POLITECNICO DI TORINO

Master's Degree Course in Data Science and Engineering

Master's Degree Thesis

Counterfactual Analysis and DEA-Based Benchmarking: Explaining ESG Performance in the Logistics Sector

Supervisors
Prof. Sara Khodaparasti
Prof. Daniele Apiletti

Candidato
Fatemeh Jamalimoghaddam

December 2025

Acknowledgements

I would like to express my heartfelt gratitude to everyone who supported me throughout this journey. My deepest thanks go to my supervisors, Prof. Sara Khodaparasti and Prof. Daniele Apiletti, whose guidance, patience, and encouragement shaped this thesis and strengthened my confidence as a researcher.

I am profoundly grateful to my partner, Siavash, whose unwavering presence has been my greatest source of strength. From the moment we began this journey together and stepped into the unknown, his kindness, patience, and constant encouragement have supported me through every challenge and uncertainty. I could not have reached this point without his belief in me.

My love and appreciation go to my family and friends, whether near or far. Their support, through their presence, their messages, and their constant encouragement,carried me through every stage of this journey. Being far from home has not always been easy, but their love and belief in me made the distance feel lighter and gave me the strength to continue.

Finally, I extend my thanks to all who, in ways big or small, contributed to this work. This thesis is not only my effort but also a reflection of every hand that lifted me up along the way.

Abstract

Enhancing the performance of institutions and other Decision Making Units (DMUs) is among the primary objectives in efficiency analysis. Conventional Data Envelopment Analysis (DEA) models offer useful tools for measuring relative efficiency, but such targets are sometimes unrealistic from a behavior-theoretical point of view, since they can be mathematically optimal but impossible to reach empirically.

This thesis presents a more powerful benchmark employing DEA in combination with counterfactual reasoning within a Bilevel Optimization (Bilevel) optimization framework. Our approach finds the smallest and most realistic changes to inputs or outputs that would make an inefficient unit efficient, which turns abstract efficiency scores into interpretable and actionable improvements.

In higher level, the model controls the sparsity of changes and magnitude and smoothness based on observed data to deviation, while standard DEA efficiency constraints are set in lower level.The resulting model also fits within the larger context of interpretable and explainable analytics, following Explainable Artificial Intelligence (XAI) principles by providing transparent explanations that can be understood by humans.

We apply the model to a case from the logistics industry, focusing on environmental, social and governance (ESG) factors in order to demonstrate its potential for eked out overall efficiency improvements derived from sustainability issues. It is shown that the counterfactual DEA provides a new way to understand efficiency and enhances the decision relevance of efficiency analysis compared with the traditional optimization perspective, promoting a shift from a rich yet opaque optimization ideology toward a more interpretable analytical framework.

Table of Contents

List of Tables VI

List of Figures VII

Acronyms VIII

1 Introduction 1

2 Literature Review 6

3 Theoretical Background: DEA and Bilevel Optimization 10
- 3.1 Data Envelopment Analysis (DEA) . . . 10
  - 3.1.1 Returns to Scale Assumptions: CRS and VRS . . . 11
  - 3.1.2 The CCR Model (CRS) . . . 12
  - 3.1.3 The BCC Model (VRS) . . . 13
  - 3.1.4 Dual (Multiplier) Form . . . 13
  - 3.1.5 The Data Envelopment Analysis (DEA) model used in this research . . . 15
- 3.2 Bilevel Optimization . . . 16
  - 3.2.1 General Formulation . . . 17
  - 3.2.2 KKT Reformulation . . . 17
  - 3.2.3 Cost Functions for Counterfactual Explanations . . . 18
  - 3.2.4 Interpretability and Actionability in DEA . . . 19
  - 3.2.5 Applications in DEA and Counterfactual Benchmarking . . . 19
  - 3.2.6 Example Formulation for Counterfactual DEA . . . 20

4 Proposed Methodology 21
- 4.1 Data and variables . . . 21
  - 4.1.1 Data Preprocessing . . . 22
  - 4.1.2 Final Dataset . . . 25
- 4.2 Problem Statement and Model Rationale . . . 26

4.3 Modeling Framework 26
4.3.1 Implementation Overview 29

5 Results Discussion, Conclusion and Future Work 30
5.1 The Data 31
5.2 Counterfactual analysis for one firm: Santos Brasil Participacoes SA 32
5.3 Aggregate Counterfactual Results across All Firms 38
5.4 Conclusion 45
5.5 Future Work 46

