A DEA-based approach to customer value analysis

Laurens Cherchye *   Bram De Rock[†]   Bart Dierynck[‡]
Pieter Jan Kerstens[§]   Filip Roodhooft[¶]

January 9, 2023

*Corresponding author. Affiliation: Department of Economics, University of Leuven (KU Leuven). E. Sabbelaan 53, B-8500 Kortrijk, Belgium. E-mail: laurens.cherchye@kuleuven.be. Laurens Cherchye gratefully acknowledges the Fund for Scientific Research - Flanders (FWO) and the Research Fund KU Leuven for financial support.

†Affiliation: ECARES, Université Libre de Bruxelles and Department of Economics, University of Leuven (KU Leuven). Av. F. Roosevelt 50, B-1050 Bruxelles, Belgium. E-mail: bram.de.rock@ulb.be. Bram De Rock gratefully acknowledges FWO and FNRS for their support.

‡Affiliation: Tilburg School of Economics and Management, Tilburg University. Warandelaan 2, 5037 AB Tilburg, Netherlands. Email: b.dierynck@tilburguniversity.edu.

§Affiliations: Flemish Institute for Technological Research (VITO) and Department of Economics, University of Leuven (KU Leuven). Boeretang 200, B-2400 Mol, Belgium. E-mail: pieter-jan.kerstens@kuleuven.be

¶Affiliation: Department of Accounting, Finance, and Insurance, University of Leuven (KU Leuven) and Accounting and Finance, Vlerick Business School. Naamsestraat 69, B-3000 Leuven, Belgium. Email: filip.roodhooft@kuleuven.be

Abstract

Firms have become increasingly customer-centric, implying that customers, rather than products, are treated as the most important asset of a firm. The switch to customer-centric strategies also implies that firms are collecting an enormous amount of customer-related data. The purpose of this paper is to propose a DEA-based methodology to determine the contribution of customer segments to firm value. We show the practical usefulness of our methodology through an application to Activity Based Costing (ABC) data collected from a large European telecom provider, which offers fixed telephone, mobile telephone, digital television and internet subscriptions. Our analysis reveals that the average cost reduction potential across all customer segments amounts to 1.26% of the total controllable costs, which represents approximately EUR 5 million when expressed in monetary terms. We also document substantial variation in the cost reduction potential across customer segments.
Keywords: data envelopment analysis, customer value, multi-output efficiency, ABC systems

1 Introduction

During the last decade, firms have become increasingly customer-centric, implying that customers, rather than products, are treated as the most important asset of a firm, and that acquiring and retaining profitable customers has become the main strategic focus (Fader, 2020; Jain and Singh, 2002; Kumar and Shah, 2009; Palmatier, Moorman, and Lee, 2019). The switch to customer-centric strategies also implies that firms are collecting an enormous amount of customer-related data (Bonacchi and Perego, 2019; Galbraith, 2005; Latinovic and Chatterjee, 2019). Currently, these customer-related data are analyzed by means of techniques such as customer profitability analysis (CPA; see for instance van Raaij, Vernooij, and van Triest (2003)) and/or customer lifetime value (CLV; see for instance Glady, Baesens, and Croux (2009)). Although these two techniques differ on some important aspects, they both aim to determine the contribution of individual customers or customer groups to firm value based on data about past customer behavior (Holm, Kumar, and Rohde, 2012). The outcomes of both techniques are used to support various managerial decisions, such as the allocation of marketing resources, pricing and customer differentiation decisions.

The purpose of the current paper is to propose an alternative DEA-based methodology to determine the contribution of customer segments to firm value in terms of (future) revenue or costs. We illustrate the practical usefulness of our newly developed methodology by using data from a large European telecom provider. To be clear, our purpose is not to develop a substitute for CLV and/or CPA but to develop a methodology that can help to extract additional insights from the customer-related data that firms collect. As such, we believe that our newly developed methodology is complementary with existing approaches to analyze customer-related data.

Our novel methodology has its origins in data envelopment analysis (DEA), which has become popular both as an analytical research instrument and as a practical decision-support tool to evaluate the efficiency of a DMU (i.e. decision-making unit, which is typically a business unit, office unit, or branch of a private or public sector company). DEA determines the efficiency of a DMU by comparing the input-output performance of the DMU to that of other DMUs operating in a similar technological environment. The outcome of a DEA exercise indicates whether the same output level can be produced with a lower level of inputs or whether a higher output level can be produced with the same level of inputs. Since the seminal work of Charnes, Cooper, and Rhodes (1978), the methodological DEA literature has mainly focused on refinements that account for uncontrollable factors, data variation, economies of scope, and the allocation of inputs to outputs. See, for example, Cook and Seiford (2009) and Zhu (2015) for reviews of DEA, and Cherchye, De Rock, and Vermeulen (2008), Cherchye et al. (2013) and Cherchye, De Rock, and Walheer (2015) for recent developments that are directly relevant to the current study.

We advocate the use of DEA to analyze the massive amounts of data that firms collect on the behavior of their customers. The aim of our methodology is to identify customer segments from which the contribution to firm value can be increased. Our methodology relies on comparing the input-output performance of similar customer segments and indicates for each customer segment the cost reduction that can be achieved without decreasing the output level of the customer segment. In the next section we argue in detail that our DEA-based methodology has several noteworthy strengths. First, it allows for heterogeneity with respect to the way in which the costs incurred to serve customer segments are transformed into outputs such as revenues, upselling, and churn rate. Second, it does not resort to parametric specifications about the way in which inputs are transformed into outputs, which makes that the results of our analysis cannot be driven by an ill-specified transformation function. Third, our DEA-based methodology allows for the inclusion of inputs and/or outputs that are not expressed in monetary terms, which increases the precision of our analysis and opens the possibility to include inputs and/or outputs that are difficult to express in monetary terms.

We show the practical use of our methodology through an empirical application that uses data collected from a large European telecom provider. Our analysis reveals that the average cost reduction potential across all customer segments amounts to 1.26% of the total controllable costs, which represents approximately EUR 5 million when expressed in monetary terms. As expected, there is substantial variation in the cost reduction potential across the customer segments. The average cost reduction potential of the ten segments with the highest cost reduction potential amounts to 28.19% of the total controllable costs. Our analysis also allows us to compute the cost reduction potential of every geographical region, product combination, or socio-demographic segment separately. Such an analysis can be useful as most marketing actions, such as advertising through radio, television, and newspapers, cannot yet be targeted on the most detailed customer segmentation level. Next, our analysis gives some guidance to managers on how they can exploit the cost reduction potential of a particular customer segment. Specifically, when our analysis identifies a cost reduction potential, this implies that there is another customer segment that realizes more outputs at lower costs. This 'dominating' customer segment can be identified and the manager can learn from the way in which this customer segment is served in order to exploit the cost reduction potential of the 'dominated' customer segment.

Before entering our analysis, it is important to remark that our methodology is inherently static; we do not model dynamic aspects that are specific to intertemporal models of managerial decision making regarding customer value. This contrasts with existing techniques such as CLV that explicitly model the customers' lifetime value. In Section 3.2 we will discuss how our DEA-based methodology relates to such a more standard CLV analysis. In the concluding section, we will also briefly digress on the possibility to incorporate dynamic aspects in our efficiency evaluation method.

The rest of this paper unfolds as follows. Section 2 advocates the use of DEA as a tool for customer value assessment. Section 3 presents our DEA-based methodology to determine the contribution of customer segments to firm value. Section 4 introduces the set-up of our empirical application and Section 5 discusses our results. Section 6 concludes and discusses some avenues for further research.

2 DEA as a tool for customer value analysis

Our DEA-based methodology for analyzing customer value determines the efficiency with which different customer segments are served by comparing the input-output performance of serving a particular customer segment to the input-output performance of other customer segments of the same firm. The inputs are typically the controllable and uncontrollable costs that are made to serve the customer segments such as, for instance, operating expenditures, acquisition costs, and development costs. The outputs are the key performance indicators that a firm defines as part of the customer-centric strategy such as, for instance, the realized revenue, the churn rate, and the upselling and cross-selling potential of the customer segment. Importantly, the data about these inputs and outputs are typically available in the firm's information system. The outcome of our analysis determines, for each customer segment, with how much the costs that are made to serve the segment can be reduced while still realizing the same output level. As our model also includes outputs that determine the future and indirect value of a customer segment, this cost reduction potential reflects the unrealized value that the segment can contribute to firm value.

