Dipartimento di Ingegneria informatica automatica e gestionale Antonio Ruberti

SAPIENZA
Università di Roma

A robust benchmarking on direct margin of the Italian energy retail market

Simone Di Leo
Marta Chicca
Cinzia Daraio
Andrea Guerrini
Stefano Scarcella

A robust benchmarking on direct margin of the Italian energy retail market

Simone Di Leo[a], Marta Chicca[b], Cinzia Daraio[a], Andrea Guerrini[b], Stefano Scarcella[b]

[a]Department of Computer, Control and Management Engineering, Sapienza University of Rome, Via Ariosto 25, 00185 Roma - Italy

[b]Regulatory Authority for Energy, Networks and Environment (ARERA), Piazza Cavour 5, 20121 Milan - Italy

Abstract

Liberalization and deregulation have fundamentally changed the energy market in Europe over the last forty years. In Italy, the retail electricity and gas markets were liberalized in the early 2000's. In 2016 retail activities were separated from other market activities. Italy in this sense is the only European country, together with Portugal, to have completely separated the retail segment from the rest of the industry. This peculiarity of the Italian market makes it interesting to analyze which operational levers or characteristics could bring advantages to Italian retail operators in achieving high margins. To do so, we propose two different models for evaluating *cost* and *commercial* efficiencies in achieving a high *direct* margin. We analyzed 120 retail operators in 2020 in the Italian retail market. By comparing the models, we identified that the groups of operators with higher revenue volumes can balance the two efficiencies and are more efficient than the others. The results of the analysis show an effect of size (in terms of revenue) on efficiency in achieving high consumer margins. The creation of demand aggregators for small operators is suggested to improve small and medium retailers' efficiency.

1 Introduction

Liberalization and deregulation have dramatically changed the energy and natural gas market in Europe over the past four decades. The liberalization process of the natural gas and energy market in Europe began in the late 1990s with the electricity (European Parliament and Council, 1996) and natural gas directives (European Parliament and Council, 1998). The liberalization process of the energy and gas market, and in particular the liberalization of the retail market, had multiple objectives. The main objectives were the stimulation of a lower final price given by competitive conditions and better quality and/or more innovation to customers. Various authors have criticized this view in the past considering the low effects that liberalization has brought. Defeuilley (2009) criticizes the view of the market as a process driven by entrepreneurship. He identifies the complexity of the determinants of choice (perceptions and decision-making protocols) and the technological paradigm of the electricity sector as two underestimated aspects of the retail market liberalization process. Earlier studies found little or no effect on prices in the liberalized markets (for a review, see Joskow, 2008), or even an increase in prices for small customers. This conclusion was also underlined by the ECME Consortium (2010), which found that where competition is allowed, significant benefits are available, but most households do not take full advantage and the generation portfolio is the main determinant of household price levels. Despite the criticism of market liberalization in the literature, the process of liberalization in various European countries (and beyond) has progressed.

In Italy, the retail electricity and gas markets were liberalised by Decree-Law No 79/1999 (also known as 'Decreto Bersani', Italian Parliament, 1999) for the electricity market and Decree-Law No 164/2000 (also known as 'Decreto Letta', Ministry of Economic Development, 2000) for the gas market. The liberalisation process has led to a clear division of the market into different industrial levels. This has allowed the creation of operators with a direct link to final customers, whose main activity is the purchase and sale of energy and natural gas, but not its distribution or production, the so-called retail market. In 2016, an important milestone was reached in terms of accounting and brand unbundling (ARERA, 2015) with the separation of retail activities from other market activities. Italy in this sense is the only European nation, together with Portugal, to have completely separated the retail segment from the rest of the industry. Despite this, it is worth noting that the retail market for domestic customers continues to coexist with two markets, a free one and a regulated

one called "maggior tutela" (scheduled to be closed in 2024). The Italian market has become very competitive in the retail sector over the years. According to the Italia Regulatory Authority for Energy Networks and Environment (ARERA) data, there are about 600 energy retailers in Italy (https://www.arera.it/it/dati/mr/mree_ccomm.htm last accessed 16/5/2023) and about 480 natural gas sellers (https://www.arera.it/it/dati/monitoraggio_retail.htm#consistenza last accessed 16/5/2023). Furthermore, from January 2020 more than 50% of the active domestic selling points are in the free market (https://www.arera.it/it/dati/mr/mree_puntiattivi.htm). In addition, many offers proposed to customers can be seen (more than a thousand offers in the first trimester of 2023) with an intense commercial activity of the retailers to attract customers. Nevertheless, the market over the years has seen a progressive reduction in its Herfindahl-Hirschman Index (HHI, Hirschman, 1980) both in the electricity and natural gas retail market. The only critical competitive dynamic is the domestic electricity market with the HHI in 2023 estimated at 2146, slightly above the critical threshold of 2000. Nevertheless, the retail market for both gas and energy appear to have a good competitive structure that is worth analyzing. In this context, the Italian competitive environment has changed radically, and retailers have implemented new strategies to maintain their market share and attract new customers. In light of these profound changes, it is interesting to analyze the performance of companies to understand what differences can be seen in the market between the various operators and whether there are features that have benefited certain operators more than others.

Extensive research has previously been conducted to analyze one specific aspect of the natural gas and energy market, such as efficiency, productivity and consumer behavior. For example, in the consumer behavior field Wieringa et al. (2007) analysed the Dutch market, Yang (2014) the Danish one, Shin et al. (2017) the Japanese one, Flores et al. (2018) analysed the British market while Fontana et al. (2019) studied the Italian market. Erdogan et al. (2022) analyse the large-scale electricity consumers in the Turkish market. These various studies have made the various national authorities also pay attention to consumer behavior in their monitoring activities. For example, in Italy, the Authority has included a whole section in the monitoring retail report (ARERA, 2022). In the area of productivity and efficiency analysis, several contributions analyzing can be found in energy generation field, e.g. Athanassopoulos et al. (1999), Chen et al. (2021), Sengupta & Mukherjee (2022) and Fleishman et al. (2009). Similarly in natural gas and electricity distribution literature, there are several contributions such as Erbetta & Rappuoli (2008), Goncharuk & Lo Storto, (2017), Hawdon (2003) in

the case of gas industry and Hjalmarsson & Veiderpass (1992) or Førsund & Kittelsen (1998) for electricity distribution. Focusing on the Italian retail case, the research group led by Capece, in several papers analysed the Italian natural gas retail market (Capece et al., 2008, Capece et al., 2010, Capece et al., 2012) and electricity (Capece et al., 2013). Capece et al. (2008) analysed the natural gas retail market using the Edgeworth index method and found that medium-sized companies perform best in terms of profitability and productivity compared to the ideal company. Large companies, on the other hand, mainly have positive profits due to the increased mark-up resulting from their bargaining power with suppliers, which allows them to purchase gas at lower prices. In Capece et al. (2010) the authors identify 4 clusters of operators and note that incumbent operators perform the least well while specialized operators or those who work purely in the north or are medium-sized perform better than the others. Similar are also the conclusions of subsequent work (Capece et al., 2012). Di Pillo et al. (2012) perform a similar analysis as Capece's group did in the Italian natural gas retail market but consider an extended period (2004-2009), which also allows them to assess the effects of regulatory decisions in the period of interest. Their main conclusion is that multi-business companies achieved the worst performance because they failed in their business diversification strategy. In this context, Di Leo et al. (2022) propose a framework to assess the retail market performance (and its sustainability). To do that, they proposed to assess the sellers' margins (expressed as the difference between the final selling price and the purchase price of raw materials).

The main research question in this article is the assessment of the efficiency in reaching high margins of the retailers using a robust benchmarking methodology and trying to understand which operational levers or characteristics could bring advantages to Italian retail operators.

To do so, we propose two different robust benchmarking models to evaluate *cost* and *commercial* efficiencies in achieving high *direct* margin. Unlike what has been proposed in the literature, the methodology used in this paper measures the relative efficiency of retailers without assuming strong hypotheses and being more robust to the influence of outliers.

This research paper is organized as follows. Section 2 provides a description of our model for the Italian energy and gas retail market and the main hypothesis behind the benchmarking carried out. Section 3 describes the methodology, the models and the dataset used for the analysis. Section 4 illustrates the results considering the

various characteristics of the retailers analyzed and Section 5 concludes the paper providing policy implications.

2   Modelling the Italian retail market

We analyze the Italian energy and natural gas retail market as a market with perfect competition in the long run (the gas and energy sales market has been liberalized for more than 10 years). In particular, the conditions we hypothesized are a large number of operators and buyers, homogeneity of products or services, perfect information (incentivized as far as possible by ARERA also through information platforms for consumers such as the offers portal https://www.ilportaleofferte.it/portaleOfferte/), zero transaction costs and free market entry and exit. According to competition theory, in a competitive market, competition between companies leads to reduced profits and profit margins. In other words, in a competitive market, companies cannot maintain high profits in the long run because competition forces companies to reduce prices to attract customers (Waldman & Jensen, 2016). In this context, the assumptions of our analysis are the following.