Bibliography 48

List of Tables

4.1 Variables used in the study and their modelling roles. 25

5.1 Descriptive statistics of the normalized ESG dataset (scaled to [0,1]). 31
5.2 Original DEA efficiency distribution. 32
5.3 Weight configurations (ν_0, ν_1, ν_2) used for counterfactual cost functions. 34
5.4 Signed input changes required under different targets and configurations. 34
5.5 Counterfactual targets for Santos Brasil Participacoes SA at $E^* = 0.8$. 35
5.6 Counterfactual targets for Santos Brasil Participacoes SA at $E^* = 1.0$. 35
5.7 Combinations of cost-function parameters (ν_0, ν_1, ν_2) used for aggregate counterfactual analysis. 39
5.8 Share of firms (%) for which each input changes when desired efficiency is $E^* = 1.0$. 40
5.9 Average absolute change by input when desired efficiency is $E^* = 1.0$. 40
5.10 Average number of inputs that change per firm at $E^* = 1.0$. 40
5.11 Share of firms (%) for which each input changes when desired efficiency is $E^* = 0.8$. 44
5.12 Average absolute change by input when desired efficiency is $E^* = 0.8$. 44
5.13 Average number of inputs that change per firm at $E^* = 0.8$. 44

List of Figures

5.1 Grouped bar chart at $E^* = 0.8$: Original, Farrell, and counterfactual targets. 36
5.2 Grouped bar chart at $E^* = 1.0$: Original, Farrell, and counterfactual targets. 36
5.3 Inputs that change for individual firms under $E^* = 1.0$ across the three cost functions. 42
5.4 Inputs that change for individual firms under $E^* = 0.8$ across the three cost functions. 43

Acronyms

DEA
: Data Envelopment Analysis

DMU
: Decision Making Unit

ESG
: Environmental, Social and Governance

VRS
: Variable Returns to Scale

BCC
: Banker–Charnes–Cooper model

CRS
: Constant Returns to Scale

MILP
: Mixed-Integer Linear Programming

MIQP
: Mixed-Integer Quadratic Programming

KKT
: Karush–Kuhn–Tucker

IQR
: Interquartile Range

MIPGap
: Mixed-Integer Programming relative optimality gap

CCR
: Charnes–Cooper–Rhodes model

LP
: Linear Programming

Bilevel
: Bilevel Optimization

MPCC
: Mathematical Program with Complementarity Constraints

CEDEA
: Counterfactual Explanation in Data Envelopment Analysis

NP
: Nondeterministic Polynomial time

OR
: Operations Research

XAI
: Explainable Artificial Intelligence

SBM
: Slack-Based Measure

DDF
: Directional Distance Function

AI
: Artificial Intelligence

ML
: Machine Learning

CF

Counterfactual Explanation

MCDM

Multiple Criteria Decision Making

QR

Quadratic Regularization

Chapter 1

Introduction

Motivation and Context

The growing availability of high-dimensional data has transformed performance evaluation into a data-driven science. Organizations, public institutions, and firms increasingly rely on advanced analytics and Artificial Intelligence (AI) tools to assess, predict, and optimize their operations. Yet, as AI systems become more pervasive, they also become more opaque. Many modern models achieve impressive predictive accuracy but offer limited transparency regarding why and how specific decisions are made. This lack of interpretability can hinder trust, adoption, and the translation of analytical insights into actionable managerial knowledge.

In the domain of efficiency and productivity analysis, DEA stands out as one of the earliest and most influential forms of interpretable AI. Since its introduction by Charnes et al. [1], DEA has offered a transparent, optimization-based method for benchmarking Decision Making Units (DMUs) using multiple inputs and outputs without requiring a predefined functional form. Over the past four decades, DEA has evolved into a cornerstone of quantitative performance assessment, applied to diverse sectors such as banking, healthcare, energy, education, and sustainability [2]. Its appeal lies in its empirical nature: DEA constructs a data-driven production frontier and measures efficiency relative to observed best practices.

However, as Bogetoft et al. [3] point out, the interpretability of DEA is limited when communicating results to non-technical decision makers. The efficiency score alone provides little insight into how a specific DMU can realistically improve. Managers often ask not only "how efficient are we?" but also "what should we do differently to become efficient?" Answering this requires moving from evaluation toexplanation—from descriptive analytics toward prescriptive and interpretable AI.

From DEA Benchmarking to Counterfactual Reasoning

In recent years, the field of AI has witnessed a paradigm shift toward explainability and human-centric reasoning. The rise of XAI stems from the need to make complex models—such as deep neural networks—more transparent and trustworthy [4, 5]. Among XAI approaches, Counterfactual Explanation (CF) have gained particular prominence because they mirror human reasoning: they describe what minimal change in the input would have led to a different outcome. For instance, in a loan decision model, a counterfactual explanation might say, "if the applicant's income were $10,000 higher, the loan would be approved." Such statements transform algorithmic predictions into actionable knowledge.

Bogetoft et al. [3] introduced a similar philosophy into DEA by proposing the concept of Counterfactual Explanation in Data Envelopment Analysis (CEDEA). Rather than simply identifying efficient peers, the model computes alternative input–output configurations that are close to the current profile of an inefficient DMU but would make it efficient. This approach unifies the interpretability of counterfactual AI with the economic rigor of DEA. The result is a framework that not only measures efficiency but also prescribes how to achieve it.