This DEA-based approach to determining the contribution of customer segments to firm value has several benefits that originate from the distinguishing features of DEA as an efficiency measurement methodology. First, as DEA is nonparametric in nature, it does not resort to some (typically unverifiable) parametric specifications of the way in which inputs are transformed in outputs. DEA thus allows for heterogeneity with respect to the way in which inputs are transformed in outputs. Allowing for such heterogeneity is relevant in the context of customer centricity as customer centricity implies that providing service to customers should be adapted to the characteristics of the customer segments (Fader, 2020; Palmatier, Moorman, and Lee, 2019). As a result, the way in which costs to serve customer segments are transformed in outputs likely differs across customer segments. Allowing for heterogeneity with respect to the way in which inputs are transformed into outputs across customer segments is also an important advantage of our DEA-based approach compared to CLV models, which often assume that the way in which inputs are transformed to outputs is homogeneous across customer segments (Holm, Kumar, and Rohde, 2012). Specifically, CLV models usually adopt a parametric approach and determine the contribution of a customer segment to firm value by using stochastic models in which customer behavior is modeled by means of probabilistic functions. Next to the fact that CLV models assume homogeneity across customer segments in transforming inputs to outputs, the parametric approach as adopted by most CLV models also implies that an identified low contribution of a particular customer segment to firm value may well be driven by an ill-specified transformation function (rather than a truly low contribution).

A final important benefit of our methodology is that it allows for heterogeneity with respect to the unit of account of the inputs and outputs. Suppose, for instance, that the outputs of serving a customer segment are the realized revenues, which are expressed in monetary terms, and customer complaints, which are typically expressed as a percentage reflecting the total number of customer complaints relative to the total number of served customers. To analyze the efficiency of the input-output transformation of the different customer segments by means of DEA, it is not required to estimate the monetary consequences of customer complaints by means of deterministic or stochastic models. The heterogeneity that DEA allows with respect to the unit of account of the inputs and/or outputs implies that it is not very restrictive with respect to the inclusion of inputs and/or outputs in the model. Moreover, the detected inefficiencies cannot be caused by errors in expressing inputs and/or outputs in monetary terms.

3 Methodology

We begin this section by introducing some necessary notations and concepts, and we subsequently explain how our methodology can be used in the context of customer analysis. In a following step, we present our cost efficiency measure, and we show its practical implementation through linear programming.

3.1 Preliminaries

The telecom operator wishes to evaluate the cost effectiveness of serving its customer segments: can resources be decreased without lowering the level of the current objectives? Its objectives are not only monetary but also include customer satisfaction, which was measured indirectly by the churn rate and the number of upsells. Detailed activity-based costing (ABC) data are available to evaluate these objectives. To align with the literature on efficiency analysis, we use the term "inputs" to refer to resources and "outputs" to refer to objectives.

We assume K observed customer segments that produce M different outputs. For each segment k ($k = 1, ..., K$), the output quantities y_k^m ($m = 1, ..., M$) are captured by $\mathbf{y}_k \in \mathbb{R}_+^M$. To open the black box of the multi-output production process, we adopt the approach of Cherchye et al. (2013) and Cherchye, De Rock, and Walheer (2015). These authors take as a starting point that each output is characterized by its own production process. They distinguish between "output-specific inputs", which are allocated to specific production processes (e.g. operations expenditures), and "subjoint inputs", which simultaneously contribute to multiple production processes. As explained below, subjoint inputs will be particularly relevant for our application in the current paper. The "public goods" nature of these inputs entails specific cost saving effects (i.e. economies of scope; see Cherchye, De Rock, and Vermeulen (2008) for a detailed discussion).

At this point, we remark that the use of subjoint inputs distinguishes our approach from the alternative method that is known as network DEA; see, for example, Tone and Tsutsui (2009) for a seminal contribution. This approach also aims at capturing inter-relations between production processes, but through the modeling of intermediate outputs and by allocating inputs to either intermediate or final output production processes. Clearly, this network DEA approach may also productively use ABC data to accurately model production processes. However, it does not allow for modeling the cost-saving effects resulting from (sub)joint inputs that are simultaneously used in multiple production processes. As a matter of fact, achieving these cost saving effects was exactly a main production objective of the telecom operator that we evaluate. Consequently, the operator's management advised us to (only) model subjoint inputs in our efficiency analysis.

We assume there are N subjoint inputs, which we denote for each customer segment k with $\mathbf{Q}_k \in \mathbb{R}^N_+$. An input is called joint if it is used in the production process of all outputs and subjoint if it is only used by some production processes. To formally model this, we introduce a binary $N \times M$ matrix $\mathbf{D}$ with:

$$D_{j,m} = \begin{cases} 1 & \text{if } Q^j_k \text{ is used to produce output } m, \\ 0 & \text{otherwise,} \end{cases}$$

using $Q^j_k \in \mathbb{R}_+$ for the j-th subjoint input quantity of customer segment k. Note that $\mathbf{D}$ is not segment-specific since we assume that the nature of the inputs is the same for all segments k and, therefore, does not depend on the specific input vector $\mathbf{Q}_k$. A specific instance is when all inputs are subjoint for all outputs, meaning that the matrix $\mathbf{D}$ contains only ones. This particular set-up will apply to our empirical application in Section 4. Example 1 below provides a simple numerical illustration.

Summarizing, we assume a dataset:

$$S = \{\mathbf{y}_k, \mathbf{Q}_k, \mathbf{D}\}_{k=1,\ldots,K} .$$

3.2 Customer lifetime value

Before presenting our DEA-based methodology, we briefly introduce the concept of customer lifetime value (CLV). The cash flow $CF_{k,t}$ a customer k generates at time t for the firm is equal to the profit the firm makes on the customer:

$$CF_{k,t} = \mathbf{w}'_{k,t}\mathbf{y}_{k,t} - \mathbf{P}'_{k,t}\mathbf{Q}_{k,t},$$

with output prices $\mathbf{w}_{k,t} \in \mathbb{R}^M_+$ and subjoint input prices $\mathbf{P}_{k,t} \in \mathbb{R}^N_+$. In general, these prices can be customer-specific. This also applies to our own application, where we will use input prices that are specific to customer segments.

Customer lifetime value for a customer k is then defined as the discounted sum of future cash flows $\{CF_{k,t+i}\}^T_{i=1}$ that the customer generates over the horizon T. Hence, the CLV of customer k at time t for the next T periods is defined as:

$$CLV_{k,t} = \sum_{i=1}^{T} \frac{CF_{k,t+i}}{(1+\rho)^i} = \sum_{i=1}^{T} \frac{\mathbf{w}'_{k,t+i}\mathbf{y}_{k,t+i} - \mathbf{P}'_{k,t+i}\mathbf{Q}_{k,t+i}}{(1+\rho)^i},$$

with discount rate ρ (which we assume to be constant for simplicity). Clearly, $CLV_{k,t}$ can be improved by (i) increasing total revenue, (ii) decreasing total costs or (iii) a combination of (i) and (ii). In this paper we opt for (ii), because cost reductions do not affect customers directly, which makes them a safe way to increase profit and customer value. However, one could easily redo the analysis for (i) or (iii) using the framework of Cherchye, De Rock, and Walheer (2016).[1]

[1]The focus on cost reduction also naturally allows us to avoid any convexity assumption on the output side. We see this need for minimal prior structure as an attractive feature of our methodology. Particularly, such a convexity assumption would be problematic given the nature of the outputs that we use in our own empirical application (see Section 4 for details).

At this point, we remark that one cannot exclude seasonal (temporary) effects on the customer lifetime value. For example, a marketing campaign affecting costs/revenues or customers switching subscriptions. Therefore, in our empirical application we will average the results over time to reduce these seasonal effects and drop the time subscripts on the variables. The result of our analysis should then be interpreted as average potential cost reductions and, accordingly, average potential customer lifetime improvements. In the concluding section we will briefly discuss the possibility to explicitly incorporate dynamic aspects of managerial decision making in our efficiency evaluation method.

3.3 Multi-output cost efficiency

We provide an alternative for analyzing customer value by using a multi-output cost efficiency framework that structurally models the interrelations between production processes. To formally introduce this model, we need a production technology in terms of input sets that represent all combinations of inputs that, for a given customer segment k, can produce a given quantity y_k^m of output m:

$$\mathcal{I}^m(y_k^m) = \{(\mathbf{d}_m \mathbf{Q_k}) \text{ can produce } y_k^m\},$$

with $\mathbf{d}_m$ denoting the m-th row of $\mathbf{D}$ and thus, more specifically, $\mathbf{d}_m\mathbf{Q}$ the inputs used in the production of output y^m. Note that the $\mathcal{I}^m$ are output-specific but are assumed to be the same for all customer segments k. We assume that these sets $\mathcal{I}^m$ satisfy the following property:

Axiom – nested input sets: For customer segment k and output m, $y_k^{m*} \geq y_k^m$ implies $\mathcal{I}^m(y_k^{m*}) \subseteq \mathcal{I}^m(y_k^m)$.

In words, this axiom implies that we can always freely dispose of some output y_k^{m*} to produce a lower output y_k^m. Put differently, if we observe a certain input-output combination, then we can always achieve lower (and worse) objectives for the same inputs.

In order to assess cost efficiency for every output separately, we need some way of assigning portions of the subjoint input costs over the different outputs. For this purpose, we use "implicit prices" defined as follows.

