot of European food retailers, by benchmarking the different firms. The robustness check, where PPPs are not considered, showed a lower average TE score, especially for high-PPPs countries. In any case, it comes to show that there is room for improvement for the European food retailing sector in terms of financial efficiency.

Secondly, a regression model was used to study the effect of environmental variables on the technical efficiency scores. Both the standard and the bootstrap approach conclude that being in the top five in terms of revenue is a significant environmental variable, and therefore a determinant of efficiency. It is discussed that the largest food retailers have an advantageous position within the food supply chain, characterised by high buyer power and access to better deals with their suppliers. Consequently, it is possible that smaller food retailers are forced to accept worse deals from the same suppliers, resulting in a Waterbed effect. Additionally, both approaches also conclude that number of outlets is also positively correlated with technical efficiency. Large food retailers may also be benefited from economies of scale and do not seem to be constrained by bureaucracy and complex management. Since both variables show a significant positive effect, their corresponding hypothesis regarding firm size (H_2) was accepted. The remaining hypotheses regarding the other environmental variables were rejected. Additionally, the robustness check reported that the only significant driver of efficiency is being a top five food retailer, which is positively correlated to efficiency.

Thirdly, the bootstrap and the standard approach of the two-stage efficiency analysis using DEA were compared during the discussion of the results. As stated before, the efficiency scores reported by each method are slightly different. The insights gained in the second stage of the analysis are alike (in terms of the effect of firm size, merge and acquisitions and number of companies in the corporate group) but also different regarding international activity and debt-to-equity ratio. In addition to verifying a good internal validity, it comes to show that, even if the methods have different theoretical backgrounds, it may be a good idea to run both since they are not mutually exclusive. Applying both approaches results in a more complete understanding of the data set, and can lead to interesting discussion points that are useful for the stakeholders. In small data sets, like in this study, the bootstrap method only takes seconds to run. However, it can take multiple hours and days to complete when a large data set is analysed.

The main goal of this research, which is completed as a result of the three specific research objectives, was to analyse what characteristics condition the efficiency of European food retailers by benchmarking their economic performance using Data Envelopment Analysis. This study shows that firm size, due to economies of scale and buyer power, heavily conditions the efficiency of the European food retailers. Therefore, the largerst retailers are benefited in the current situation. In the recent years, the largest food retailers have kept growing rapidly, increasing the difference between them and smaller firms. On contrast, this situation may not be desirable for policy makers and smaller retailers. The current food retail market puts smaller firms in a disadvantageous position, where they have to deal with worse conditions and deals that result in less efficiency. It cannot be firmly concluded that debt-to-equity ratio and international activity have a negative correlation with efficiency but since both approaches differ on the significance of these variables.

This study may provide managers of food retail firms with a valuable description of the state of the industry in 2020. Defining precisely the context where their business operates can be useful to maximize its performance and efficiency. Managers may particularly benefit from the obtained technical efficiency scores if their firm was included in the analysis, because in that case it has been benchmarked with competing firms. Nevertheless, this study describes in detail how to conduct an efficiency analysis, and, depending on the interests of the manager, its methodology can be used with a different set of food retailers. Regarding the studied environmental variables, firm size has been demonstrated to have a significant effect on technical efficiency. Large corporate firm managers may want to keep exploiting this advantage, in order to maintain efficiency. On the other hand, managers of small corporate food retailers might prefer to work with suppliers that help them exploit economies of scale or that offer a unique advantage. Additionally, firm managers may want to pay close attention to financial leverage and international activity since they could have a significant effect on efficiency.

Lastly, there are several recommendations for future studies that aim to develop the insights gained in this research. Firstly, a possible improvement of this research is to incorporate a different reliable source of information, so that it is possible to include missing food retailers. Secondly, including at least one food retailer headquartered in each country would improve the external validity of the study, since all the countries that are part of the scope would be represented. Additionally, an analysis that considers only the EU would ensure a more similar economic framework amongst the firms and therefore increase the validity of benchmarking. Increasing the number of firms included could have captured new or different significant effects that are not reported in this analysis, potentially leading to new conclusions and insights. Lastly, some of the environmental variables (such as international activity and merge and acquisitions) used can be substituted by quantitative variables that may capture the effect more efficiently.