In practical terms, counterfactual DEA formulates a Bilevel optimization problem.The lower level corresponds to the standard DEA model that assesses technical efficiency at a given return to scale—either under Constant Returns to Scale (CRS) or Variable Returns to Scale (VRS)—whereas the upperstage balances the cost or effort needed to get from where we are now to an alternative position. The distance is computed by mixtures of ℓ_0, ℓ_1 and ℓ_2 norms resulting in flexible trade-offs between sparsity (fewest changes), magnitude (smallest amount by which the parameters should be adjusted) and realism (smooth and plausible transitions).This formulation parallels recent developments in interpretable Machine Learning (ML), where counterfactuals are used to explain black-box models such as neural networks, support vector machines, or tree ensembles [6, 4].

Problem Statement and Research Gap

While classical DEA provides valuable efficiency benchmarks, it lacks behavioral realism: the recommended targets may be mathematically valid but economically or operationally infeasible. At the same time, most XAI frameworks—though strong in interpretability—focus on classification and prediction tasks, not on multi-input multi-output production systems. The intersection of these domains remains underexplored.

There is thus a growing need for methods that combine:

- the theoretical robustness and interpretability of DEA,
- the explanatory power of counterfactual AI, and
- the optimization precision of modern data science.

Integrating DEA with counterfactual reasoning allows the model to answer the "what-if" questions that traditional efficiency analysis cannot. Instead of producing static efficiency scores, it generates dynamic, scenario-based insights: what minimal, data-supported changes can make an inefficient firm reach the frontier? This shift from evaluation to simulation aligns DEA with contemporary trends in data science and AI, where interpretability, fairness, and algorithmic recourse are central concerns [5, 4].

Yet, practical implementations of counterfactual DEA are scarce. Existing studies mainly focus on theoretical formulations or small illustrative datasets. Moreover, there is limited empirical evidence on how the integration of Bilevel optimization and counterfactual reasoning performs on complex, real-world data—such as Environmental, Social and Governance (ESG) indicators or industrial performance metrics—where interpretability and accountability are crucial. This thesis addresses these gaps by operationalizing the counterfactual DEA model within a modern AI pipeline.

Objectives and Contributions

The main objective of this research is to develop and apply a counterfactual-based DEA framework that combines the transparency of Operations Research (OR) models with the interpretability standards of modern AI. By integrating explainable and data-driven principles into efficiency benchmarking, the thesis aims to enhance both methodological rigor and practical decision support. Specifically, it seeks to:

1. Formulate an interpretable DEA model incorporating counterfactual reasoning through Bilevel optimization.
2. Implement the model using Python and Gurobi, integrating data preprocessing, scaling, and visualization within a data science workflow.
3. Evaluate the model on real datasets, such as ESG performance indicators, to derive meaningful and actionable efficiency improvements.
4. Experiment with different regularization schemes (ℓ_0, ℓ_1, and ℓ_2) in the counterfactual optimization to evaluate trade-offs between sparsity, total adjustment, and smoothness—key factors influencing interpretability and realism in data-driven target setting.

This combination of DEA, counterfactual reasoning, and XAI positions the work at the intersection of optimization and data science, contributing to the development of transparent, explainable, and accountable AI for performance analysis.

By combining principles from OR, optimization, and XAI, this work contributes to the emerging field of transparent decision analytics. The proposed framework bridges the gap between prescriptive modeling and human understanding, enabling decision makers to explore multiple realistic pathways toward efficiency improvement.

Scientific and Practical Relevance

From a scientific perspective, this study extends the frontier of DEA research toward hybrid AI systems. It demonstrates that counterfactual reasoning, originally developed for interpretable ML, can enrich performance evaluation by providing local, human-understandable explanations. It also contributes to the broader debate on how optimization models can support ethical and accountable AI—an issue increasingly recognized in European and international regulatory frameworks.

From a practical standpoint, organizations are under pressure to justify their decisions and sustainability claims using transparent and explainable analytics. In this context, counterfactual DEA offers a decision-support tool that transforms abstract efficiency metrics into concrete managerial actions. It enables firms to understand not only their relative position but also the feasible steps required to reach best practice, all while maintaining interpretability and auditability.

Structure of the Thesis

The thesis is organized as follows:

- Chapter 1 gives an introduction to the research work, motivation and objectives of this study and presents the research problem as well as the overall thesis structure.
- Chapter 2 presents a review of the literature on DEA, target setting, and counterfactual and interpretability techniques, situating our work within this wider established context.
- Chapter 3 presents the backgrounds of DEA and Bilevel optimization for theoretical preparation in our model.
- Chapter 4 presents the proposed counterfactual DEA framework, detailing its mathematical formulation, model design, and computational implementation.

- Chapter 5 discusses the empirical application of the model, analyzes the results, and provides managerial insights in addition to the conclusion and future work.