Definition 1 (Implicit prices). *For customer segment k with subjoint input prices $\mathbf{P}_k \in \mathbb{R}_+^N$ and binary matrix $\mathbf{D} \in \mathbb{R}^{N\times M}$, implicit prices are any vectors $\mathbf{P}_k^m \in \mathbb{R}_+^N$ that satisfy $\sum_{m=1}^M \mathbf{P}_k^m \mathbf{d}_m = \mathbf{P}_k$.*

The vector $\mathbf{d}_m$ in the above definition ensures that the price/cost of a subjoint input is only distributed over the outputs that use it. These implicit prices allow us to allocate the cost of shared inputs over the different outputs. In some applications, ABC data or managerial knowledge may provide sufficient information for allocating the costs directly to the outputs. In what follows, we assume that such information is not available. In the absence of prior knowledge of the output-specific implicit prices, we then choose the prices that imply a "most favorable" cost allocation (i.e., maximizing the cost efficiency of the customer segment under evaluation; see further).[2]

We are now in a position to adapt the cost efficiency definition of Cherchye et al. (2013) and Cherchye, De Rock, and Walheer (2015) to our specific set-up.

[2]This practice of using most favorable prices in efficiency evaluations reflects the "benefit-of-the-doubt" interpretation of DEA methods. See, for example, Cherchye et al. (2007) for a detailed discussion.

Definition 2 (Cost efficiency)**.** *Customer segment k is multi-output cost efficient if, for each output m, there exist an input set $\mathcal{I}^m(y_k^m)$ that satisfies Axiom nested input sets and implicit prices $\mathbf{P}_k^m \in \mathbb{R}_+^N$ such that:*

- $\mathbf{d}_m \mathbf{Q}_k \in \mathcal{I}^m(y_k^m)$,
- $(\mathbf{P}_k^m \mathbf{d}_m)'\mathbf{Q}_k = \min_{\mathbf{Q} \in \mathcal{I}^m(y_k^m)} (\mathbf{P}_k^m \mathbf{d}_m)'\mathbf{Q}$.

Intuitively, this definition says that a customer segment is multi-output cost efficient if, for every output m, its chosen input combination is (i) technically feasible and (ii) is the lowest cost combination for the given implicit prices.

For given implicit prices and Ainput sets, Definition 2 can be easily operationalized: it suffices to find $\mathbf{Q}$ that produces y_k^m at minimal cost. As such, the first task is to reconstruct the input sets from the dataset S. Cherchye et al. (2013) show that the reconstructed set $\mathcal{I}^m$ is completely characterized by $\mathcal{D}_k^m = \{s | y_k^m \leq y_s^m\}$, which contains all observed customer segments s that dominate segment k in terms of the m-th output (i.e. $y_s^m \geq y_k^m$).

Let $c_k^m \equiv \min_{s \in \mathcal{D}_k^m} (\mathbf{P}_k^m \mathbf{d}_m)'\mathbf{Q}_s$ denote the minimal cost to produce y^m. Then, it suffices to check whether:

$$CE_k^m \equiv \frac{c_k^m}{(\mathbf{P}_k^m \mathbf{d}_m)'\mathbf{Q}_k}$$

equals 1. If CE_k^m is smaller than 1, then costs can be reduced for customer segment k, which means that the customer lifetime value can be improved.

In practice, price information is often not available to the empirical analyst, which hinders the above calculation. As indicated above, in such a case we can evaluate cost efficiency using "most favorable" (shadow) prices $\mathbf{P}_k$ and implicit prices $\mathbf{P}_k^m$. In that case, we compute:

$$\begin{aligned}
CE_k \equiv \max \ & \frac{\sum_{m=1}^{M} c_k^m}{\sum_{m=1}^{M} (\mathbf{P}_k^m \mathbf{d}_m)'\mathbf{Q}_k} \\
s.t. \quad & c_k^m \leq (\mathbf{P}_k^m \mathbf{d}_m)'\mathbf{Q}_s \qquad \forall s \in \mathcal{D}_k^m, \ \forall m = 1, \dots, M, \\
& \sum_{m=1}^{M} \mathbf{P}_k^m \mathbf{d}_m = \mathbf{P}_k, \\
& c_k^m \in \mathbb{R}_+, \mathbf{P}_k \in \mathbb{R}_+^N, \mathbf{P}_k^m \in \mathbb{R}_+^N.
\end{aligned}$$

This program evaluates customer segment k in the best possible light by choosing (shadow and implicit) prices that make the segment appear as efficient as possible (summed over all outputs) when compared to its peers in $\mathcal{D}_k^m$.

In its original form, the above programming problem is nonlinear, as free variables enter the denominator of the objective function. Following standard practices, we can circumvent this by normalizing the denominator to be equal to 1. This results in the following linear program that we have used in our empirical exercise:

$$
\begin{aligned}
CE_k \equiv \max & \sum_{m=1}^{M} c_k^m \\
s.t. \quad & c_k^m \leq (\mathbf{P}_k^m \mathbf{d}_m)' \mathbf{Q}_s \qquad \forall s \in \mathcal{D}_k^m, \ \forall m = 1, \ldots, M, \\
& \sum_{m=1}^{M} \mathbf{P}_k^m \mathbf{d}_m = \mathbf{P}_k, \\
& \sum_{m=1}^{M} (\mathbf{P}_k^m \mathbf{d}_m)' \mathbf{Q}_k = 1, \\
& c_k^m \in \mathbb{R}_+, \mathbf{P}_k \in \mathbb{R}_+^N, \mathbf{P}_k^m \in \mathbb{R}_+^N.
\end{aligned}
$$

The program computes a minimal cost c_k^m for every output m. The sum of these output-specific costs gives our measure CE_k of overall cost efficiency of segment k. Intuitively, this overall cost efficiency can be decomposed as a weighted sum of output-specific efficiencies. That is, $CE_k = \sum_{m=1}^{M} w_k^m CE_k^m$ where the weights w_k^m represent the share of the output-specific cost in the overall cost:

$$
w_k^m = \frac{(\mathbf{P}_k^m \mathbf{d}_m)' \mathbf{Q}_k}{\sum_{m=1}^{M} (\mathbf{P}_k^m \mathbf{d}_m)' \mathbf{Q}_k}.
$$

As a final note, we remark that our above definition of $\mathcal{D}_k^m$ overlooks an important aspect: it does not control for the relative size of the customer segments. Particularly, customer segments with many customers generate a lot of profits (output) and, therefore, will appear in the dominating set $\mathcal{D}_k^m$ of many k. Arguably, however, it is unfair to compare these large customer segments against the smaller ones. To account for this, we can adapt an original proposal of Ruggiero (1996) to our setting, and control for differences in the customer environment (such as size of the segment) by using a slightly modified definition of $\mathcal{D}_k^m$:

$$
\mathcal{D}_k^m = \{s | y_k^m \leq y_s^m\} \cap \{s | \mathbf{z}_s \leq \mathbf{z}_k\},
$$

where the vector $\mathbf{z} \in \mathbb{R}_+^{Env}$ captures environmental conditions, with higher values indicating a more favorable environment.

To conclude our methodological exposition, we illustrate the above concepts and tools through the numerical Example 1.

Example 1. *We consider a simple setting with two customer segments ($k = 1, 2$) that use one input ($N = 1$) to produce two outputs ($M = 2$):*

$$
\begin{aligned}
\mathbf{Q}_1 = \bar{\mathbf{Q}} \text{ and } \mathbf{Q}_2 = 2\bar{\mathbf{Q}} & \text{ as inputs for the two segments,} \\
y_1^1 = 1 \text{ and } y_1^2 = 5 & \text{ as outputs for segment 1, and} \\
y_1^2 = 5 \text{ and } y_2^2 = 1 & \text{ as outputs for segment 2,}
\end{aligned}
$$

where $\bar{\mathbf{Q}}$ represents an arbitrarily fixed constant (representing input quantities). We note that customer segment 2 consumes twice the input of customer segment 1. Further, we assume that the input is subjoint for both outputs (i.e., $\mathbf{d}_1 = \mathbf{d}_2 = 1$) and the two customer segments operate in the same environment (i.e., $\mathbf{z}_1 = \mathbf{z}_2$).

Using the output information, we can define the sets:

$$\mathcal{D}_1^1 = \{1,2\}, \mathcal{D}_1^2 = \{1\}, \mathcal{D}_2^1 = \{2\} \text{ and } \mathcal{D}_2^2 = \{1,2\},$$

which contain the observed customer segments that dominate the two segments in the respective outputs.[3]

Then, the linear program for segment 1 simplifies to:

$$\begin{aligned} CE_1 \equiv \max\, & c_1^1 + c_1^2, \\ s.t.\ \ & c_1^1 \leq \mathbf{P}_1^1 \bar{\mathbf{Q}}, \\ & c_1^1 \leq 2\mathbf{P}_1^1 \bar{\mathbf{Q}}, \\ & c_1^2 \leq \mathbf{P}_1^2 \bar{\mathbf{Q}}, \\ & \mathbf{P}_1^1 + \mathbf{P}_1^2 = \mathbf{P}_1, \\ & (\mathbf{P}_1^1 + \mathbf{P}_1^2)\bar{\mathbf{Q}} = 1, \\ & c_1^1, c_1^2, \mathbf{P}_1, \mathbf{P}_1^1, \mathbf{P}_1^2 \in \mathbb{R}_+. \end{aligned}$$

It is readily clear that solving this program obtains $CE_1 = 1$, hereby using $c_1^1 = \mathbf{P}_1^1\bar{\mathbf{Q}}$ and $c_1^2 = \mathbf{P}_1^2\bar{\mathbf{Q}}$. We conclude that segment 1 is identified as efficient. In other words, there exists a (shadow) price $\mathbf{P_1}$ and implicit prices $\mathbf{P}_1^1$ and $\mathbf{P}_1^2$ that make segment 1 cost efficient when compared to segment 2.