By defining the price P, the marginal cost (MC) and the markup ε, it is possible to summarize the relationships between energy and gas sellers and their suppliers. These relationships form the basis of the proposed models, defining in detail the 'decision levers' that an operator can move to improve performance.

Given the price of energy P_{1E} purchased from the seller defined as

$$P_{1E} = MC_{1E} + \varepsilon_{1E}$$

and the price of natural gas P_{1G} purchased from the seller defined as

$$P_{1G} = MC_{1G} + \varepsilon_{1G}$$

we define the resale price of electricity to the customer P_{2E} by the retailer as

$$P_{2E} = P_{1E} + \varepsilon_{2E} = CM_{1E} + \varepsilon_{1E} + \varepsilon_{2E}$$

and we define the resale price of natural gas to the customer by the seller as

$$P_{2G} = P_{1G} + \varepsilon_{2G} = CM_{1G} + \varepsilon_{1G} + \varepsilon_{2G}$$

Assuming that the seller's main cost for determining the price is exclusively the purchase price of energy and natural gas we can define the total retailer revenue R_{TOT} as:

$$R_{TOT} = P_{2E} * Q_E + P_{2G} * Q_G$$

Where Q_E is the amount of energy sold and Q_G is the amount of gas sold.

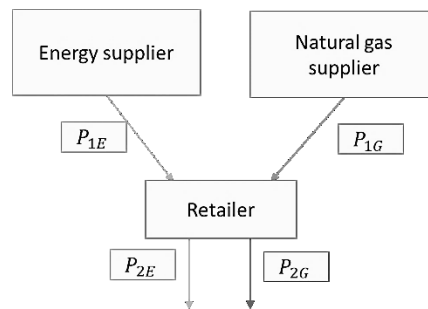The total raw material cost for a single operator can be defined as:

$$RAW\ MATERIALS\ COST_{TOT} = P_{1E} * Q_E + P_{1G} * Q_G$$

For simplicity we are assuming that the quantity of gas and energy sold to end customers is the same as that purchased from suppliers. Given the assumptions presented above, we define the *direct* margin as

$$Direct\ Margin = R_{TOT} - RAW\ MATERIALS\ COST_{TOT}$$
$$= (P_{2E} * Q_E + P_{2G} * Q_G) - (P_{1E} * Q_E + P_{1G} * Q_G) = (P_{2E} - P_{1E})Q_E + (P_{2G} - P_{1G})Q_G$$
$$= \varepsilon_{2E} * Q_E + \varepsilon_{2G} * Q_G$$

which we consider to be the closest possible approximation to the retailer's profit because in the case of resale activity only we consider fixed costs to be negligible (and distribution and regulatory costs set equal for all). It goes without saying that, in the case of a free market with perfect competition in long run, P_2 prices for both gas and energy will be similar for all operators with margins and mark-ups close to zero. This simplifying assumption makes it possible to evaluate the performance of the operators by considering only and intermediation process towards final customers. Figure 1 presents an outline of the relations between the retailer and the energy and gas suppliers described above.

Fig 1. Summary of the relation between suppliers and retailer

3 Methodology: a nonparametric robust benchmarking approach

One of the most known family of approaches for performance measurement is efficiency analysis, which consists in the estimation of an efficient (or optimal or best benchmarking) frontier against which to compare the performance of a unit with respect to its comparison set. As described in the seminal paper by Farrell (1957), an efficiency analysis starts with the estimation of a production or a cost frontier. In the literature, there are two main approaches for the estimation of an efficient frontier: the *non-parametric* and the *parametric* frontier approaches (for an introduction see e.g., Coelli et al., 2005). *Parametric* models require the assumption of a functional form for the production (or cost) frontier and the assumption of a distribution for the inefficiency. On the other hand, *non-parametric* models do not assume neither a functional form for the efficient frontier nor any distribution for the inefficiencies.

For these reasons, nonparametric models are more flexible than parametric ones. Different non-parametric methods are available in the literature, with the most popular being Data Envelopment Analysis (DEA) and Free Disposal Hull (FDH). Charnes et al. (1978) introduced the DEA method. The main objective of DEA is to evaluate the performance of a given sample of units (so called Decision-Making Units, DMU) to operate close to the *efficient* boundary of the production set.

DEA is a nonparametric approach, hence does not require prior specification of a functional form for the production technology and can manage multiple inputs and outputs but it requires the free disposability (meaning the possibility to destroy goods without costs) and the convexity assumptions. The efficient frontier estimated with DEA is constructed as a piecewise function over the data through linear programming. A DEA analysis can be conducted in an *input* or in an *output orientation*. An input-oriented DEA aims to analyze, given the outputs produced by the DMUs, how much the inputs could be contracted to reach the efficient frontier estimated empirically on the analyzed sample. In an output-oriented DEA, the aim is to maximize the outputs of each DMU, given the level of inputs used. Free Disposal Hull (FDH), introduced by Deprins et al. (1984) is a nonparametric frontier approach that assumes only the free disposability and does not impose the convexity assumption as DEA does.

However, nonparametric methods in efficiency analysis, including DEA and FDH, suffer for the influence of outliers or errors in the data and of the so-called "curse of dimensionality", meaning that for having a

meaningful estimation of the efficiency scores it is necessary to have a lot of data (for more details, see e.g. Daraio and Simar, 2007). This is why we chose a non-parametric and robust methodology called *Order-m efficiency analysis* for benchmarking Italian retailer operators in reaching high margins. Cazals et al. (2002) introduced Order-m efficiency analysis that is a nonparametric approach because it does not make any assumption about the inefficiency and the functional form of the frontier. In addition to the advantages of nonparametric approaches, Order-m efficiency scores have nice statistical properties that make them very attractive for empirical analysis: they are root-n consistent estimators, asymptotically unbiased and asymptotically normally distributed (see more details in Daraio and Simar, 2007, p. 77).

For its ability of not enveloping all data points, Order-m efficiencies are more robust to outliers and extremes in the data, and this is another attractive property of this approach. For empirical applications, this property means that we can avoid one of the more important limitations of the traditional nonparametric estimators, related to their deterministic nature. In addition, as a consequence of their statistical properties, robust order-m measures of efficiency do not suffer of the curse of dimensionality shared by most nonparametric estimators and by the DEA/FDH efficiency estimators. This is a relevant property for empirical analysis since it states that we can work with samples of moderate size and we do not require large samples to avoid imprecise estimation (e.g. large confidence intervals). Finally, the value of m could be considered as the number of potential competitors against which to carry out the benchmarking and also the trigger parameter to set the desired degree of robustness, i.e. the percentage of high performers of the population we want to exclude in our more realistic and robust benchmarking comparison (see more details in Daraio and Simar, 2007, p. 78).

An Order-m analysis, in the *output orientation*, consists in benchmarking a DMU (x, y) against the average maximal output reached by *m* peers randomly drawn from the population of DMUs that are producing a level of output higher or equal to that of the analyzed DMU. When y is univariate the full frontier for an input level x is defined as:

$$\psi(x) = \lambda(x, y)y = \sup\{y | S_{Y|X}(y|x) > 0\}$$

Then the expected order-m frontier is defined, for a fixed integer value of $m \geq 1$ as the expected value of the maximum of m random variables $Y \ldots Y$ drawn from the conditional distribution function Y given that $X \leq x$.

Its formal definition is:

$$\psi_m(x) = E[\max(Y^1, ..., Y^m)|X \leq x] = \int_0^\infty (1 - [F_{Y|X}(y \mid x)]^m) \, dy$$

Its non-parametric estimator can be defined and computed by:

$$\hat{\psi}_{m,n}(x) = \hat{E}[\max(Y^1, ..., Y^m)|X \leq x] = \int_0^\infty (1 - [\hat{F}_{Y|X,n}(y \mid x)]^m) \, dy$$

We carry out an efficiency analysis with output-oriented order-m frontiers and report the results à la Shephard, that is efficient DMUs have a score of one while super-efficient units have a score greater than one, and inefficient units have a score less than one. The analysis is performed on R using the FEAR package version 3.1 (Wilson, 2008).