Through this structure, the thesis aims to advance DEA from a traditional benchmarking technique to a human-centered, interpretable AI framework—one capable of transforming data-driven evaluation into transparent, actionable, and explainable decision support.

Chapter 2

Literature Review

This chapter reviews the theoretical and empirical foundations that motivate the development of the proposed counterfactual–based DEA framework. It traces the evolution of efficiency analysis from early concepts of productive efficiency to modern extensions of DEA, highlighting the progression from measurement to explanation. The chapter then explores the growing intersection between optimization, counterfactual reasoning, and XAI, emphasizing how interpretability and realism have become central goals in data science. The review concludes by identifying the research gaps that this thesis seeks to address.

From Productive Efficiency to Data Envelopment Analysis

The origin of efficiency measurement dates to Farrell [7], who defined efficiency as the ratio between observed performance and the theoretical best practice. Farrell's geometric framework, representing a production frontier through isoquants, established the conceptual foundation for modern nonparametric benchmarking.

Charnes et al. [1] translated this concept into an operational method through DEA, introducing the Charnes–Cooper–Rhodes model (CCR) model under constant returns to scale. By constructing an empirical frontier enveloping all observed DMUs, DEA measures relative efficiency without assuming a specific functional form. The subsequent Banker–Charnes–Cooper model (BCC) model by Banker et al. [8] relaxed the scale assumption, allowing variable returns and thereby improving realism in empirical studies.

DEA's flexibility led to widespread adoption across public and private sectors, supported by the comprehensive expositions of Cooper et al. [9] and Cooper et al. [2]. These works systematized extensions such as additive models, slacks-based measures, and directional distance functions. Recent bibliometric analyses confirm

DEA's continued growth, identifying new trends such as dynamic and network DEA, environmental efficiency, and robust formulations that incorporate uncertainty and noise [10]. Collectively, these developments portray DEA as one of the most enduring and adaptive tools in quantitative performance analysis.

Target Setting and Distance–Based Extensions

Classical DEA identifies efficiency scores but does not specify the minimal changes required for an inefficient unit to reach the frontier. Research in the early 2000s addressed this limitation by reformulating DEA as a distance-minimization problem. Aparicio et al. [11] proposed the "closest-target" model, which minimizes the distance from each inefficient unit to the efficient frontier, producing realistic and attainable benchmarks. Related works introduced multiple distance norms and penalty structures to ensure the suggested adjustments remained feasible and interpretable.

Parallel to distance approaches, the Slack-Based Measure (SBM) family of models emerged, emphasizing input and output slacks rather than aggregate distance. Surveys of SBM developments show how inefficiency can be decomposed into factor-specific components and integrated with undesirable outputs [12]. These contributions collectively pushed DEA from descriptive assessment toward prescriptive, action-oriented modeling—an intellectual trajectory that foreshadows the counterfactual perspective explored later in this thesis.

Optimization Perspectives and Bilevel Formulations

DEA naturally aligns with mathematical programming and optimization theory. The field of Bilevel optimization, formalized by Bard [13] and Dempe [14], provides the theoretical foundation for hierarchical problems in which one optimization task is nested within another. Such problems capture leader–follower or upper–lower decision structures and are inherently nonconvex and computationally challenging. Later reviews, including Sinha et al. [15], categorize bilevel methods into classical, evolutionary, and hybrid approaches, emphasizing their applicability to nonconvex, multiobjective problems.

These advances have enabled new formulations of DEA that explicitly integrate behavioral realism, regulatory goals, and multi-level decision processes. The same mathematical logic underpins the recent wave of counterfactual DEA models, in which an upper-level problem minimizes deviation from observed data while a lower-level problem enforces efficiency constraints.

Counterfactual Analysis and Target Setting in DEA

The first studies to incorporate counterfactual reasoning in DEA dates back to Bogetoft et al. [3]. Their model, called Counterfactual DEA, casts improving the efficiency in terms of identifying a nearby point in input–output space that would make an inefficient DMU efficient. The objective function is a mixture of the ℓ_0, ℓ_1, and ℓ_2 norms to control trade-offs between sparsity, total adjustment, and smoothness. This changes DEA from being a static benchmarking mechanism to an interpretive and normative system, which is able to provide actionable recommendations.

Building on this foundation, Carrizosa et al. [6] interpreted counterfactual target setting within the broader context of mathematical optimization. They framed counterfactual explanations as constrained optimization problems seeking minimal perturbations that reverse model outcomes, thereby unifying perspectives from economics, machine learning, and operations research. These developments converge on a shared goal: to generate small, plausible, and interpretable adjustments that link quantitative results with managerial insight.

Counterfactual Explanations and Explainable Artificial Intelligence

Beyond efficiency analysis, counterfactual reasoning has become central to XAI. A counterfactual explanation specifies the minimal change to an input that would alter a model's output, offering intuitive "what-if" reasoning. Guidotti [4] provide one of the most extensive surveys of counterfactual methods, showing that validity, actionability, sparsity, and plausibility often trade off against each other. Verma et al. [5] review over 350 algorithms and emphasize counterfactuals' role in algorithmic recourse, fairness, and transparency.