9. References

Abu-Abbas, B. M., Alhmoud, T., & Algazo, F. A. (2019). Financial leverage and firm performance evidence from Amman stock exchange. *The European Journal of Comparative Economics*, *16*(2), 207–237.

Arnold, G. (2005). *Handbook of corporate finance: A business companion to financial markets, decisions & techniquе*nova, T. (2007). The Problem of Efficiency Measurement and Its Solutions. *Riga Technical University 53rd International Scientific Conference*, 97.

Attig, N., Boubakri, N., El Ghoul, S., & Guedhami, O. (2016). Firm Internationalization and Corporate Social Responsibility. *Journal of Business Ethics*, *134*(2), 171–197. https://doi.org/10.1007/s10551-014-2410-6

Babatunde, Y., & Ahmed, M. (2015). Corporate Social Responsibility and Company Performance. *Journal of Business Studies Quarterly*, *7*(1).

Baffes, J. (1991). Some Further Evidence on the Law of One Price: The Law of One Price Still Holds. *American Journal of Agricultural Economics*, *73*(4), 1264–1273. https://doi.org/10.2307/1242454

Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis. *Management Science*, *30*(9), 1078–1092. https://doi.org/10.1287/mnsc.30.9.1078

Banker, R. D., & Natarajan, R. (2008). Evaluating Contextual Variables Affecting Productivity Using Data Envelopment Analysis. *Operations Research*, *56*(1), 48–58. https://doi.org/10.1287/opre.1070.0460

Barros, C. (2006). Efficiency measurement among hypermarkets and supermarkets and the identification of the efficiency drivers. *International Journal of Retail & Distribution Management*, *34*(2), 135–154. https://doi.org/10.1108/09590550610649795

Brynjolfsson, E., Hu, Y. (Jeffrey), & Rahman, M. S. (2009). Battle of the Retail Channels: How Product Selection and Geography Drive Cross-Channel Competition. *Management Science*, *55*(11), 1755–1765. https://doi.org/10.1287/mnsc.1090.1062

Burt, S. (2010). Retailing in Europe: 20 years on. *The International Review of Retail, Distribution and Consumer Research*, *20*(1), 9–27. https://doi.org/10.1080/09593960903497773

Cambridge Dictionary. (2022). *Retail - English meaning*. https://dictionary.cambridge.org/dictionary/english/retail

Canback, S., Samouel, P., & Price, D. (2006). Do Diseconomies of Scale Impact Firm Size and Performance? A Theoretical and Empirical Overview. *Icfai Journal of Managerial Economics*, *4*(1), 27–70.

Castro, V. F., & Frazzon, E. M. (2017). Benchmarking of best practices: an overview of the academic literature. *Benchmarking: An International Journal*, *24*(3), 750–774. https://doi.org/10.1108/BIJ-03-2016-0031

Çerkezi, M. S. A. (2013). A literature review of the trade−off theory of capital structure. *ILIRIA International Review*, *3*(1), 125. https://doi.org/10.21113/iir.v3i1.103

Charnes, A., Cooper, W. W., Lewin, A. Y., & Seiford, L. M. (1994). Basic DEA Models. *Data Envelopment Analysis: Theory, Methodology, and Applications*, 23–47. https://doi.org/10.1007/978-94-011-0637-5_2

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, *2*(6), 429–444. https://doi.org/10.1016/0377-2217(78)90138-8

Coelli, T. J., Prasada Rao, D. S., O'Donell, C. J., & Battese, G. E. (2005). An introduction to efficiency and productivity analysis, Second edition. *Springer* .