To summarize, thanks to this robust methodology we can identify efficient operators, analyze best practices, compare best practices with less efficient ones to investigate existing differences. After that we extend the analysis considering different factors (e.g. size, vertical integration etc.) to evaluate their influence on the efficiency of operators in achieving high *direct* margin. In order to that we consider the classifications of the operators proposed by Capece et al. (2008). To make a descriptive comparison of the efficiency scores between more than two groups of classification we use the nonparametric Kruskal-Wallis test for the equality of the medians (Kruskal & Wallis, 1952). The asymptotic properties of order-m estimators that are root-n consistent, asymptotically unbiased and asymptotically normally distributed allow us to avoid the statistical inconsistencies associated with DEA/FDH estimators (see for more details Kneip et al., 2016).

After that, Dunn's test (Dunn, 1964) was adopted to determine exactly which groups are different after the results of Kruskal-Wallis test. Dunn's Test performs pairwise comparisons between each independent group and tells you which groups are statistically significantly different. In the case of a comparison between two classification groups, we adopted the Mann-Whitney test (Mann & Whitney, 1947). This method makes it possible to assess whether two statistical samples come from the same population. The tests are performed using the R package ggstatsplot (Patil, 2021). More details on the classifications for each operator is reported in the Section 3.2.

3.1   Data

To have as homogeneous a context as possible, the operators considered are only those who, according to the authority's listing and according to their economic activity (using the ATECO code of 2017) only carry out energy and/or natural gas retail activities in Italy We selected comparable and homogeneous operators. Starting from the official list of the Italian energy and gas authority (ARERA) of operators (available at https://www.arera.it/ModuliDinamiciPortale/elencooperatori/elencoOperatoriHome) and integrating the balance sheet data of 2020 from the database "Analisi informatizzata delle aziende italiane" (AIDA, 2023) (last date accessed 20/04/2023) we examined for our work 120 energy and natural gas retail operators. For these operators we obtained data on:

- Raw material costs in 2020.

- Total Revenue, Added Value (VA), Earnings Before Interests Taxes Depreciation and Amortization (EBITDA) in 2020.

- Direct Margin in 2020.

- Kilowatt per hour (kWh) consumed by customers in 2020.

- Cubic meters (m^3) of natural gas consumed by customers in 2020.

- Profit/Losses at the end of 2020.

We also classified the operators similarly to what was proposed by Capece et al. (2008). Each retailer was classified in:

- **Business classification** ("Double business" or "Only energy"). The classification is based on the energy and/or natural gas sold by an operator in 2020. This is a proxy to consider the *"horizontal integration"* of an operator in both the energy and natural gas retail market.

- **Size classification** by 2020 total revenue quartiles, dividing the operators into large (if revenues greater than 8703056), medium-large (if revenues greater than 2665873), medium-small (if revenues greater than 579354), and small for the others.

- **Society classification.** This classification is based on belonging to a corporate group or not according to what is reported in the ARERA national registry. This classification is a proxy for the *"vertical integration"* of operators.

- **Time classification**, dividing between operators entered in the market after 2016 (*post* unbundling) and pre 2016 *(pre* unbundling*)*. This classification and a proxy for the assessment of *'incumbents'* in the market.

In the dataset there is a high number of the double business operator, 96, and 24 are only electricity retailers (also called specialized operators). There are also 25 operators that are part of a corporate group and 95 not belonging to a corporate group. In Table 1 we reported the number of operators also for the other classification group.

Tab. 1 Different types of classification of the operators analyzed

Classification type	Classification description	Number of operators
Society classification	Belonging to a corporate group	25
	Not belonging to a corporate group	95
Business classification	Double Business	96
	Only Energy	24
Time classification	After 2016	61
	Pre 2016	59
Size Classification	Large	30
	Large-Medium	30
	Small	31
	Small-Medium	29

3.2 **Specification of the Models**

Based on this data and the assumption explained in section 2, we propose two models: one to assess *cost efficiency* and the other to assess *commercial* efficiency.

The cost efficiency model aims to assess which operators, for the same raw material cost, manage to have a higher margin. That is, they manage to sell at a higher P_{2E} and/or P_{2G} price than their competitors by working on the two possible levers, i.e., markups ε_{2E} and ε_{2G} and the raw material supply price P_{1E} and P_{1G}.

On the contrary, the *commercial* model aims to assess which operators, for the same quantities sold, manage to have a higher margin. That is, they manage to sell at a higher P_{2E} and/or P_{2G} price than their competitors for the same quantities sold Q_E and Q_G. In this case, the strategy adopted on the supply price side is not affected. The exclusion of the procurement cost optimization strategies only allows an evaluation relative to the quantity sold.

There are limitations on these assumptions that are made in our model. The first relates to an 'aggregate view' of both the energy and gas market without considering differences between the type of energy purchased (e.g., renewables only, mix or fossil fuels only) and the end customer to whom the P_2 price is being applied (e.g. domestic gas customer or low voltage domestic energy customer). These assumptions are derived from the limited unbundling of operators' annual financial statements which, by their nature, only allow for an aggregated view of operators' operations. Another assumption that limits the proposed models concerns the quantities sold and purchased of natural gas and energy. For simplicity's sake, the two Q's are considered equal over the period considered since natural gas and energy are not easily stored by sellers.

Therefore, the first model (the ***cost*** efficiency model) considers as input in the cost of raw materials (X_c above) and as output in the *direct* margin (Y above). The model is output oriented. The second model, the ***commercial*** efficiency model, we use as input the number of m³ of natural gas consumed (X_{t_1} above) and the kWh consumed by customers (X_{t_2} above) in 2020. As output, the use of *direct* margin *(Y)* is proposed as in the previous model. In both models, the data were divided by the standard deviation to reduce the different magnitude of the variables considered[1]. In Table 2 we reported the descriptive statistics of the data used in the models.

---

[1] We added 1 to the value of cubic meters (m³) sold for greater discriminating power in the case of operators who do not sell gas.

Tab. 2 Descriptive statistics on the variables used in the analysis

Descriptive Statistic	X_c	X_{t_1}	X_{t_2}	Y
Minimum	1056	0	8678	587
First quartile	414571	28778	1339912	95143
Median	1718871	487001	5729321	517876
Mean	7897059	3603785	28106493	1889356
Third quartile	6912813	2321846	18315715	1544940
Maximum	84890569	71451154	427962632	29238794
Standard Deviation	15768410	9428332	68803054	4249359