Recent research extends these ideas toward robustness and causality. The survey by Delaney and Greene [16] synthesizes strategies for ensuring that counterfactuals remain stable under data perturbations, while Hancox-Li and Lipton [17] discuss the distinction between mere correlation-based counterfactuals and truly causal explanations. Ethical considerations have also emerged: Kusner et al. [18] warn that ill-defined counterfactuals can misrepresent feasible actions, especially when sensitive attributes are involved.

Another growing theme concerns user perception. Empirical studies such as Poyiadzi et al. [19] show that simplicity and interpretability strongly influence human trust in counterfactual explanations. These insights are particularly relevant

when counterfactual principles are applied to management and policy contexts, where interpretability and realism are as critical as mathematical validity.

Extensions of DEA and Opportunities for Integration

Parallel to developments in explainability, the DEA community has explored numerous methodological extensions. Reviews such as Kao [20] document the rise of network DEA, which models multi-stage and interlinked processes; others emphasize dynamic and stochastic DEA, addressing temporal change and statistical noise. These advances broaden DEA's applicability but rarely address interpretability explicitly. Integrating counterfactual reasoning with such models could enhance their transparency and decision relevance.

Moreover, hybrid approaches combining data-driven learning and optimization logic are gaining prominence in AI. The concept of neurosymbolic AI—surveyed by Kosasih et al. [21]—illustrates how neural networks can capture complex patterns while symbolic components maintain interpretability. Counterfactual DEA aligns with this hybrid vision: it retains optimization rigor while providing interpretable, localized explanations, effectively bridging OR and AI.

Identified Research Gap

Despite the conceptual alignment between DEA, counterfactual reasoning, and XAI, practical implementations of counterfactual DEA remain scarce. Existing studies mainly demonstrate feasibility on small synthetic datasets, leaving empirical validation on complex, high-dimensional data largely unexplored. Few works systematically analyze how different regularization norms (ℓ_0, ℓ_1, ℓ_2) affect interpretability, realism, or the managerial usefulness of the recommended adjustments.

Additionally, connections with broader XAI themes—robustness, causability, and human perception—are seldom examined in efficiency analysis. Bridging these perspectives offers a promising avenue for advancing both methodological rigor and interpretability in performance benchmarking. The framework proposed in this thesis addresses these gaps by integrating norm-regularized counterfactual optimization with DEA, emphasizing transparent, realistic, and data-driven target setting.

Chapter 3

Theoretical Background: DEA and Bilevel Optimization

This chapter presents the theoretical foundations that underpin the modeling framework proposed in this thesis. It is structured in two main parts: the first introduces DEA, a method for measuring the relative efficiency of DMUs; the second explores Bilevel optimization, a powerful framework for modeling hierarchical decision processes. The integration of these two approaches forms the basis for counterfactual-based target setting, as proposed in the CEDEA model discussed in the next chapter.

3.1 Data Envelopment Analysis (DEA)

DEA is a non-parametric method in operations research and economics for assessing the relative efficiency of DMUs, such as firms, public-sector agencies, or production systems, which use multiple inputs to produce multiple outputs. DEA originated from Farrell's seminal work on efficiency measurement [7], and was later formalized and extended by Charnes, Cooper, and Rhodes through the development of the CCR model [1], and subsequently the BCC model by Banker, Charnes, and Cooper [8].

The DEA methodology constructs a best-practice frontier by enveloping observed data points through Linear Programming (LP), without requiring explicit assumptions on the functional form of the production process [9]. Efficiency is defined as the distance of a DMU from this frontier.

Farrell Efficiency Measure (1957)

The term technical efficiency was first defined formally by M. J. Farrell in 1957 [7], who introduced an analog to the ratio of productivity and suggested a method for comparing the relative efficiency of firms with respect to a constructed production frontier. The basic thought is that a firm is technically efficient if it works on the frontier and technically inefficient if it lies below.

Farrell suggested measuring efficiency as the maximum possible proportional reduction in inputs that still allows the firm to produce the same level of outputs. In modern notation, for a firm using an input vector x_0 to produce an output vector y_0, the Farrell input efficiency score E is defined as:

$$E = \min \theta \quad \text{such that } (\theta x_0, y_0) \in T \tag{3.1}$$

Where:

- T is the feasible production set (i.e., the set of all input-output combinations observed or assumed).
- θ is the scalar representing the maximal feasible contraction of inputs.

If $E = 1$, the firm is efficient. If $E < 1$, the firm is inefficient and can reduce all inputs by a proportion $(1 - E)$ without reducing output.

This model assumes radial reductions and laid the groundwork for the DEA methodology, which later operationalized this idea through LP using observed data.