Cooper, W. W., Seiford, L. M., & Zhu, J. (2011). *Handbook on Data Envelopment Analysis* (2nd ed.). Springer. https://doi.org/10.1007/978-1-4419-6151-8

Corelli, A. (2018). *Capital Structure* (pp. 223–255). https://doi.org/10.1007/978-3-319-95762-3_7

Cox, T. L., & Wohlgenant, M. K. (1986). Prices and Quality Effects in Cross-Sectional Demand Analysis. *American Journal of Agricultural Economics*, *68*(4), 908–919. https://doi.org/10.2307/1242137

Cruz Roche, I., Romero, J., & Sellers-Rubio, R. (2019). Retail services efficiency: impact of country-specific factors. *International Journal of Retail & Distribution Management*, *47*(8), 774–792. https://doi.org/10.1108/IJRDM-12-2018-0275

Deloitte. (2022). *Global Powers of Retailing Report*. https://www2.deloitte.com/content/dam/Deloitte/at/Documents/consumer-business/at-global-powers-retailing-2022.pdf

Diaz, M. A., & Sanchez, R. (2008). Firm size and productivity in Spain: a stochastic frontier analysis. *Small Business Economics*, *30*(3), 315–323. https://doi.org/10.1007/s11187-007-9058-x

Emrouznejad, A., Yang, G., Khoveyni, M., & Michali, M. (2022). Data Envelopment Analysis: Recent Developments and Challenges. In *The Palgrave Handbook of Operations Research* (pp. 307–350). Springer International Publishing. https://doi.org/10.1007/978-3-030-96935-6_10

Eurocommerce, & McKinsey & Company. (2022). *Navigating the market headwinds: The state of grocery retail 2022 - Europe*.

Eurostat. (2018). *Glossary: Purchasing power parities*. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:Purchasing_power_parities_(PPPs)

Eurostat. (2022). *Statistics: Population on 1 January*. https://ec.europa.eu/eurostat/databrowser/view/TPS00001/bookmark/table?lang=en&bookmarkId=c0aa2b16-607c-4429-abb3-a4c8d74f7d1e

Farrell, M. J. (1957). The Measurement of Productive Efficiency. *Journal of the Royal Statistical Society. Series A (General)*, *120*(3), 253. https://doi.org/10.2307/2343100

Gandhi, A., & Shankar, R. (2014). Efficiency measurement of Indian retailers using Data Envelopment Analysis. *International Journal of Retail & Distribution Management*, *42*(6), 500–520. https://doi.org/10.1108/IJRDM-10-2012-0094

Gauri, D. K., Jindal, R. P., Ratchford, B., Fox, E., Bhatnagar, A., Pandey, A., Navallo, J. R., Fogarty, J., Carr, S., & Howerton, E. (2021). Evolution of retail formats: Past, present, and future. *Journal of Retailing*, *97*(1), 42–61. https://doi.org/10.1016/j.jretai.2020.11.002

Ge, D., Pan, Y., Shen, Z.-J., Wu, D., Yuan, R., & Zhang, C. (2019). Retail supply chain management: a review of theories and practices. *Journal of Data, Information and Management*, *1*(1–2), 45–64. https://doi.org/10.1007/s42488-019-00004-z

Guo, H., Legesse, T. S., Tang, J., & Wu, Z. (2021). Financial leverage and firm efficiency: the mediating role of cash holding. *Applied Economics*, *53*(18), 2108–2124. https://doi.org/10.1080/00036846.2020.1855317

Hamilton, S. F., Liaukonyte, J., & Richards, T. J. (2020). Pricing Strategies of Food Retailers. *Annual Review of Resource Economics*, *12*(1), 87–110. https://doi.org/10.1146/annurev-resource-101619-094219

Hanousek, J., Kočenda, E., & Shamshur, A. (2015). Corporate efficiency in Europe. *Journal of Corporate Finance*, *32*, 24–40. https://doi.org/10.1016/j.jcorpfin.2015.03.003

Hirsch, S., Lanter, D., & Finger, R. (2021). Profitability and profit persistence in EU food retailing: Differences between top competitors and fringe firms. *Agribusiness*, *37*(2), 235–263. https://doi.org/10.1002/agr.21654

Inderst, R., & Valletti, T. M. (2011). Buyer power and the "Waterbed effect." *The Journal of Industrial Economics*, *59*(1), 1–20. https://doi.org/10.1111/j.1467-6451.2011.00443.x

Jensen, M. C. (1986). Agency Costs of Free Cash Flow, Corporate Finance, and Takeovers. *The American Economic Review*, *76*(2).