4   Results and discussion

4.1. Results of the Cost efficiency model

The first step in carrying out an order m analysis concerns the choice of m. Following a simulation for m=1 to m=120, we selected m= 90 given the balance between super-efficient and efficient, which in our case shows 4 efficient units (about 3.3%) and 23 super-efficient units (19.2%). Figure 2 shows a bimodal distribution of the efficiency scores of this first model with a group of units concentrated toward the frontier and another spread toward the lowest values of efficiency.
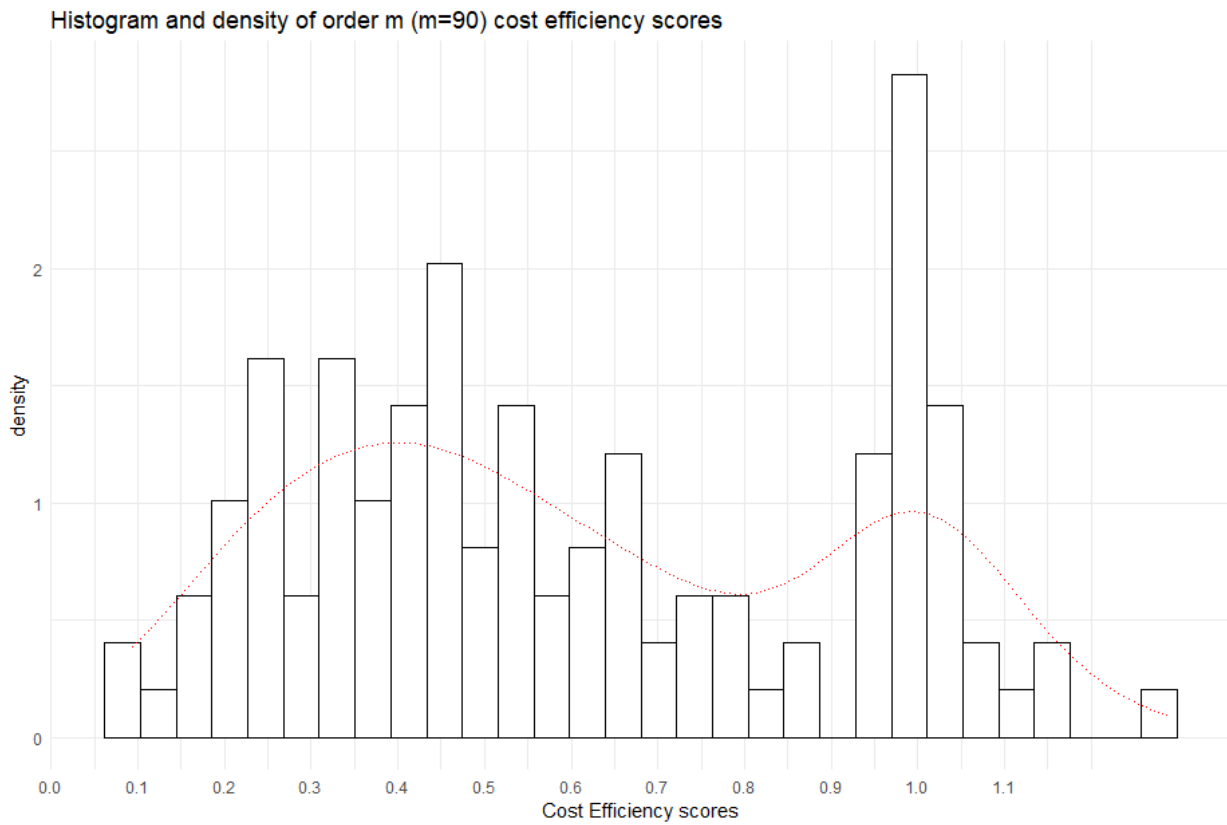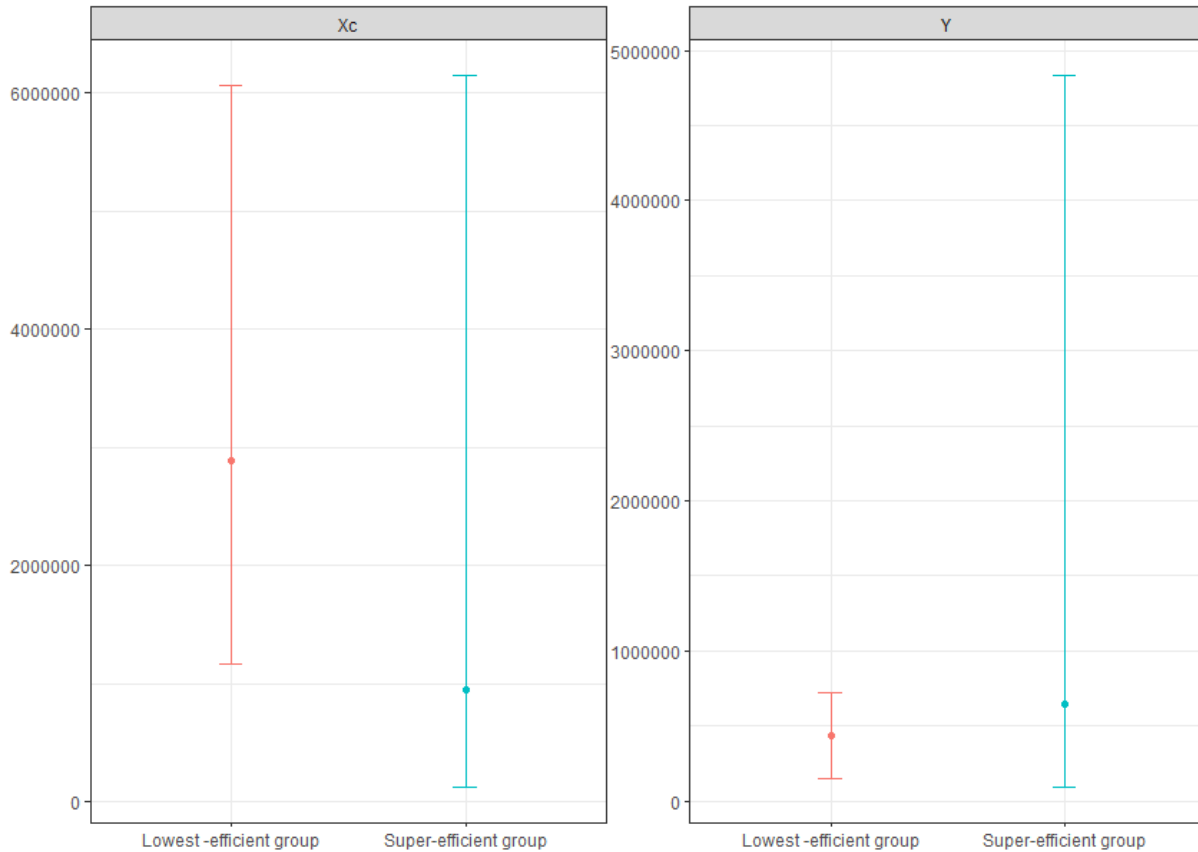
Fig. 1 Histogram and distribution (in red) of the cost efficiency scores

To understand the results, we focus our attention on the Super-efficient operators (i.e. operators with an efficiency greater than 1, hence called Super-efficient group) and the lowest-efficient operators (hence called Lowest-efficient group), so that best practices or characteristics can be identified. Analyzing the Super-efficient group (23 operators) and the 23 operators in the lowest-efficient group (the same number as the super-efficient group) we found differences between the best performers and the worst performers in margin creation at the same raw material cost.

As shown in Table 3 the super-efficient units have a X_c with a median of 956000 Euros (raw material costs) for a median Y of 643000 Euros. In contrast, the lowest-efficient group have a median of 4235354 and 370557.3 Y. A remarkable difference in the results obtained even when considering the minimum and maximum (both input and output) employed by the less efficient operators. This difference can also be easily seen in Figure 3, which shows the first quartile, median and third quartile values of raw material cost and direct margin of both groups.

Fig. 3 Cost efficiency model: visualization of descriptive statistics of super-efficient and lowest-efficient operators.

The red lines represent the Lowest-efficient group, the blue lines represent the super-efficient group. The points represent the median, the upper bars represent the third quartile, the lower bars represent the first quantile.

In the super-efficient group, we do not notice any influences related to belonging or not belonging to one of the proposed classifications. We have, according to the classification by business, 16 Double business operators and 7 only energy retailers. According to the size classification, 6 are large, 4 are large-medium, 4 are small-medium and 9 are small-medium. According to the classification by group, 17 are not part of a corporate while 6 are in a corporate group and according to the time classification, 16 are incumbent operators (after 2016) and 7 are historical operators (Pre 2016). They also turn out to be all operators in profit except one. Almost the same can be said for the lowest efficient operators. In detail, in the lowest efficient group we have 21 double business and 2 only energy. For the size classification, 4 are large, 10 are medium-large, 6 are
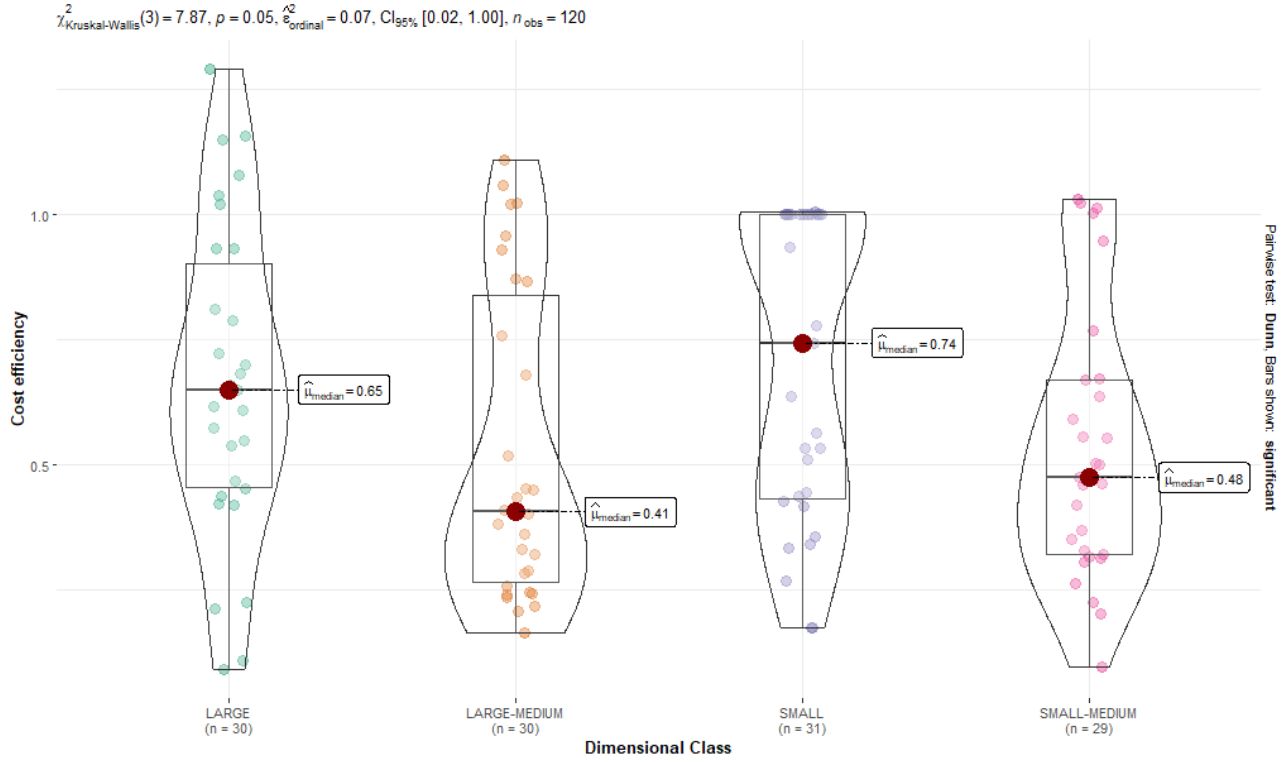
medium-small and 3 are small. Three operators are in a corporate group and 20 are not in a corporate group; 7 are after 2016 and 16 are pre 2016.