3.1.1 Returns to Scale Assumptions: CRS and VRS

In the early development of DEA, two fundamental assumptions about production technology were established: Constant Returns to Scale (CRS) and Variable Returns to Scale (VRS)[8, 1]. Under CRS, a proportional increase in all inputs leads to an equivalent proportional increase in all outputs, implying that all DMUs operate at an optimal and efficient scale. In contrast, the VRS assumption allows for economies or diseconomies of scale, acknowledging that production efficiency may vary with the size of operations or external conditions. These assumptions are essential because they determine how the production frontier is constructed and how efficiency is interpreted. The classical CCR model proposed by Charnes, Cooper, and Rhodes (1978) represents the DEA formulation under CRS, whereas the later BCC model introduced by Banker, Charnes, and Cooper (1984) incorporates VRS to capture differences in scale among DMUs.

Model Orientation: Input- and Output-Oriented Approaches

Another important modeling aspect in DEA is the **orientation** of the analysis, which defines the direction in which efficiency is measured [9]. An input-oriented model evaluates how much input usage can be proportionally reduced while maintaining the same level of outputs. This perspective is suitable when decision makers have greater control over resource utilization than over output generation. Conversely, an output-oriented model assesses how much outputs can be proportionally expanded without increasing input consumption. This formulation is more appropriate when the focus lies on improving productivity or service delivery given fixed resources. Although these two orientations take different perspectives, they generally yield consistent efficiency classifications under convex production technologies [9].

3.1.2 The CCR Model (CRS)

The CCR model assumes CRS, implying that a proportional increase in all inputs results in a proportional increase in outputs. The CRS assumption is appropriate when all DMUs are believed to operate at an optimal scale [1].

The input-oriented CCR model evaluates how much input quantities can be proportionally reduced without decreasing output levels. It is formulated as:

$$\begin{aligned} \min_{\theta,\lambda} \quad & \theta \\ \text{s.t.} \quad & Y\lambda \geq y_0 \\ & X\lambda \leq \theta x_0 \\ & \lambda \geq 0 \end{aligned} \tag{3.2}$$

The output-oriented CCR model evaluates how much outputs can be proportionally increased without increasing input usage:

$$\begin{aligned} \max_{\phi,\lambda} \quad & \phi \\ \text{s.t.} \quad & Y\lambda \geq \phi y_0 \\ & X\lambda \leq x_0 \\ & \lambda \geq 0 \end{aligned} \tag{3.3}$$

Where:

- $x_0 \in \mathbb{R}^m$ is the input vector of the evaluated DMU.
- $y_0 \in \mathbb{R}^s$ is the output vector.
- X, Y are the input/output matrices.

- λ is a non-negative intensity vector.
- θ is the input efficiency score, ϕ is the output efficiency score.

3.1.3 The BCC Model (VRS)

The BCC model removes the assumption of CRS and instead allows for VRS, making it suitable when DMUs operate at different scales of production [8]. The key addition is the convexity constraint:

$$\sum_{j=1}^{n} \lambda_j = 1 \tag{3.4}$$

This ensures that the reference set is a convex combination of existing DMUs, allowing the efficiency frontier to exhibit variable returns to scale. The input-oriented BCC model becomes:

$$\begin{aligned} \min_{\theta,\lambda} \quad & \theta \\ \text{s.t.} \quad & Y\lambda \geq y_0 \\ & X\lambda \leq \theta x_0 \\ & \sum_j \lambda_j = 1 \\ & \lambda \geq 0 \end{aligned} \tag{3.5}$$

And the output-oriented BCC model:

$$\begin{aligned} \max_{\phi,\lambda} \quad & \phi \\ \text{s.t.} \quad & Y\lambda \geq \phi y_0 \\ & X\lambda \leq x_0 \\ & \sum_j \lambda_j = 1 \\ & \lambda \geq 0 \end{aligned} \tag{3.6}$$

3.1.4 Dual (Multiplier) Form

Every linear programming formulation in DEA has a corresponding dual problem that provides a complementary interpretation of efficiency [1, 9]. While the primal (envelopment) form constructs the efficiency frontier by enveloping the observed data, the dual—also known as the multiplier form—focuses on determining the optimal weights assigned to each input and output. In this view, a DMU is

considered efficient if it can choose a set of non-negative weights that maximize its own weighted output-to-input ratio without exceeding one for any other unit.

The dual of the input-oriented CCR model is expressed as:

$$\begin{aligned} \max_{u,v} \quad & u^T y_0 \\ \text{s.t.} \quad & v^T x_0 = 1 \\ & u^T Y - v^T X \leq 0 \\ & u, v \geq 0 \end{aligned} \tag{3.7}$$

In this formulation:

- u and v are the non-negative weight vectors for outputs and inputs, respectively;
- the constraint $v^T x_0 = 1$ normalizes the input side to avoid trivial scaling;
- the inequality $u^T Y - v^T X \leq 0$ ensures that no DMU achieves an efficiency score greater than one using the same set of weights.