Jensen, M. C. (1999). Agency Cost Of Free Cash Flow, Corporate Finance, and Takeovers. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.99580

Junius, K. (1997). Economies of Scale: A Survey of the Empirical Literature. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.8713

Kalirajan, K. P., & Shand, R. T. (1999). Frontier Production Functions and Technical Efficiency Measures. *Journal of Economic Surveys*, *13*(2), 149–172. https://doi.org/10.1111/1467-6419.00080

Kamilaris, A., Fonts, A., & Prenafeta-Boldύ, F. X. (2019). The rise of blockchain technology in agriculture and food supply chains. *Trends in Food Science & Technology*, *91*, 640–652. https://doi.org/10.1016/j.tifs.2019.07.034

Kaynak, E., & Cavusgil, S. T. (1982). The evolution of food retailing systems: Contrasting the experience of developed and developing countries. *Journal of the Academy of Marketing Science*, *10*(3), 249–268. https://doi.org/10.1007/BF02729966

Konuk, F. (2018). Financial and performance analysis of food companies: Application of Topsis and DEA. *MANAS Journal of Social Studies*, *7*(3).

Kumbhakar, S. C., Parmeter, C. F., & Zelenyuk, V. (2021). Stochastic Frontier Analysis: Foundations and Advances I. In *Handbook of Production Economics* (pp. 1–40). Springer Singapore. https://doi.org/10.1007/978-981-10-3450-3_9-2

Kuosmanen, T., Cherchye, L., & Sipiläinen, T. (2006). The law of one price in data envelopment analysis: Restricting weight flexibility across firms. *European Journal of Operational Research*, *170*(3), 735–757. https://doi.org/10.1016/j.ejor.2004.07.063

Kuosmanen, T., & Kortelainen, M. (2012). Stochastic non-smooth envelopment of data: semi-parametric frontier estimation subject to shape constraints. *Journal of Productivity Analysis*, *38*(1), 11–28. https://doi.org/10.1007/s11123-010-0201-3

Lampe, H. W., & Hilgers, D. (2015). Trajectories of efficiency measurement: A bibliometric analysis of DEA and SFA. *European Journal of Operational Research*, *240*(1), 1–21. https://doi.org/10.1016/j.ejor.2014.04.041

Lindsay, C. M., & Maloney, M. T. (1996). Do rate and volume matter? Transaction cost limits to economies of scale. *Studies in Economics and Finance*, *17*(1), 88–109. https://doi.org/10.1108/eb028728

Macfadyen, S., Tylianakis, J. M., Letourneau, D. K., Benton, T. G., Tittonell, P., Perring, M. P., Gómez-Creutzberg, C., Báldi, A., Holland, J. M., Broadhurst, L., Okabe, K., Renwick, A. R., Gemmill-Herren, B., & Smith, H. G. (2015). The role of food retailers in improving resilience in global food supply. *Global Food Security*, *7*, 1–8. https://doi.org/10.1016/j.gfs.2016.01.001

Maddala, G. S. (1983). *Limited-dependent and qualitative variables in econometrics*. Cambridge University Press. https://doi.org/10.1017/CBO9780511810176

Marks, M. L., & Mirvis, P. H. (2011). Merge Ahead: A Research Agenda to Increase Merger and Acquisition Success. *Journal of Business and Psychology*, *26*(2), 161–168. https://doi.org/10.1007/s10869-011-9219-4

Martić, M., Novaković, M., & Baggia, A. (2009). Data Envelopment Analysis - Basic Models and their Utilization. *Organizacija*, *42*(2), 37–43. https://doi.org/10.2478/v10051-009-0001-6

McCorriston, S. (2014). *Background Note. Competition issues in the food chain industry, OECD Directorate for Financial and Enterprise Affairs.* 9–46.