Tab. 3 Cost efficiency model: Descriptive statistics of the input and output used in the comparison of the Super-Efficient Group vs the Lowest-efficient group

	Mean	Standard Deviation	Minimum	First quartile	Median	Third quartile	Max
Super-efficient group							
X_c	12210506	22151230	41139	126683	956000	6145257	66400442
Y	3644375	6687520	20057	92318	643000	4831830	29238794
Lowest -efficient group							
X_c	4372376	4235354	54362	1166400	2885378	6069710	15151062
Y	466942	370557.3	3543	152170	433791	723697	1244495

After analyzing the best practices, we analyze at macro level all the operators to assess whether the characteristics of the operators can influence the efficiency score. It is worth emphasizing that the robust frontier methodology used is not conditioning the data-generating process on the characteristics considered. The Mann-Whitney test results shows that there is a significant difference between operators by business class (p-value=0.02) with specialized operators on average more efficient than double business operators (the median of cost efficiency for the double business operators is 0.52; for the only energy operators the median is 0.94). There is a significant difference in efficiencies in time groups (p value 0.04) with after 2016 operators on average more efficient than the old ones (median of 0.65 for the after 2016 vs 0.48 for the pre 2016). There is also a *slightly* significant difference in efficiencies in size groups as shown in Figure 4 (p value approximate to 0.05) but no significant differences between groups are reported. No others significant differences are noted for classes by corporate group.
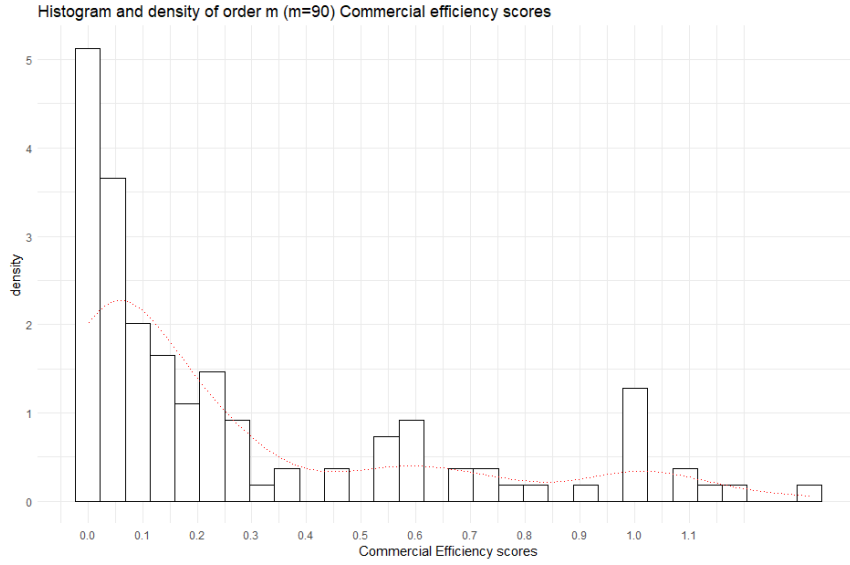
Fig. 4 Comparison of cost efficiencies between size classes. *Large (in green), Large-Medium (in orange), Small (in blue), Small-Medium (in pink). For each class, the boxplot, violin plot and median of the efficiencies are presented. The dots represent the efficiency scores of the operators and the red dot is the median. The first row of the figure shows the result of the Kruskal-Wallis test, the p value, the confidence interval and the number of operators analyzed (120).*

$\chi^2_{\text{Kruskal-Wallis}}(3) = 7.87, p = 0.05, \hat{e}^2_{\text{ordinal}} = 0.07, CI_{95\%} [0.02, 1.00], n_{\text{obs}} = 120$
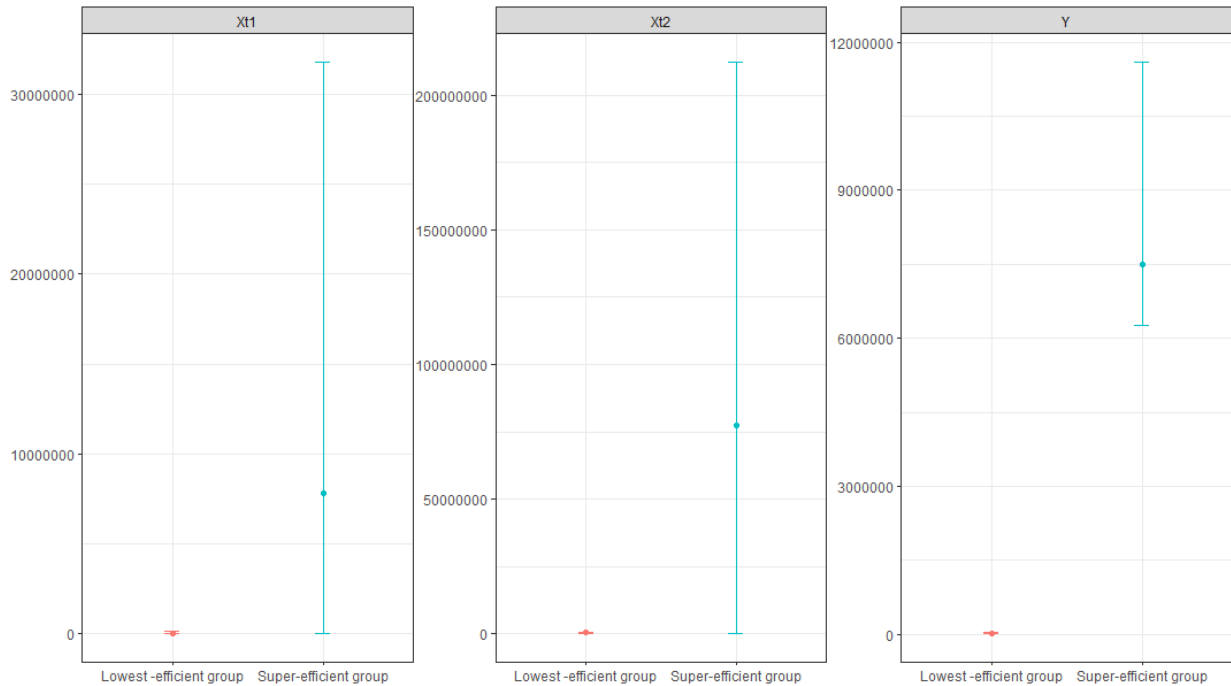
4.2. Results of the Commercial efficiency model

As in the previous case we need to select m. Following the simulation for m equals 1 to m equals 120, the choice of m fell to 90 to have some discriminant power for the identification of over efficient unit. With m=90 there are 3 efficient units (about 2.5% of the total) and 9 super-efficient units (7.5% of the total). The results of the *commercial* model are different in distribution from the previous model, as shown in Figure 5.

Fig. 5 Histogram and distribution (in red) of the *commercial* efficiency scores

Histogram and density of order m (m=90) Commercial efficiency scores

The distribution of the *commercial* efficiency scores shows a strong presence of operators with low efficiencies close to 0-0.1. This strong concentration points to widespread *commercial* inefficiency on the part of operators. As shown in Table 4 the super-efficient group have a median of 7825619 X_{t_1}; 204587 X_{t_2} and 6268115 Y. In contrast, the lowest-efficient group have a median of 836.9 X_{t_1} (i.e. no gas consumed/sold to the customers), 415713 X_{t_2}, and 23758 Y. This difference can also be easily seen in Figure 6, which shows the first quartile, median and third quartile values of X_{t_1}, X_{t_2} and Y for both groups.