A directly similar reasoning obtains the same efficiency conclusion for DMU 2 (i.e. $CE_2 = 1$). In that case, the linear program becomes:

$$\begin{aligned} CE_2 \equiv \max\, & c_2^1 + c_2^2, \\ s.t.\ \ & c_2^1 \leq 2\mathbf{P}_2^1 \bar{\mathbf{Q}}, \\ & c_2^2 \leq \mathbf{P}_2^2 \bar{\mathbf{Q}}, \\ & c_2^2 \leq 2\mathbf{P}_2^2 \bar{\mathbf{Q}}, \\ & \mathbf{P}_2^1 + \mathbf{P}_2^2 = \mathbf{P}_2, \\ & 2(\mathbf{P}_2^1 + \mathbf{P}_2^2)\bar{\mathbf{Q}} = 1, \\ & c_2^1, c_2^2, \mathbf{P}_2, \mathbf{P}_2^1, \mathbf{P}_2^1 \in \mathbb{R}_+. \end{aligned}$$

We obtain $CE_2 = 1$ for $c_1^2 = 2\mathbf{P}_2^1\bar{\mathbf{Q}} = 1$ and $c_2^2 = \mathbf{P}_2^2\bar{\mathbf{Q}} = 0$. Particularly, customer segment 2 is identified as efficient because it dominates the first segment in the first output. The linear program then obtains the maximum value of unity by setting the implicit prices $\mathbf{P}_2^1 = 1/2\bar{\mathbf{Q}}$ and $\mathbf{P}_2^2 = 0$. Basically, this implies an (extreme) allocation in which the full input cost is assigned to the first (best performing) output, which reflects the principle of choosing a "most favorable" cost allocation.[4]

[3]Our use of output-specific dominating sets contrasts with more standard DEA methods, which typically define dominating sets in terms of all outputs simultaneously.

[4]The fact that the full input cost is allocated to a single output reflects the simplified set-up of our hypothetical example. Obviously, such an extreme cost allocation is usually not a realistic representation of the true (but unobserved) allocation over customer segments. In practice, we can exclude such unrealistic scenarios by restricting the possible values of the implicit prices in the linear program; these price restrictions can reflect the (limited) price information that is available (e.g. through ABC data). This parallels the frequent practice of using so-called "weight restrictions" in DEA (see, for example, Zhu, 2015).

4 Empirical application: set-up

We demonstrate the practical applicability of our methodology by means of a unique data set collected from a large European telecom provider that offers fixed telephone, mobile telephone, digital television and internet subscriptions. The inputs of our model are controllable costs, such as operating expenditures, acquisition costs and development costs, and uncontrollable costs, such as interconnection costs with other operators, roaming costs, IT-costs and billing costs. The outputs of our model are the revenue streams realized in the different product categories of the company, the churn rate of a customer segment, and the number of upsells of a customer segment. In total, our model contains 20 inputs and 7 outputs. In what follows, we first introduce the way in which the telecom provider segments its customer base. Next, we describe in more detail the inputs and the outputs that we use in our application.

4.1 Customer segments

The telecom operator segments its customer base on the basis of the product combination the customer has, the region in which the customer lives, and the socio-demographic category to which the customer belongs. The telecom operator offers fixed telephone, mobile telephone, digital television and internet, and customers can choose any possible combination. The main distinction for the product combination is the number of products, leading to product combinations, which we will label 'X-play packs', with 4, 3, 2, or 1 product respectively. The 0-play pack is a rest category, with all customers that can not grouped in one of the previous categories. Furthermore, the telecom operator distinguishes 11 regions and 6 socio-demographic groups. As our newly developed methodology boils down to comparing the input-output performance of different customer segments, it is important to only compare customer segments that operate in a similar environment. For that reason, we only compare customer segments within a particular X-play pack. Specifically, we only compare customer segments that have the same number of products (i.e. 1, 2, 3, or 4) in their X-play pack. Obviously, as the telecom operator offers 4 products, there are 4 different combinations for the 3-play pack, 6 different combinations for the 2-play pack and 4 different combinations for the 1-play pack. Combining the number of combinations within each X-play pack with the 11 regions and 6 socio-demographic groups leads to 66 customer segments for the 4-play pack, 264 customer segments for the 3-play pack, 396 customer segments for the 2-play pack, and 264 customer segments for the 1-play pack.

4.2 Data

The telecom operator provided us with data for the year 2014. For each month, we have detailed data on all costs and all revenues associated with every customer segment. We also have data about the total number of customers in each segment as well as about the migration of customers from one customer segment to another customer segment. The efficiency scores are computed on a monthly basis by comparing each customer segment with all similar customer segments in all periods (i.e. we assume no change in technology over time). While this is a strong assumption, the advantage of this approach is that we have much more observations to compare with. These results were then averaged over all months, because the telecom operator did not see benefits in analyzing the contribution of customer segments on a monthly basis. This averaging also reduces potential seasonal (temporary) effects. The telecom operator has $32,121,558$ customers in total and the average net margin per customer, which we calculate as (total revenues minus total costs)/number of customers, amounts to -0.5850 EUR. Upon looking into more detail, a clear pattern emerges: the net margin per customer is almost always negative for socio segment A. In order to illustrate this and the considerable heterogeneity in net margin per customer, we present heat maps in Figure 1 for all 2-play packs.

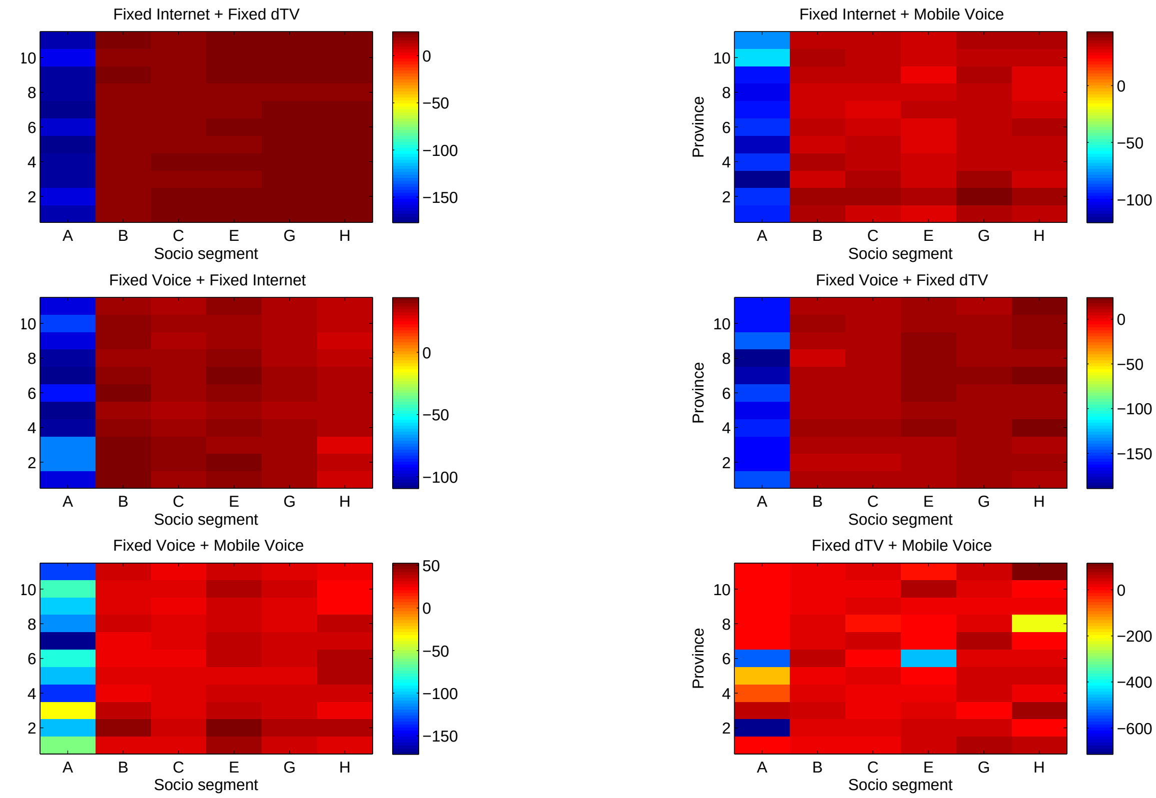

Figure 1: 2-play packs, average net margin per customer

Our model contains different outputs which together reflect the current contribution of a customer segment and the future potential of a customer segment. The current contribution of a customer segment is reflected by five revenue streams, which are the revenues for each customer segment for fixed telephone, mobile telephone, digital television, internet and other revenues. The future potential of a customer segment is reflected by the churn rate and the number of upsells for each customer segment. The churn rate of an X-play pack represents the percentage of customers that cancel their subscription entirely. The number of upsells for every customer segment is constructed from the monthly migration data and is defined as the number of existing customers of the telecom operator that change their subscription to that particular customer segment.