This dual representation provides valuable managerial insight into how each input and output contributes to the efficiency score of a DMU. It also highlights one of the strengths of DEA: each unit is evaluated using the most favorable set of weights within the feasible space, thereby avoiding subjective or pre-specified weighting schemes.

Intuitive Understanding of the Efficiency Frontier

In the simplest case with one input and one output, the DEA frontier can be understood conceptually as a piecewise linear and convex boundary that envelops all observed DMUs [9]. Units that lie directly on this boundary are considered efficient ($\theta = 1$), while those located below it are inefficient. The efficiency score represents the proportional distance of a unit from this frontier—that is, the ratio between its current performance and the performance of an efficient peer operating on the boundary.

DEA Extensions

Over the years, DEA has been extended in multiple directions to better reflect real-world production settings and to increase its discriminatory power among efficient units [9]. The most relevant extensions include:

- Slack-Based Measure (SBM) models: which assess efficiency using input and output slacks rather than purely radial proportional changes [22].

- Directional Distance Function (DDF) models: which allow simultaneous input reductions and output augmentations in specified directions [23].
- Super-efficiency Models: which enable ranking of efficient DMUs by excluding the evaluated unit from the reference set [24].
- Malmquist Productivity Index and Window Analysis: which evaluate productivity and efficiency changes over time [25].
- Robust and Stochastic DEA: which incorporate noise or uncertainty in the data [26].

These formulations broaden the applicability of DEA across domains such as banking, energy, education, and healthcare [9]. However, while these models improve measurement and discrimination, they still provide limited interpretability regarding how an inefficient DMU should adjust its inputs or outputs to become efficient. This limitation motivates the integration of Bilevel optimization, which enables the generation of interpretable and actionable counterfactual targets, as discussed in the next section.

3.1.5 The Data Envelopment Analysis (DEA) model used in this research

The empirical analysis in this thesis builds upon the classical Data Envelopment Analysis (DEA) framework, originally introduced by Charnes et al. [1] and based on the efficiency concept formulated by Farrell [7]. DEA is a nonparametric method for measuring the relative efficiency of a set of homogeneous Decision Making Units (DMUs), each using multiple inputs to produce multiple outputs.

Input-oriented DEA and Farrell efficiency

In the input-oriented formulation, efficiency is defined as the maximum proportional reduction in inputs that allows the DMU to continue producing the same output levels. For a DMU_0 with input vector $\mathbf{x}_0$ and output vector $\mathbf{y}_0$, the model is written as:

$$\begin{aligned} \min_{\theta,\boldsymbol{\lambda}} \quad & \theta \\ \text{s.t.} \quad & Y\boldsymbol{\lambda} \geq \mathbf{y}_0, \\ & X\boldsymbol{\lambda} \leq \theta \mathbf{x}_0, \\ & \boldsymbol{\lambda} \geq 0, \end{aligned} \tag{3.8}$$

where X and Y are matrices of observed inputs and outputs for all DMUs, and $\boldsymbol{\lambda}$ represents the intensity variables. The scalar $\theta \in (0,1]$ corresponds to the Farrell efficiency score, indicating the feasible proportion of input use.

A DMU is efficient if $\theta = 1$, while values below unity imply inefficiency. The corresponding Farrell projection identifies the efficient input combination:

$$\mathbf{x}^* = \theta^* \mathbf{x}_0,$$

which lies on the efficient frontier. This projection defines the classical DEA benchmark used as a baseline for counterfactual comparison in later chapters.

Returns to scale and model orientation

The analysis in this thesis applies the input-oriented model under both Constant Returns to Scale (CRS) and Variable Returns to Scale (VRS) assumptions, following the CCR and BCC formulations respectively. The choice of orientation reflects the ESG benchmarking context, where firms aim to reduce environmental and social input intensities while maintaining the same output performance.

Transition to counterfactual DEA

Although the Farrell model provides an objective benchmark, it assumes uniform proportional contraction across all inputs, which may not correspond to realistic or interpretable managerial adjustments. To address this, the counterfactual DEA model proposed in Chapter 4 reformulates the efficiency projection as an optimization problem that incorporates regularization norms (L_0, L_2, and L_0+L_2) to capture sparsity, smoothness, and realism in improvement paths.

3.2 Bilevel Optimization

Bilevel optimization is a mathematical programming framework for modeling hierarchical decision-making processes where two decision-makers (called the leader and the follower) interact. The leader makes the first decision, anticipating the reaction of the follower, who then optimizes their own objective in response to the leader's action. This hierarchical structure was first formalized in optimization by Bard [13] and later developed into a comprehensive theoretical framework by Dempe [14]. Bilevel formulations naturally occur in many practical settings such as pricing and regulation, game theory, transportation planning, and network design [13, 14, 15]. More recently, they have also been applied in explainable analytics and performance benchmarking, where upper-level interpretability constraints guide lower-level optimization [15, 3].