Fig. 6 Cost efficiency model: visualization of descriptive statistics of super-efficient and lowest-efficient operators. *The red lines represent the lowest efficient operators, the blue lines represent the super-efficient operators. The points represent the median, the upper bars represent the third quartile, the lower bars represent the first quartile.*

A notable difference in the results obtained even when considering the minimum and maximum (both input and output) used by the lowest-efficient group (Table 4). In terms of characteristics, the super-efficient operators are 7 double business and 2 only energy; 7 large and 2 small; 2 are in a corporate group and 7 are not in a corporate group. Moreover, there are 6 historical operators (pre-2016) and 3 incumbent operators (after 2016). Considering their ability to make a final profit, only 2 operators are loss-making. Of these latter, one has high positive EBITDA and VA, indicating that financial activities weighed heavily on their ability to make a profit. The super-efficient group have a very high VA and EBITDA compared to the worst (VA 4878681 vs 2341; EBITDA 2889152 vs 2338). Their median profit is also much higher than that of less efficient operators (520506 Euros vs 1416 Euros).

	Mean	Standard Deviation	Minimum	First quartile	Median	Third quartile	Max
Super-efficient group							
X_{t_1}	21403012	26693758	1	16881	7825619	31775420	71451155
X_{t_2}	131219177	135944149	174874	204587	77473738	212104107	365529399
Y	10821582	9926885	3543	6268115	7497013	11585260	29238794
Lowest-efficient group							
X_{t_1}	45940,2	63540,64	1	1	836.9	92284	157025.4
X_{t_2}	651483	741208.1	258666	342507	415713	535533	2608188
Y	26829	8919.248	16822	20523	23758	36593	39468

As a feature, it is worth noting that the units with the lowest efficiency are all small; 4 of them are loss-making, with negative EBITDA, indicating how these operators have problems in the day-to-day running of their businesses. Regarding the others classification of the lowest efficient group, there are 5 double business operators and 4 only energy retailers; 2 are part of a corporate group and 7 are not in a corporate group; 6 are pre 2016 operators and 3 are after 2016 operators. The analysis of best practices thus shows how the size of the operators plays a very important role compared to other classifications to have a high *commercial* efficiency.

Expanding the focus of the analysis to all 120 retailers we identify significant differences between operators with similar characteristics using the Kruskal-Wallis test. Test results are presented in Figure 6 and indicate a significant difference between the dimensional class with a p-value of 1.17e-14. Significant differences between the groups are also identified by the Dunn's test. Large operators have higher efficiencies than the other groups. Also, the large medium operators are more efficient then the small and small medium operators. The Mann-Whitney test identifies a significant difference in efficiency between pre-2016 and after 2016 operators (P-value 4.45e-04) with pre-2016 operators more efficient than after 2016 operators (median efficiency 0.22 vs. 0.06). No differences are identified by the tests between different businesses and between operators in a corporate group or not. From the results of the tests, we can understand that historicity or incumbency in the market and the size of revenues affect *commercial* efficiency.
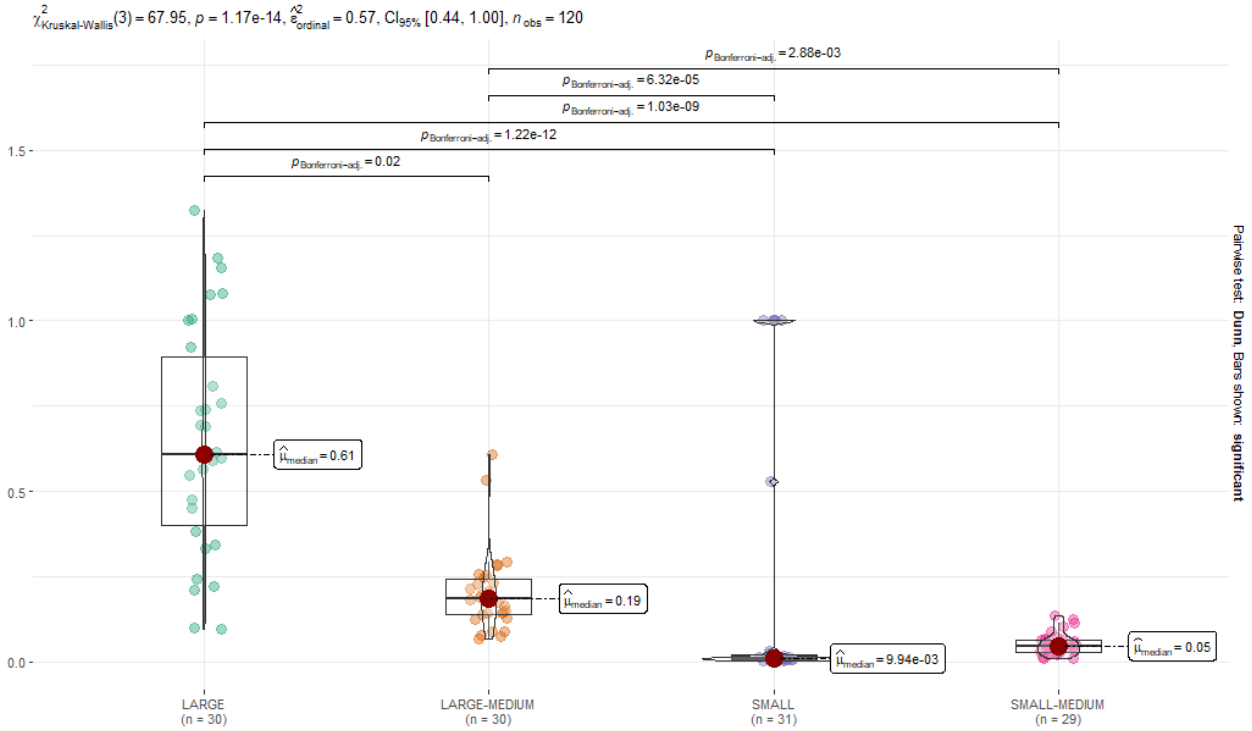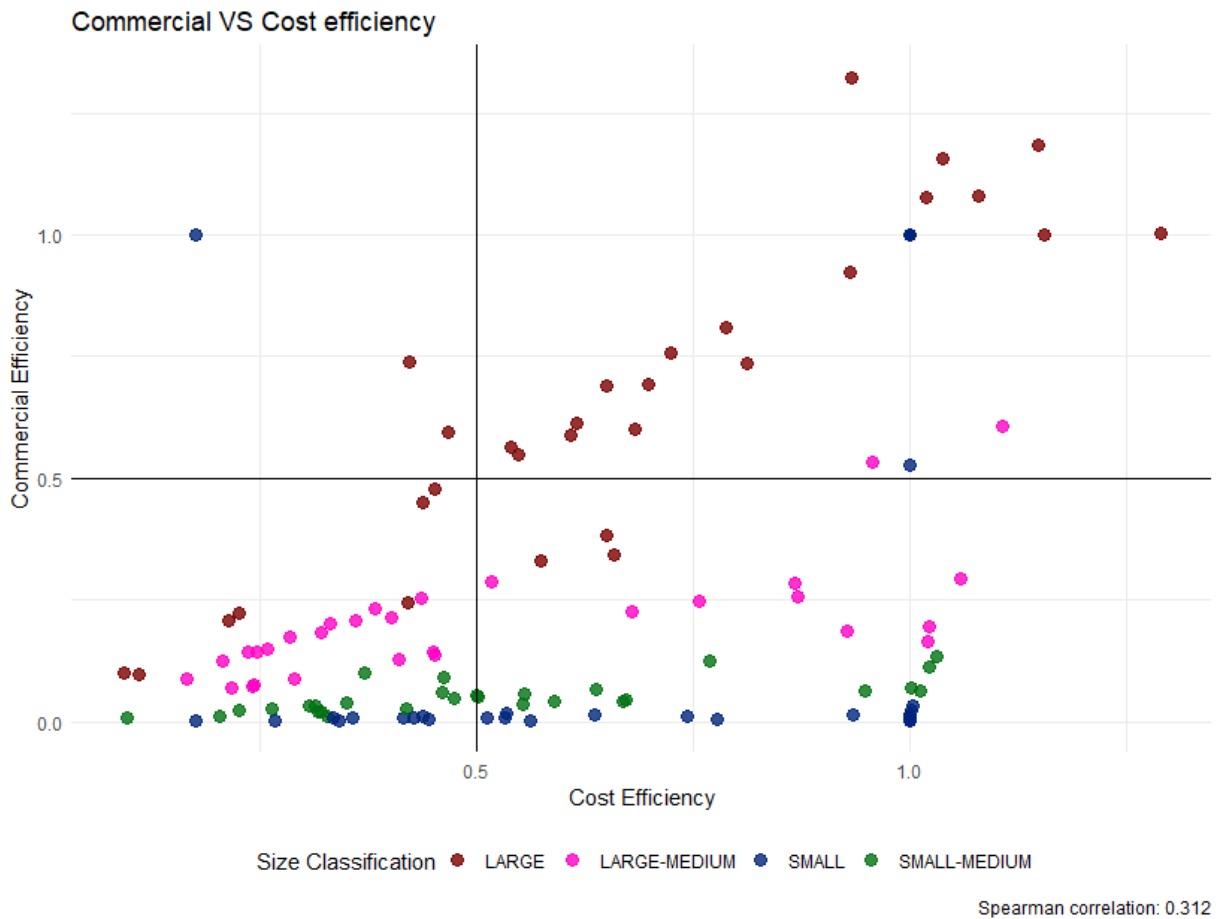
Fig. 7 Comparison of commercial efficiencies by size classes of the operators. *Large (in green), Large-Medium (in orange), Small (in blue), Small-Medium (in pink). For each class, the boxplot, violin plot and median of the efficiencies are presented. The dots represent the efficiency scores of the operators and the red dot is the median. The first part of the figure shows the result of the Kruskal-Wallis test, the p value, the confidence interval and the number of observations (120). The lines between the boxplots represent significant differences between the groups identified by Dunn's test (Bonferroni-adjusted)*