The inputs in our model are the costs that the telecom operator makes to realize the outputs. These costs typically consist of controllable and uncontrollable costs. Uncontrollable costs are usage costs, such as the interconnection costs between telecom operators and roaming costs, or fixed costs that are only controllable in the long run, such as billing costs, IT costs and costs for bad debt. After consulting with the management team of the telecom operator, we decided to ignore the uncontrollable costs for our analysis, as these costs can never be used to realize cost reductions in the short run. The controllable costs include various operating expenditures, acquisition costs and development costs. The acquisition and development costs for every customer segment are constructed from the monthly migration data by multiplying the acquisition cost, respectively the development cost, for a particular customer segment with the number of customers that migrate to that particular customer segment.

In total, we have 14 cost categories that serve as an input in our model. The name of every cost category including a short description of the cost is presented in Table 7 in the Appendix. In cooperation with the management team of the telecom operator, we assigned inputs to outputs to ensure that our model is a better reflection of reality (i.e. we derived the matrix $\mathbf{D}$ describing the subjoint inputs). Table 7 present an overview of the descriptive statistics of the different inputs and outputs that we include in our model. Table 8 in the Appendix does the same at X-play pack level. These descriptive statistics learn that there is a lot of variation in both the inputs and the outputs across the different customer segments.

A first approach to distill management recommendations based on the available data is to look at key performance indicators such as the number of customers in the different segments or the number of upsells in different customer segments. A second approach is to construct key performance indicators that combine revenues and costs such as the gross margin of a customer segment, which we define here as the difference between the total revenues of a customer segment and the usage costs (i.e. interconnection costs, roaming costs and costs of content), or the average net margin per customer for each customer segment, which we define as the difference between total revenues and total costs per customer segment divided by the total numbers of customers in the particular customer segment. The main conclusion of such an analysis is that the results very strongly depend on the key performance indicator one is analyzing, which calls for an approach in which the input-output performance of different customer segments is analyzed in a more structural way. This is what we do next.

5 Main results and managerial implications

We begin this section by presenting the main results of our empirical analysis. Subsequently, we discuss alternative managerial applications of our methodology.

5.1 Main results

The results we present are the outcome of analyses in which we only compare the customer segments offering a play pack with the same number of products to each other, which are 66 customer segments for 4-play pack, 264 customers segments for 3-play pack, 396 customer segments for 2-play pack, 264 customer segments for 1-play pack, and 52 customer segments for 0-play pack. For each of these customer segments, our methodology identifies a potential cost reduction, which reflects the cost reduction that can be realized in that particular customer segment while achieving the same output level. By doing so, our methodology analyzes the potential increase in customer lifetime value for each customer segment.

The results of our analysis reveal that the total potential cost reduction amounts to approximately EUR 5 million, which equals 1.26% of the total controllable costs. For each customer segment based on socio segment, play pack, and geographical region, our methodology calculates a particular potential cost reduction. Table 1 presents the summary statistics of the aggregated potential cost reductions for each play pack. A minority of 244 customer segments is efficient. As it can be argued that it is difficult to target individual customer segments to address the potential cost reductions, analyzing the aggregated potential cost reductions can be useful. Given the variation in the number of customers across play packs, we also present the aggregated potential cost reduction per customer. We observe that the highest aggregated potential cost reduction can be realized by focusing on the 2-play pack. Overall, our methodology allows us to calculate the potential cost reduction of each individual customer segment and thus also allows for aggregating the potential cost reduction to a level that fits with the level at which the firm targets its customers. Importantly, our methodology calculates potential cost reductions and it is up to the telecom operator to verify (1) the extent to which these potential cost reductions can be realized in particular customer segments and (2) the impact of realizing the potential cost reductions. To determine whether a potential cost reduction in a particular customer segment can be realized, operational knowledge concerning the particular segment should be combined with the results of our analysis.

Play pack	Mean	Std.	Max	EUR/`nb_cust`
4-play	4823.4558	8757.2108	37114.6421	0.3379
3-play	4872.5900	10480.9505	64859.3471	1.1683
2-play	5793.3951	20800.1563	243108.9998	2.5960
1-play	4392.2554	20551.9157	289958.2872	1.2699
0-play	248.0972	245.4731	1024.4908	0.0649

Table 1: Summary statistics on potential cost reductions

Not every customer segment carries equal weight in terms of economic importance for the telecom operator: the histogram in Figure 2 shows that the vast majority of customer segments only represent a very small fraction of the total controllable costs. More specifically, we compute that 19 (out of 1042) customer segments represent 20.62% of the total controllable costs, and 21.98% (229/1042) of customer segments represent 80.03% of the total controllable costs. Unweighted summary statistics, such as an arithmetic average, assign equal weight to every customer segment and would provide a distorted representation of reality.

Table 2 complements these results by providing summary statistics on cost efficiencies and economic weights per X-play pack. We find that 2-play packs, which represent the largest fraction of customer segments (396/1042), show the most variation in efficiency scores, which falls in line with our conclusion from Table 1. Next, we observe that 0-play packs have the lowest average efficiency score. At first sight, this may seem to contradict our findings in Table 1, from which we concluded that these play packs are generally characterized by the least cost reductions. The explanation is that 0-play packs generally have very low economic weights, which appears from the bottom row in Table 2. Figure 4 in the Appendix provides further insights into the variation of efficiency scores per X-play pack (summarized in the form of boxplots).

As a final remark, we note that the average efficiency scores in Table 2 are rather close to one. Given the reported standard deviations, this implies it could be important to control for statistical noise in order to obtain more robust conclusions in terms of inefficiencies. This could be done along the lines of Simar and Zelenyuk (2018) for bias correction or Nguyen, Simar, and Zelenyuk (2022) for data sharpening. Given the focus of the current paper, we leave these interesting extensions for further research.

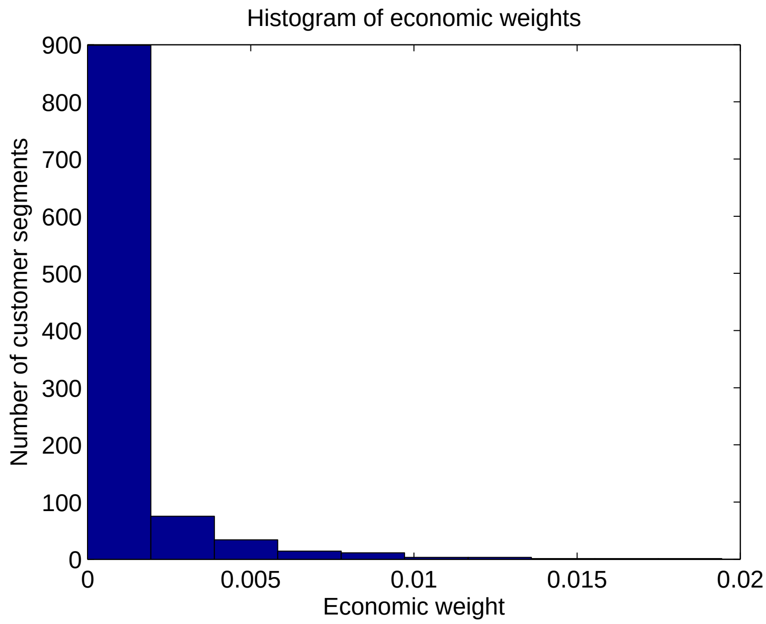

Figure 2: Distribution of economic weights of the customer segments.

Play pack	Number	Mean	Std.	Min	Max
Efficiency score					
4-play	66	0.9967	0.0056	0.9707	1.0000
3-play	264	0.9845	0.0258	0.8521	1.0000
2-play	396	0.9553	0.0568	0.6324	1.0000
1-play	264	0.9770	0.0362	0.8016	1.0000
0-play	52	0.9523	0.0433	0.8379	1.0000
Economic weight					
4-play	66	0.0039	0.0036	0.0002	0.0175
3-play	264	0.0014	0.0024	0.0000	0.0194
2-play	396	0.0005	0.0009	0.0000	0.0092
1-play	264	0.0008	0.0013	0.0000	0.0070
0-play	52	0.0000	0.0000	0.0000	0.0001

Table 2: Summary statistics on efficiency scores and economic weights

5.2 Managerial implications

So far, we have focused on identifying customer segments where potential cost savings could be realized. Of course, only the firm can now analyze if it is indeed doable (and desirable) to realize these cost savings. After all, these are clearly strategic decisions. Below we show how we can guide the management of the firm in this process.