Mehdiloo, M., & Podinovski, V. V. (2019). Selective strong and weak disposability in efficiency analysis. *European Journal of Operational Research*. https://doi.org/10.1016/j.ejor.2019.01.064

Miljkovic, D. (1999). The Law of One Price in International Trade: A Critical Review. *Applied Economic Perspectives and Policy*, *21*(1), 126–139. https://doi.org/10.2307/1349976

Milone, L. M. (1986). The Law of One Price: further empirical evidence concerning Italy and the United Kingdom. *Applied Economics*, *18*(6), 645–661. https://doi.org/10.1080/00036848600000083

Modigliani, F., & Miller, M. H. (1958). The Cost of Capital, Corporation Finance and the Theory of Investment. *The American Economic Review*, *58*(3), 261–297.

Mostafa, M. M. (2009). Benchmarking the US specialty retailers and food consumer stores using data envelopment analysis. *International Journal of Retail & Distribution Management*, *37*(8), 661–679. https://doi.org/10.1108/09590550910966178

Murillo-Zamorano, L. R. (2004). Economic Efficiency and Frontier Techniques. *Journal of Economic Surveys*, *18*(1), 33–77. https://doi.org/10.1111/j.1467-6419.2004.00215.x

Myers, S. C. (1984). The Capital Structure Puzzle. *The Journal of Finance*, *39*(3), 574–592. https://doi.org/10.1111/j.1540-6261.1984.tb03646.x

Neves, F. L., Sampaio, R. M. B., & Sampaio, L. M. B. (2018). Efficiency, productivity gains, and the size of Brazilian supermarkets. *International Journal of Production Economics*, *197*, 99–111. https://doi.org/10.1016/j.ijpe.2017.12.016

Olesen, O. B., & Petersen, N. C. (2016). Stochastic Data Envelopment Analysis—A review. *European Journal of Operational Research*, *251*(1), 2–21. https://doi.org/10.1016/j.ejor.2015.07.058

Organisation for Economic Co-operation and Development. (2023). *Purchasing Power Parities: 2020*. https://data.oecd.org/conversion/purchasing-power-parities-ppp.htm

Perrigot, R., & Barros, C. P. (2008). Technical efficiency of French retailers. *Journal of Retailing and Consumer Services*, *15*(4), 296–305. https://doi.org/10.1016/j.jretconser.2007.06.003

Pinckaers, M., & Phillips, S. (2019). *The Dutch Food Retail Report 2019*.

Rittenberg, L., & Tregarthen, T. (2012). *Economics Principles*.

Robinson, J. (1956). *The Accumulation of Capital*.

Sanchez-Robles, B., Herrador-Alcaide, T. C., & Hernández-Solís, M. (2022). Efficiency of European oil companies: an empirical analysis. *Energy Efficiency*, *15*(8), 63. https://doi.org/10.1007/s12053-022-10069-2

Sial, M., Zheng, C., Khuong, N., Khan, T., & Usman, M. (2018). Does Firm Performance Influence Corporate Social Responsibility Reporting of Chinese Listed Companies? *Sustainability*, *10*(7), 2217. https://doi.org/10.3390/su10072217

Simar, L. (1992). Estimating Efficiencies from Frontier Models with Panel Data: A Comparison of Parametric, Non-Parametric and Semi-Parametric Methods with Bootstrapping. *International Applications of Productivity and Efficiency Analysis*, 167–199. https://doi.org/10.1007/978-94-017-1923-0_11

Simar, L., & Wilson, P. W. (2007). Estimation and inference in two-stage, semi-parametric models of production processes. *Journal of Econometrics*, *136*(1), 31–64. https://doi.org/10.1016/j.jeconom.2005.07.009

Simar, L., & Wilson, P. W. (2008). Statistical inference in nonparametric frontier models: recent developments and perspectives. *The Measurement of Productive Efficiency and Productivity Growth*, 421–521.