4.3. Discussion

The two efficiency scores obtained from the proposed models can be used together to see which operators are simultaneously efficient on raw material costs (i.e., higher margin for the same raw material costs) and volume sold (i.e., higher margin for the same volume sold). It is difficult, however, to rely on a single model to understand which players are very efficient or which characteristics may influence the market. Analyzing the market one-dimensionally could lead to an incomplete and distorted view, as also suggested by Capece et al. (2008). Therefore, it is possible to combine the two model results together to assess which operators are efficient and which are not from a *commercial* and *cost* perspective. Between the two efficiency scores there is no strong Spearman correlation (0.312). Figure 6 shows the efficiency scores of the two models. On the x-axis we have the *cost* efficiency score, and, on the y-axis, we have the *commercial* efficiency score. The points correspond to the intersection of these two efficiencies for each operator analyzed. We have divided Figure 6 into four quadrants: the North-East quadrant identifies efficient operators in both models; the North-West quadrant identifies *commercial* efficient but not *cost* efficient operators; the south-east quadrant identifies *cost*

efficient but not *commercial* efficient operators; finally, the south-west quadrant identifies inefficient operators in both models. We assessed whether the classification adopted influences the two efficiencies and it turns out that only size has an influence. In Figure 6, the colors represent an operator's belonging to the size classification we adopted: in red the large operators, in green the small operators, in blue the medium-small operators and in pink the medium-large operators. A comparison of the two scores shows that most operators are oriented toward being more cost-efficient than commercial-efficient. It can be seen from Figure 6 that operators classified as large manage to balance commercial and cost efficiency better than the others. Large operators balance both efficiencies well and are very present in the North-East quadrant of Figure 6. As the size decreases, the operators fail to balance the efficiencies and always lean towards cost efficiency over commercial efficiency. Figure 6 shows the details of this analysis.

Fig. 6 Scatter plot of cost and commercial efficiencies of Italian retail operators. *Each point represents cost and commercial efficiency of an operator. The points correspond to the intersection of these two efficiencies for each operator analysed. We have divided the graph into four quadrants: the north-east quadrant identifies cost efficient operators in both models; the north-west quadrant identifies commercial efficient operators but not cost efficient operators; the south-east quadrant identifies cost efficient but not commercial efficient operators; finally, the south-west quadrant identifies inefficient operators in both models The colours in the graph represent the size classification of the operator. The lines for dividing into quadrants are positioned at 0.5.*

5 Conclusion and policy implications

This article responds to the research question on the characteristics that can influence the efficiency in reaching high *direct margin* of the Italian energy and natural gas retailers. Two robust benchmarking models for evaluating cost and *commercial* efficiency in achieving high *direct* margin were developed. The use of the robust benchmarking methodology allowed us to evaluate some of the aspects of performance evaluation expressed by Thanassoulis (2001):

i)   identification of good operational practices (i.e. the best performer) both for individual models and by comparing models together.

ii)   an analysis of the most productive operational scale, comparing the results of efficiencies in individual models and comparing them subsequently.

We found out that the specialization, incumbent status, and the dimension of the retailer influences the *cost* efficiency. For *commercial* efficiency, it is influenced by dimension and incumbent status in the market. The results are partially in agreement with the research of Capece et al. (2008; 2010). We agree that the efficiency in achieving large margins depends on the size (in terms of total revenues), specialization and incumbent status of the retailers. Contrary to the conclusions of previous studies, business group operators do not seem to affect the achievement of the efficiency. These results therefore show that the long work to separate retail activities between different operators has avoided advantages for vertically integrated companies. The relevance of size is highlighted in both models by considering analyses of super-efficient operators.

The focus on size is further emphasised by comparing the results of the two models (see Figure 6). This comparison allows us to identify 4 quadrants. In these quadrants, the operators with high *commercial* and *cost* efficiencies are the large operators. It is also noticeable that small and medium-sized operators suffer from competition of large operators preluding being *cost*-efficient more than *commercially* efficient. A similar problem in the Italian context has already been identified by Stagnaro et al. (2020) which identified improper behaviors by large operators in Italy. This behavior may limit the expansion of small operators and consequently also limit their efficiency by not being able to increase their sales volume.

Based on our findings, our policy proposal is to promote a strengthening of small and small-medium retailers and incentive for joint purchase of energy or natural gas for small/small-medium operators to increase their scale and efficiency. In Italy, to contribute to the strengthening of retailers, a barrier to entry for operators has been decided by the Ministry of Business and Enterprise (MISE). The MISE has decided to set up a list of electricity retailers (https://www.mase.gov.it/pagina/elenco-venditori-energia-elettrica-eve), requiring a minimum share capital of 100000 euro for registration. Parallel to these efforts, the aggregation of demand through a demand aggregator (e.g. through the creation of consortia) could represent an opportunity for small and medium-sized retailers in the electricity sector to gain advantages in the purchase of raw materials. This

intermediation instrument would bring together the purchasing needs of several operators, allowing them to benefit from economies of scale and more advantageous contractual conditions when negotiating with energy and gas suppliers. Aggregating demand through an aggregator offers numerous advantages to small operators:

a) **Greater bargaining power:** A demand aggregator can negotiate wholesale contracts with suppliers representing a group of small players. Demand aggregation enables small sellers to obtain more favorable purchasing conditions through increased bargaining power.

b) **Simplified access to the wholesale market**: The aggregator can simplify the process of access to the wholesale market for small sellers by removing administrative barriers and minimum participation requirements that might be difficult for individual traders to meet.

c) *Cost reduction*: Demand aggregation allows operational costs to be shared among participants, thus reducing individual costs for small sellers.

d) *Risk mitigation*: Through demand aggregation, small sellers can mitigate the risk associated with the supply of energy or natural gas. An aggregator can diversify sources of supply and ensure a reliable supply to participating operators.

To stimulate demand aggregation of small operators, regulators and policy makers can take measures such as:

- Creation of a regulatory framework that encourages the participation of demand aggregators and establishes clear rules for their operation.

- Promoting information and awareness among small retailers on the opportunity and benefits of demand aggregation.

- Support financially or provide tax incentives for demand aggregators that cater for small sellers.

- Promote the creation of digital platforms or portals that facilitate demand aggregation and simplify the negotiation and contracting process.

Another option could be the mergers of small companies with each other (e.g., through mergers or acquisitions). The increase in scale due to this type of operations would benefit the new entrant, who would have both more bargaining power on the raw material purchasing side and a larger market share, which would mitigate the effects of misbehavior by larger competing operators. Further growth opportunities could also come from commercial innovation in the side of offers to customers. Offering additional services, for the same

price, can lead to better market positioning and customer attractiveness as also emphasized in the framework proposed by Di Leo et al. (2022). This could involve also offering energy contracts with flexible options or incentive programs for the gradual adoption of environmentally friendly technologies. Some operators have already started to do this, as can be seen on the Platform *"Portale Offerte"* provided by ARERA (https://www.ilportaleofferte.it/portaleOfferte/).

Although the results lead us to emphasize the beneficial effect of size, it is important to recognize that small operators can distinguish themselves in the crowded and competitive Italian market through customer loyalty having a direct and personalized contact with users thanks to a commercial network of agents. This aspect cannot be overlooked because it can be both a significant competitive advantage and disadvantage for small operators. An increase in the size of these operators could be hampered by this type of customer relationship. Maintaining a network of agents establishing direct contact with customers may entail higher costs than operators relying on other means such as a centralized call center. Furthermore, the agent network may be effective at a local or regional level but can become problematic when it comes to expanding nationwide. Therefore, although size remains the main factor influencing efficiency in achieving high direct margins, it is essential to also consider the importance of customer retention through direct contact for smaller operators. This topic, together with the discussion on commercial innovation, goes beyond the scope of this article, however, but offers an interesting cue for further research on business innovation and competitive strategies of small operators.