5.2.1 Highest potential cost savings

As the resources to realize the potential cost reductions are limited, managers have to make choices regarding the customer segments on which they will focus. One criterion to determine the customer segments one wants to focus on is the potential cost reduction of the customer segments. Table 3 presents the ten customer segments that have the highest potential cost reduction. These ten customer segments represent a total potential cost reduction of EUR 1 409 309.20, which is 28.19% of the total potential cost reduction identified by our methodology. The results reveal that a lot of cost reductions can be realized in socio segment A among customers that have a 2-play pack. This is not surprising given the pattern established earlier where we found negative net margins per customer for socio segment A. Given the high amount of potential cost reductions that can be realized in these customer segments, the telecom operator should have a close look at these customer segments and question whether and how operational improvements can be made and/or whether these customer segments should be kept in the customer portfolio.

Cost reduction	X-play pack	Socio-segment	Province
289958.29	1-play	A	2
243109.00	2-play	A	2
139757.09	2-play	A	5
127931.43	2-play	A	10
127316.70	2-play	A	9
125255.35	2-play	A	1
106027.85	2-play	A	4
88498.77	2-play	A	8
82489.43	2-play	A	11
79038.53	1-play	A	1

Table 3: Overall top-10 of potential cost reductions accounting for 28.19% of total potential cost reductions.

Table 4 lists the 5 customer segments with the largest potential cost reduction for every play pack. It also shows the total potential cost saving over all play packs as well as the share of these cost savings represented by the top 5 customer segments. Many potential cost savings are located in segment A and E.

The ranking of potential cost reductions of the different customer segments can also be used to verify whether the focus of certain strategic and marketing actions is justified. That is, in some cases, firms decide to focus on particular customer segments because of actions by a competitor or because external events increase the saliency of a particular customer segment. Before investing resources in marketing actions targeted towards these customer segments, it can be useful to examine the amount of potential cost reductions that can be realized in the segments. If the potential cost reduction is high, targeting that

Cost reduction	Socio segment	Province
4-play		
Total cost saving = 318348.08 EUR, top 5 = 46.84%		
Cost reduction	Socio segment	Province
37114.64	E	1
36296.76	A	9
33858.32	E	10
23166.66	E	9
18683.10	E	8
3-play		
Total cost saving = 1286363.75 EUR, top 5 = 20.75%		
Cost reduction	Socio segment	Province
64859.35	E	6
51428.07	A	1
51069.36	E	8
50002.94	A	9
49499.06	A	11
2-play		
Total cost saving = 2236250.49 EUR, top 5 = 34.14%		
Cost reduction	Socio segment	Province
243109.00	A	2
139757.09	A	5
127931.43	A	10
127316.70	A	9
125255.35	A	1
1-play		
Total cost saving = 1141986.41 EUR, top 5 = 48.85%		
Cost reduction	Socio segment	Province
289958.29	A	2
79038.53	A	1
67698.20	A	10
61533.83	A	5
59595.35	A	9
0-play		
Total cost saving = 16374.42 EUR, top 5 = 24.70%		
Cost reduction	Socio segment	Province
1024.49	B	1
869.79	C	9
775.94	B	2
699.00	B	8
675.59	A	2

Table 4: Top 5 of largest potential cost savings per X-play pack

particular customer segment seems warranted and one can do a more extensive analysis of the particular customer segment in order to develop a marketing strategy that also enables realizing the potential cost reductions. If the potential cost reduction is rather low, one should question whether resources should be invested in developing marketing actions targeted towards these particular customer segments.

5.2.2 Output-specific efficiencies

Now that we have identified the customer segments with the largest potential cost reductions, we can dig a bit deeper into the results and explore in which outputs these potential improvements are located. The analysis of the overall top-10 revealed that much of the improvements are situated in 2-play. It would be even more useful for 2-play to identify the specific outputs in which cost improvements are possible. Figure 3 shows histograms of the output-specific cost efficiencies. More specifically it uses our output-specific decomposition to show the frequency of $w_k^m CE_k^m$ on the x-axis and the number of observations on the y-axis. It turns out that a lot of improvement is possible in the "Fixed Access revenues" and "Other revenues" outputs for all 2-play customer segments. The most variation in the efficiency scores is in the "churn rate" output. The remaining outputs have heavy left tails but are less extreme than the "Fixed Access revenues" and "Other revenues" outputs.

Although customer segments can have a similar overall cost efficiency CE_k, their inefficiencies can be located in different outputs. Table 5 illustrates this for three customer segments with overall efficiency scores of approximately 0.8 but with different output-specific efficiency scores. This heterogeneity in output-specific efficiencies across customer segments with similar overall efficiencies likely implies different managerial actions for improvement. This highlights the strength of our methodology that allows for starting with a helicopter view and gradually zooming in to a more detailed level in order to establish managerial implications.

Province	Socio segment	Mobile	Access	Internet	Digital TV	Other	Churn rate	Upsells
9	C	0.7899	0.0000	0.9738	0.8178	0.0000	0.7563	0.0000
6	E	0.8299	0.0000	0.0000	0.5910	0.0000	0.7514	0.0000
6	A	0.8802	0.0000	0.0000	0.0000	0.0000	0.7957	0.0000

Table 5: Output-specific efficiency scores CE_k^m for 2-play customer segments with $CE_k \approx 0.80$

5.2.3 Dominating peers

A core aspect of our methodology is that the input-output performance of a particular customer segment is compared against the input-output performance of other customer segments in the same play pack. This implies that at least one dominating peer customer segment exists for every customer segment with a non-zero potential cost reduction. Such a dominating peer has a better input-output performance, implying that this customer segment uses less inputs for equal (or greater) amounts of the outputs. Analyzing the dominating peer customer segment(s) can be instrumental to guide managers in realizing the potential cost reductions. That is, by considering how inputs are transformed into outputs at the dominating peer, it can become clear how the potential cost reduction of a particular customer segment can be realized. As a specific illustration, we analyze the

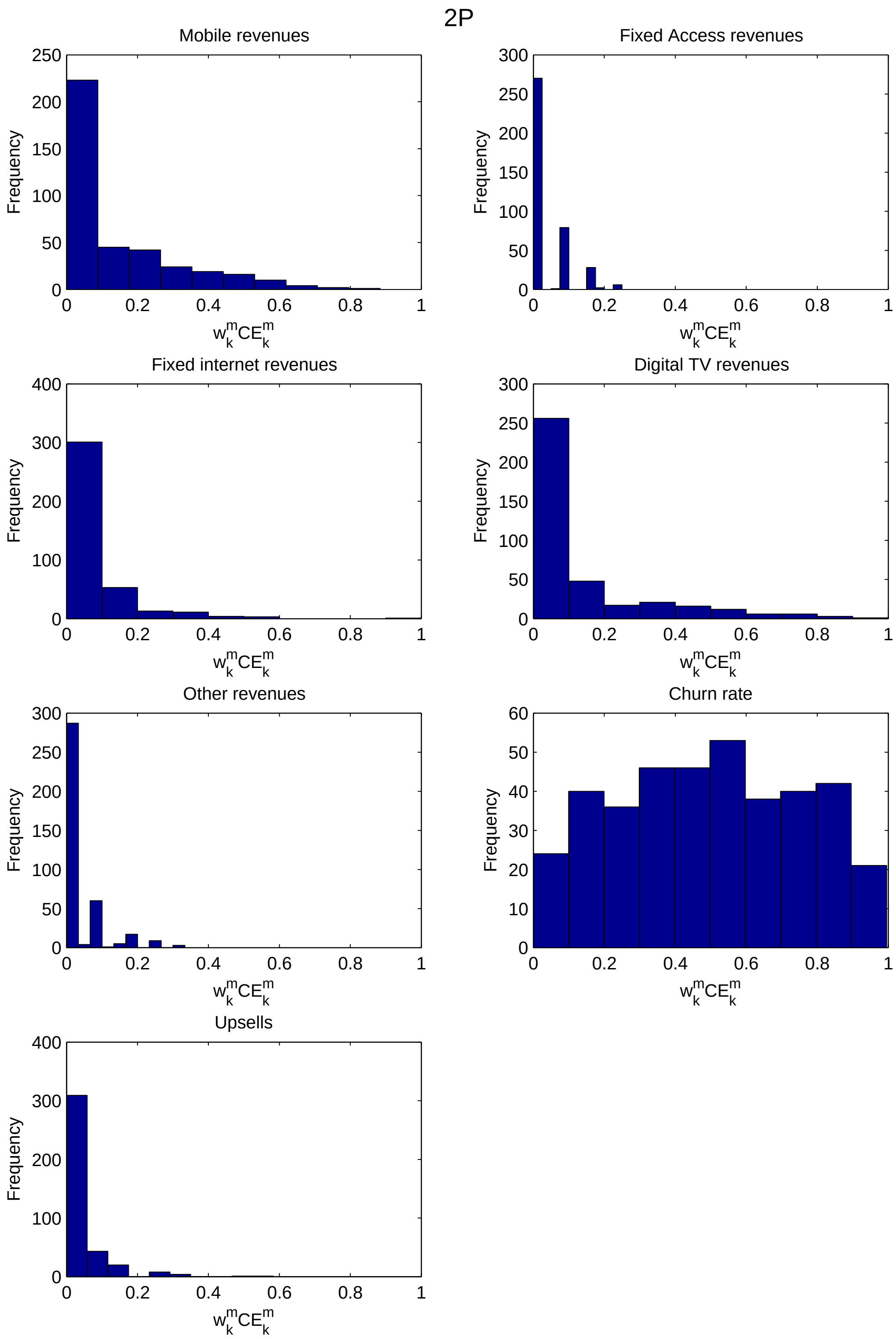

Figure 3: 2-play packs, potential output-specific cost reductions

dominating peers for a specific customer segment in Table 6. The table shows the dominating peer for every output of this specific customer segment. We learn that the Fixed Voice + Fixed Internet and Fixed Voice + Mobile Voice in segment G are dominating on all but one output.