Simar, L., & Wilson, P. W. (2011). Two-stage DEA: caveat emptor. *Journal of Productivity Analysis*, *36*(2), 205–218. https://doi.org/10.1007/s11123-011-0230-6

Simm, J., & Besstremyannaya, G. (2020). *rDEA: Robust Data Envelopment Analysis (DEA) for R*. https://github.com/jaak-s/rDEA

Singh, K., & Misra, M. (2021). Linking Corporate Social Responsibility (CSR) and Organizational Performance: the moderating effect of corporate reputation. *European Research on Management and Business Economics*, *27*(1), 100139. https://doi.org/10.1016/j.iedeen.2020.100139

Stanton, J. L. (2018). A brief history of food retail. *British Food Journal*, *120*(1), 172–180. https://doi.org/10.1108/BFJ-01-2017-0033

Statista Research Department. (2022). *Total resident population of Russia from 1959 to 2022*. https://www.statista.com/statistics/1009271/population-size-russia/

Stigler, G. J. (1958). The Economies of Scale. *The Journal of Law and Economics*, *1*, 54–71. https://doi.org/10.1086/466541

Teece, D. J. (1982). Towards an economic theory of the multiproduct firm. *Journal of Economic Behavior & Organization*, *3*(1), 39–63. https://doi.org/10.1016/0167-2681(82)90003-8

Verhoef, P. C. (2021). Omni-channel retailing: some reflections. *Journal of Strategic Marketing*, *29*(7), 608–616. https://doi.org/10.1080/0965254X.2021.1892163

Wei, Q. (2001). Data envelopment analysis. *Chinese Science Bulletin*, *46*(16), 1321–1332. https://doi.org/10.1007/BF03183382

Yafeh, Y. (2003). An international perspective of corporate groups and their prospects. In Structural impediments to growth in Japan. *University of Chicago Press.*, 259–284.

Yu, W., & Ramanathan, R. (2009). An assessment of operational efficiency of retail firms in China. *Journal of Retailing and Consumer Services*, *16*(2), 109–122. https://doi.org/10.1016/j.jretconser.2008.11.009

Zentes, J., Morschett, D., & Schramm-Klein, H. (2017). The Internationalisation of Retailing. In *Strategic Retail Management* (pp. 163–184). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-10183-1_8

Appendix 1: R studio code for data analysis

App. 1.1. R code of the standard two-step analysis

```
#Description of the data

summary(data)


#Determine the inputs (X) and outputs (Y) of the analysis

X = data[c('employees', 'fixed_assets', 'inventory')]

Y = data[c('turnover')]


#Standard DEA

firms=1:18

standard <-dea(XREF=X, YREF=Y,X=X[firms,],Y=Y[firms,], model="output", RTS="variable")

standard


#The results obtained are copied to an excel filed, named “results_rdea”, which is then imported


#Regression standard DEA

rall <- lm(results_rdea$Reciprocal~ data$group + data$debt_to_equity + data$desquared + data$international + data$outlets + data$top5 + data$`m&a`)

summary(rall)
```

App. 1.2. R code of the double bootstrap procedure

```
#Description of the data

summary(data)


#Determine the inputs (X), outputs (Y) and environmental variables (Z) of the analysis

X = data[c('employees', 'fixed_assets', 'inventory')]

Y = data[c('turnover')]

Z = data[c('group', 'international', 'debt_to_equity','desquared', 'outlets', 'top5', 'm&a')]


#Bias-corrected DEA with environmental variables

bev <-dea.env.robust (X, Y, W=NULL,Z, model="output", RTS="variable", L1=100, L2=2000, alpha=0.05)

bev
```