As shown in Section 2, there are simplifying assumptions we have made in modelling this market. As possible future developments, the assumptions made can be developed further and other more complex models can be proposed. Furthermore, the models developed do not consider the differences between different customers and between different businesses. A possible development could be modelling by type of customer (domestic or non-domestic, etc.) or by individual business. In this case, to develop such models, it is necessary to have separate accounting data to have the necessary details for modelling and robust benchmarking analysis.

CREdit Author Statement:

Simone Di Leo: Conceptualization, Methodology, Writing, Data curation **- Marta Chicca:** Data curation, Reviewing, Policy implication **– Cinzia Daraio:** Supervision, Methodology, Validation **- Andrea Guerrini-** Reviewing, Editing, Validation**,** Policy implication**- Stefano Scarcella:** Data curation, Reviewing, Policy implication.

Disclaimer: The content of this paper is the expression of the individual authors who assume all responsibility for the work and does not represent the opinions of the regulatory agency ARERA.

References

Analisi informatizzata delle aziende italiane (AIDA). [data from Italian retailer of energy and natural gas in Italy][Data set]. https://www.bvdinfo.com/it-it/le-nostre-soluzioni/dati/nazionali/aida, last accessed 20/04/2023.

ARERA, (2015), Disposizioni in merito agli obblighi di separazione funzionale (unbundling) per i settori dell'energia elettrica e del gas (Delibera 22 giugno 2015 296/2015/R/com), https://www.arera.it/it/docs/15/296-15.html (last accessed 21/04/2023)

ARERA, (2022), Monitoraggio sull'evoluzione dei mercati di vendita al dettaglio dell'energia elettrica e del gas. Rapporto di aggiornamento di luglio 2022 (37/2022/I/com, 327/2021/I/com, ARG/com 151/2011), https://www.arera.it/it/docs/22/342-22.htm (last accessed 20/04/2023).

Athanassopoulos, A. D., Lambroukos, N., & Seiford, L. (1999). Data envelopment scenario analysis for setting targets to electricity generating plants. European journal of operational research, 115(3), 413-428.

Coelli, T. J., Rao, D. S. P., O'Donnell, C. J., & Battese, G. E. (2005). An introduction to efficiency and productivity analysis. Springer science & business media.

Capece, G., Cricelli, L., Di Pillo, F., & Levialdi, N. (2008). A productivity analysis of the Italian gas retail market. Aravossis, Brebbia, Gomez editors. Environmental Economics and Investment Assessment II, 108, 43-52.

Capece, G., Cricelli, L., Di Pillo, F., & Levialdi, N. (2010). A cluster analysis study based on profitability and financial indicators in the Italian gas retail market. Energy Policy, 38(7), 3394-3402.

Capece, G., Cricelli, L., Di Pillo, F., & Levialdi, N. (2012). New regulatory policies in Italy: Impact on financial results, on liquidity and profitability of natural gas retail companies. Utilities Policy, 23, 90-98.

Capece, G., Di Pillo, F., & Levialdi, N. (2013). Measuring and comparing the performances of energy retail companies: Firm strategies following the liberalization. International Journal of Energy Sector Management.

Cazals, C., Florens, J. P., & Simar, L. (2002). Nonparametric frontier estimation: a robust approach. *Journal of econometrics*, *106*(1), 1-25.

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European journal of operational research*, *2*(6), 429-444.

Chen, Z., Zhao, W., & Zheng, H. (2021). Potential output gap in China's regional coal-fired power sector under the constraint of carbon emission reduction. Energy Policy, 148, 111888.

Corsatea, T. D., & Giaccaria, S. (2018). Market regulation and environmental productivity changes in the electricity and gas sector of 13 observed EU countries. Energy, 164, 1286-1297.

Daraio, C., & Simar, L. (2007). Advanced robust and nonparametric methods in efficiency analysis: Methodology and applications. Springer, New York.

Defeuilley, C. (2009). Retail competition in electricity markets. Energy Policy, 37(2), 377-386

Di Leo, S., Chicca, M., Daraio, C., Guerrini, A., & Scarcella, S. (2022). A Framework for the Analysis of the Sustainability of the Energy Retail Market. Sustainability, 14(12), 7320.

Di Pillo, F., Capece, G., Cricelli, L., & Levialdi, N. (2012). Business diversification strategies in the Italian energy markets. International Journal of Industrial and Systems Engineering, 6(11), 2814-2820.

Dunn, O. J. (1964). Multiple comparisons using rank sums. Technometrics, 6(3), 241-252.

Erbetta, F., & Rappuoli, L. (2008). Optimal scale in the Italian gas distribution industry using data envelopment analysis. Omega, 36(2), 325-336.

Erdogan, M. R., Camgoz, S. M., Karan, M. B., & Berument, M. H. (2022). The switching behavior of large-scale electricity consumers in The Turkish electricity retail market. Energy Policy, 160, 112701.

European parliament and council Directive 98/30/EC of 22 June 1998 concerning common rules for the internal market in natural gas.

European Parliament and Council, (1996). Directive 96/92/CE concerning common rules for the internal market in electricity, Bruxelles, Belgium.

Farrell, M. J. (1957). The measurement of productive efficiency. Journal of the royal statistical society: series A (General), 120(3), 253-281

Fleishman, R., Alexander, R., Bretschneider, S., & Popp, D. (2009). Does regulation stimulate productivity? The effect of air quality policies on the efficiency of US power plants. Energy Policy, 37(11), 4574-4582.

Flores, M., & Price, C. W. (2018). The role of attitudes and marketing in consumer behaviours in the British retail electricity market. The Energy Journal, 39(4).

Fontana, M., Iori, M., & Nava, C. R. (2019). Switching behaviour in the Italian electricity retail market: Logistic and mixed effect Bayesian estimations of consumer choice. Energy Policy, 129, 339-351.

Førsund, F. R., & Kittelsen, S. A. (1998). Productivity development of Norwegian electricity distribution utilities. Resource and energy economics, 20(3), 207-224.

Goncharuk, A. G., & lo Storto, C. (2017). Challenges and policy implications of gas reform in Italy and Ukraine: Evidence from a benchmarking analysis. Energy Policy, 101, 456-466.

Hawdon, D. (2003). Efficiency, performance and regulation of the international gas industry-a bootstrap DEA approach. Energy policy, 31(11), 1167-1178.

Hjalmarsson, L., & Veiderpass, A. (1992). Productivity in Swedish electricity retail distribution. The Scandinavian journal of economics, S193-S205.

Hirschman, A. O. (1980). *National power and the structure of foreign trade* (Vol. 105). Univ of California Press.

Italian Parliament, 1999. Legislative Decree N° 79, "Attuazione della direttiva 96/92/CE recante norme comuni per il mercato interno dell'energia elettrica", Rome, Italy.

Kneip, A., Simar, L., & Wilson, P. W. (2016). Testing hypotheses in nonparametric models of production. Journal of Business & Economic Statistics, 34(3), 435-456.

Kruskal, W. H., & Wallis, W. A. (1952). Use of Ranks in One-Criterion Variance Analysis. Journal of the American Statistical Association, 47, 583-621.

Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. The annals of mathematical statistics, 50-60.

Ministry of Productive Activities Legislative Decree n. 164 of 23 May 2000, concerning the implementation of Directive 98/30/EC and the common rules for the internal market in natural gas.

Patil, I. (2021). Visualizations with statistical details: The 'ggstatsplot' approach. Journal of Open Source Software, 6(61), 3167, doi:10.21105/joss.03167

Sengupta, D., & Mukherjee, D. (2021). By-production of electricity and particulates: efficiency of Indian thermal power plants revisited. International Journal of Energy Sector Management.

Shin, K. J., & Managi, S. (2017). Liberalization of a retail electricity market: Consumer satisfaction and household switching behavior in Japan. Energy Policy, 110, 675-685.

Stagnaro, C., Amenta, C., Di Croce, G., & Lavecchia, L. (2020). Managing the liberalization of Italy's retail electricity market: A policy proposal. Energy Policy, 137, 111150.

Thanassoulis, E. (2001). *Introduction to the theory and application of data envelopment analysis*. Dordrecht: Kluwer Academic Publishers.

Waldman, D. E., & Jensen, E. J. (2016). *Industrial organization: theory and practice*. Routledge.

Wieringa, J. E., & Verhoef, P. C. (2007). Understanding customer switching behavior in a liberalizing service market: an exploratory study. Journal of Service Research, 10(2), 174-186

Wilson, Paul W. (2008), "FEAR 1.0: A Software Package for Frontier Efficiency Analysis with R," Socio-Economic Planning Sciences 42, 247--254.

Yang, Y. (2014). Understanding household switching behavior in the retail electricity market. Energy policy, 69, 406-414.