Output	X-play pack	socio segment	province
Mobile revenues	Fixed Voice + Fixed Internet	G	10
Fixed access revenues	Fixed Voice + Mobile Voice	G	4
Fixed internet revenues	Fixed Voice + Fixed Internet	G	3
Fixed TV revenues	Fixed Internet + Fixed dTV	H	11
Other revenues	Fixed Voice + Fixed Internet	G	10
Churn rate	Fixed Voice + Mobile Voice	G	4
Upsells	Fixed Voice + Fixed Internet	G	3

Table 6: Dominating peers characteristics for Fixed Internet + Fixed dTV in socio segment A and province 1

6 Conclusion

We have presented a DEA-based methodology to analyze customer value. We argue that this method provides a useful complement to existing methods such as CLV and CPA, stemming from the distinguishing features of DEA as a nonparametric efficiency evaluation tool. We have demonstrated our newly proposed methodology through an empirical application to customer segments of a large European telecom provider. In this application, we illustrated our method in identifying potential cost reductions as well as alternative managerial applications.

The practical relevance of our DEA-based methodology stems from the fit between our methodology and the importance of customer centricity in today's business environment. A first important aspect of customer centricity is that the way in which customer segments are served needs to be tailored to the characteristics of the customer segment. Our DEA-based methodology allows to incorporate heterogeneity in the way customer segments are served. Using our DEA-based methodology will thus provide managers with more realistic insights into the improvement potential of customer segments allowing managers to take better decisions regarding for instance the allocation of marketing resources and pricing.

A second important aspect of today's customer-centric business environment is that an enormous amount of data about customers is available. For instance, firms collect data about the revenues that customers generate, customer complaints, website visits, as well as behavior on social media such as whether customers speak positively about the firm. However, these data are not always expressed in monetary terms, making them difficult to use for existing methodologies such as CLV and CPA. Our DEA-based methodology can deal with data that vary with respect to the unit of account and does not require that data are expressed in, or are transformed into, monetary terms. By including more diverse data, we believe that our DEA-based methodology can generate insights regarding the improvement potential of customer segments that are difficult to generate from existing methodologies such as CLV and CPA.

We see multiple avenues for future research. At the methodological level, one can integrate into our framework the many existing theoretical insights from the DEA literature in order to better grasp specific features of the business environment. A notable natural next step is to focus on the dynamic or intertemporal aspects of many managerial decisions. The detailed nature of the available ABC data in many big firms and the existing theoretical results could lead to several novel (empirical) insights. See, for instance, Färe and Grosskopf (1996) for a thorough overview, and Cherchye, De Rock, and Kerstens (2018) for a recent contribution that is closely related to the DEA approach that we followed in the current paper.

Next, to keep our exposition simple and to focus our discussion, we have abstracted from properly accounting for specific features of data generating processes that may often complicate empirical efficiency analyses. Correspondingly, we have not explicitly considered statistical inference and the integration of environmental variables in practical applications. We refer to the extensive and very active DEA literature on these different aspects. See, for example, Daraio and Simar (2007) and, more recently, Simar and Wilson (2020) for reviews. The tools and insights that have been developed in this existing work can be readily adapted to the framework that we set out in the previous sections.

Finally, at the application level, we look forward to applications of our DEA-based methodology in other environments that adapt the way customers are served based on characteristics of the customer segment. Given that customer centricity seems to be an imperative in both the for-profit and not-for-profit environment, we believe that our DEA-based methodology has potential to be applied in, for instance, online retailing, banking, health care, and education. Applications in other environments should be encouraged not only because they will help to discover the usefulness and boundaries of our DEA-based methodology but also because such applications can help to develop methodological refinements which may inspire the DEA literature.

References

Bonacchi, M. and P. Perego (2019): "Customer accounting: creating value with customer analytics," *SpringerBriefs in Accounting*, Edition 1, 1 – 87.

Charnes, A., W. W. Cooper, and E. Rhodes (1978): "Measuring the efficiency of decision making units," *European journal of operational research*, 2, 429–444.

Cherchye, L., B. De Rock, and P. J. Kerstens (2018): "Production with storable and durable inputs: nonparametric analysis of intertemporal efficiency," *European journal of operational research*, 270, 498–513.

Cherchye, L., B. De Rock, and F. Vermeulen (2008): "Analyzing cost-efficient production behavior under economies of scope: A nonparametric methodology," *Operations Research*, 56, 204–221.

Cherchye, L., B. De Rock, and B. Walheer (2015): "Multi-output efficiency with good and bad outputs," *European journal of operational research*, 240, 872–881.

——— (2016): "Multi-output profit efficiency and directional distance functions," *Omega*, 61, 100–109.

Cherchye, L., W. Moesen, N. Rogge, and T. Van Puyenbroeck (2007): "An introduction to 'benefit of the doubt'composite indicators," *Social Indicators Research*, 82, 111–145.

Cherchye, L., B. D. Rock, B. Dierynck, F. Roodhooft, and J. Sabbe (2013): "Opening the black box of efficiency measurement: input allocation in multi-output settings," *Operations research*, 61, 1148–1165.

Cook, W. D. and L. M. Seiford (2009): "Data envelopment analysis (DEA)–Thirty years on," *European journal of operational research*, 192, 1–17.

Daraio, S. and L. Simar (2007): *Advanced Robust and Nonparametric Methods in Efficiency Analysis: Methodology and Applications*, Springer.

Fader, P. (2020): *Customer centricity: Focus on the right customers for strategic advantage*, Wharton Digital Press: Philadelphia.

Färe, J. and S. Grosskopf (1996): *Intertemporal Production Frontiers: With Dynamic DEA*, Kluwer Academic Publishers, London.

Galbraith, j. (2005): *Designing the customer-centric organization. A guide to strategy, structure, and process*, Jossey-Bass: San Francisco.

Glady, N., B. Baesens, and C. Croux (2009): "Modeling churn using customer lifetime value," *European Journal of Operational Research*, 197 (1), 402 – 411.

Holm, M., V. Kumar, and C. Rohde (2012): "Measuring customer profitability in complex environments: an interdisciplinary contingency framework." *Journal of the Academy of Marketing Science*, 40, 387 – 401.

Jain, D. and S. S. Singh (2002): "customer lifetime value research in marketing: a review and future directions." *Journal of Interactive Marketing (John Wiley & Sons)*, 16, 34 – 46.

Kumar, V. and D. Shah (2009): "Expanding the Role of Marketing: From Customer Equity to Market Capitalization." *Journal of Marketing*, 73, 119 – 136.

Latinovic, Z. and S. Chatterjee (2019): "Customer centricity in the digital age," *MIT Sloan Management Review*, 60 (4), 1 – 2.

Nguyen, B. H., L. Simar, and V. Zelenyuk (2022): "Data sharpening for improving CLT approximations for DEA-type efficiency estimators," *European Journal of Operational Research*, 303, 1469–1480.

Palmatier, R., C. Moorman, and J.-Y. Lee (2019): *Handbook on customer centricity: strategies for building a customer-centric organization*, Edward Elgar Publishing: Cheltenham United Kingdom.

Ruggiero, J. (1996): "On the measurement of technical efficiency in the public sector," *European Journal of Operational Research*, 90, 553–565.

Simar, L. and P. Wilson (2020): "Technical, allocative and overall efficiency: Estimation and inference," *European Journal of Operational Research*, 282, 1164–1176.

Simar, L. and V. Zelenyuk (2018): "Central Limit Theorems for Aggregate Efficiency," *Operations research*, 166, 139–149.

Tone, K. and M. Tsutsui (2009): "Network DEA: A slacks-based measure approach," *European Journal of Operational Research*, 197, 243–252.

van Raaij, E. M., M. J. Vernooij, and S. van Triest (2003): "The implementation of customer profitability analysis: A case study," *Industrial marketing management*, 32 (7), 573 – 583.

Zhu, J. (2015): *Data Envelopment Analysis a Handbook of Models and Methods*, Springer.

Appendix: Additional data and results

Input	Mean	Std	Min	Max
ICX_COst	126534,7	558300,1	0	7691876
ROAM_COST	5983,757	21220,79	0	263047,4
COGS_CONTENT	115619,1	334864,7	1,5956	3136804
OPEX_BILLING	17124,49	38740,08	0,456	363408,3
OPEX_BAD_DEBT	16094,74	34516,34	-0,8384	345770,6
OPEX_IT	11929,29	39141,26	-160,649	411506,4
OPEX_REPAIR	114447,1	263642,8	0	2061820
OPEX_CPE	51,11962	222,2961	-172,135	3210,98
OPEX_CCA	77834,9	173275,8	1,2973	1772943
OPEX_OWN_SHOPS	19497,31	45203,92	0,4157	470081,2
OPEX_ECH	6281,529	14644,32	0,1264	152493,6
OPEX_COMMISSIONS	6869,482	18229,62	0,0905	173461,4
SAC_CHANNEL	45381,83	181909,1	0	2315726
SAC_INSTALL	26449,44	114983,7	0	1955994
SAC_TERMINAL	19450,96	96713,02	0	1771770
SAC_CCA_BACK_OFFICE	4152,572	19661,31	0	374602,8
SDC_CHANNEL	24048,75	68878,44	0	983809,9
SDC_INSTALL	18407,64	39758,43	0	445857,9
SDC_TERMINAL	14045,75	34255,48	0	383237,4
SDC_CCA_BACK_OFFICE	2617,542	6804,588	0	80528,87
Output	Mean	Std	Min	Max
Mobile revenues	404850,2	1394090	-77,3397	15637947
Fixed access revenues	396283,5	1167457	-605,956	13472855
Fixed internet revenues	351610,5	867189,8	-68,2616	8210710
Fixed TV revenues	213808,5	617779,2	-187,173	5921540
Other revenues	2831,775	8740,084	-415,099	139795,4
Churn rate	-0,13577	0,13623	-0,48667	0
Upsells	29919,82	71029,37	0	603054
Number of customers	30826,83	72117,02	1	620704

Table 7: Summary statistics of costs and revenues

	4-play		3-play		2-play		1-play		0-play	
Input	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
Repair costs	454499.29	506769.11	162391.00	307091.52	67038.17	121062.10	78552.40	240117.62	1292.63	1286.11
Rental costs devices	143.31	435.27	68.27	281.72	44.28	185.36	30.44	111.65	11.82	119.46
Call center costs	341731.08	392709.48	99664.88	180986.28	32838.94	56822.25	74216.30	145802.74	4032.03	3430.03
Shops costs	90802.42	104468.34	26351.87	48045.70	7950.84	14812.51	16251.70	34125.03	1089.01	930.04
Web costs	29349.02	33748.44	8493.28	15425.38	2534.66	4817.63	5244.59	11334.93	365.49	314.28
Commissions	32187.31	37880.92	6423.97	9096.37	1811.28	3000.56	9934.00	25645.21	844.17	719.37
SAC CHANNEL	84181.39	132197.61	53732.74	215607.91	15206.82	61934.33	83371.60	268187.65	0.00	0.00
SAC INSTALL	48508.31	76176.96	44440.04	181988.87	18849.24	77574.83	20579.95	91401.23	0.00	0.00
SAC TERMINAL	43939.58	69002.28	39712.93	164955.68	15039.30	69769.69	4148.08	24514.12	0.00	0.00
SAC CCA BACK OFFICE	9518.34	14947.51	8175.63	34831.72	2501.36	10691.17	2211.08	7725.90	0.00	0.00
SDC CHANNEL	197355.11	183100.06	32953.60	40251.84	7628.21	11691.40	1496.62	3466.20	0.00	0.00
SDC INSTALL	105024.60	88256.50	28910.82	36839.06	10537.02	18080.53	2113.03	8069.98	0.00	0.00
SDC TERMINAL	90139.03	75916.81	22959.44	32025.80	6450.71	15284.73	520.11	2204.44	0.00	0.00
SDC CCA BACK OFFICE	18854.18	15833.80	4141.30	6092.69	898.36	1493.36	165.50	665.20	0.00	0.00
Output	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
Mobile revenues	1609028.16	1932590.72	211823.12	336505.18	49896.88	128735.41	915997.39	2417011.43	35109.44	27414.26
Fixed access revenues	1294707.19	1497576.23	463405.67	1019595.33	205638.11	470249.40	482764.07	1794354.21	3676.24	3428.55
Fixed internet revenues	1485713.15	1703256.65	592675.31	1147821.29	225368.95	491208.91	94561.38	300294.55	4188.88	4301.56
Fixed TV revenues	1119748.99	1282514.40	394958.02	809351.93	111568.24	309895.42	3400.10	5195.58	10101.49	12169.81
Other revenues	21045.20	25447.77	3693.96	6200.80	685.34	1358.57	1133.67	2469.05	412.51	570.50
Churn rate	-0.03	0.00	-0.06	0.04	-0.09	0.09	-0.22	0.10	-0.49	0.00
Upsells	59741.65	68229.48	24126.11	45126.88	12371.78	21999.89	60341.74	120496.68	6058.38	6008.06
Number of customers	60993.89	68906.95	24991.31	46034.43	12821.23	22390.94	61963.66	122203.20	6647.06	6132.40

Table 8: Summary statistics of controllable costs and revenues per X-play pack

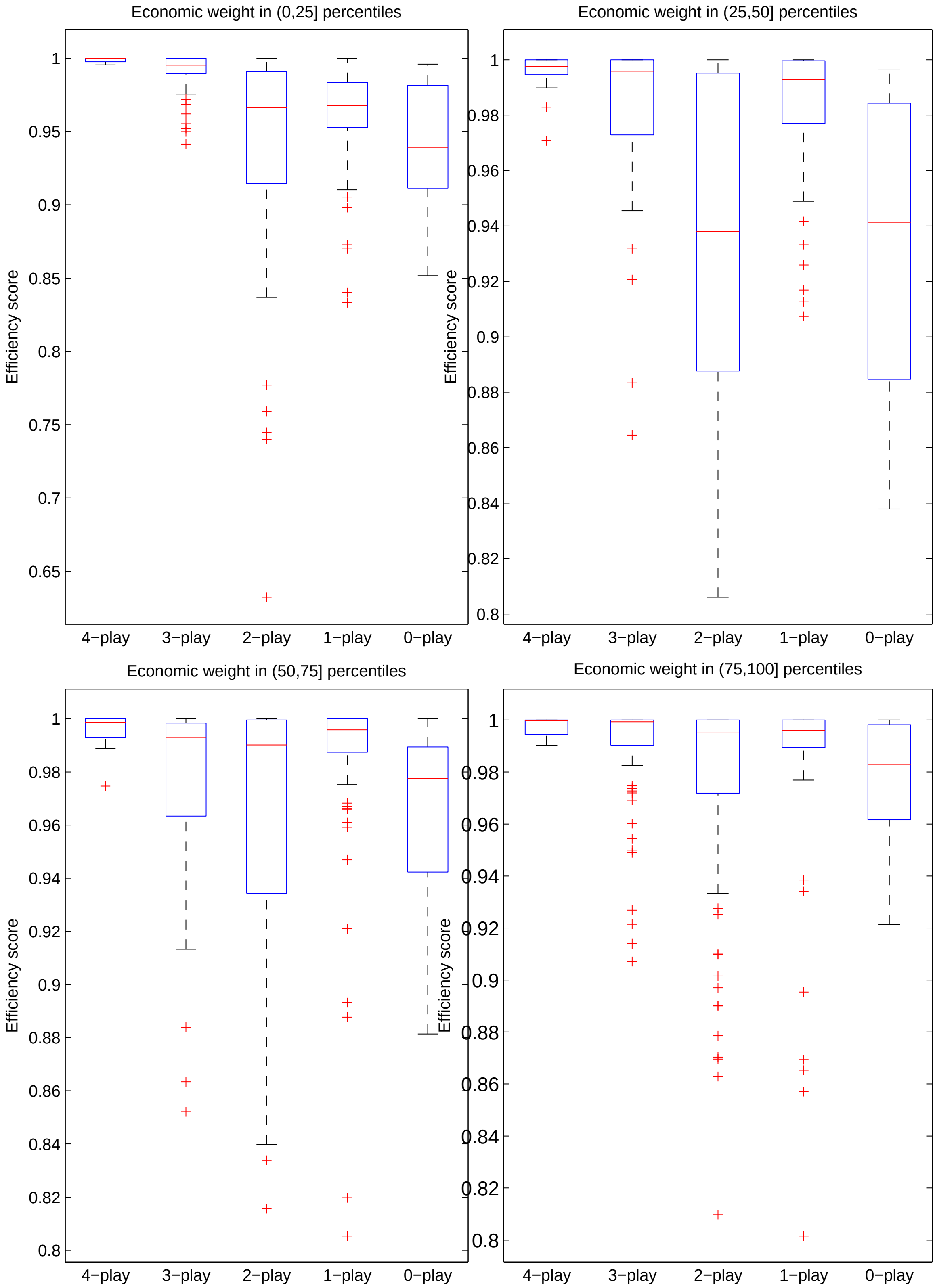

Figure 4: Boxplot of efficiency scores for different quartiles of observations according to economic weight.