3.2.1 General Formulation

A general Bilevel program consists of two nested optimization problems—an upper-level problem (leader) and a lower-level problem (follower)—that can be expressed as follows [13, 14]:

Upper-level problem (leader):

$$\min_{x \in X,\, y \in Y(x)} \quad F(x, y) \tag{3.9}$$

Subject to:

Lower-level problem (follower):

$$y \in \arg\min_{y \in Y(x)} \{f(x, y) : g(x, y) \leq 0\} \tag{3.10}$$

Where:

- $F(x, y)$ is the objective function of the upper-level (leader) problem;
- $f(x, y)$ is the objective function of the lower-level (follower) problem;
- $g(x, y) \leq 0$ are the constraints of the lower-level problem;
- x and y denote the upper- and lower-level decision variables, respectively.

Bilevel problems are challenging because of their nested structure, which makes them generally non-convex and Nondeterministic Polynomial time (NP)-hard [14, 15]. Specialized algorithms, relaxations, and reformulations have therefore been developed to handle such models in practical applications.

3.2.2 KKT Reformulation

One common approach for solving Bilevel problems with convex and differentiable lower-level problems is to replace the follower's optimization problem with its optimality conditions, typically the Karush–Kuhn–Tucker (KKT) conditions [13, 14]. The lower-level problem:

$$\begin{aligned} \min_{y} \quad & f(x, y) \\ \text{s.t.} \quad & g_i(x, y) \leq 0 \quad \forall i \end{aligned} \tag{3.11}$$

can be replaced by the KKT system:

$$\begin{aligned} \nabla_y f(x,y) + \sum_i \mu_i \nabla_y g_i(x,y) &= 0 \\ g_i(x,y) &\leq 0 \quad \forall i \\ \mu_i &\geq 0 \quad \forall i \\ \mu_i g_i(x,y) &= 0 \quad \forall i \end{aligned} \tag{3.12}$$

These KKT conditions transform the original Bilevel problem into a single-level Mathematical Program with Complementarity Constraints (MPCC) (Mathematical Program with Complementarity Constraints), which can be solved using relaxation or penalty-based techniques [14, 15].

3.2.3 Cost Functions for Counterfactual Explanations

An important recent application of bilevel optimization arises in counterfactual analysis and benchmarking. One of the key innovations in the counterfactual DEA model by Bogetoft et al. [3] is the use of flexible cost functions to model the "effort" or "burden" of improving performance. The idea is to generate counterfactual targets (i.e., improved inputs or outputs) that are not only efficient but also interpretable and actionable [4]. This is achieved by minimizing a convex combination of three norms:

$$\min \quad \nu_0 \|x - \hat{x}\|_0 + \nu_1 \|x - \hat{x}\|_1 + \nu_2 \|x - \hat{x}\|_2^2 \tag{3.13}$$

- ℓ_0-norm ($\|x - \hat{x}\|_0$): Promotes sparsity—it counts the number of variables that change, leading to counterfactuals requiring few modifications [3].
- ℓ_1-norm ($\|x - \hat{x}\|_1$): Encourages small total change and is widely used in interpretable modeling for feature selection [4].
- ℓ_2-norm ($\|x - \hat{x}\|_2^2$): Penalizes large deviations and promotes smooth, realistic recommendations [3].

The weights ν_0, ν_1, and ν_2 can be tuned to reflect user preferences. For example, a regulator may prioritize sparse explanations (high ν_0), while a manager may prefer gradual improvements (high ν_2). This flexible cost structure enables customized and practical benchmarking solutions [3].

Principle of Least Action in DEA Benchmarking

The use of composite cost functions aligns with the economic and philosophical principle of least action, which suggests that optimal adjustments should occur with

minimal total effort. In the DEA context, this principle implies that an inefficient DMU should not be forced to make arbitrary or excessive changes to become efficient. Instead, counterfactual recommendations should identify the "nearest" efficient point—one requiring minimal effort, disruption, or cost [3]. This notion also supports incentive compatibility, meaning that the proposed targets are more likely to be accepted and implemented in practice. It further allows benchmarking to be integrated into policy or regulatory frameworks where minimal invasiveness is crucial.

This principle is formalized in the CEDEA model through the minimization of distance functions subject to DEA efficiency constraints [3]. It complements traditional DEA models by offering not only evaluation but also explanation and prescription.

3.2.4 Interpretability and Actionability in DEA

Traditional DEA models have been criticized for their "black-box" nature, particularly in managerial or regulatory decision contexts. While DEA provides relative efficiency scores and target benchmarks, it does not clarify how or why those targets are chosen [9]. Counterfactual DEA addresses this limitation by producing interpretable and justified recommendations. For instance, it can answer questions such as:

- What is the smallest change required to become efficient?
- Which input contributes most to inefficiency?
- How do the suggested targets compare with peers?

These questions are particularly relevant in sectors where transparency and accountability are essential (e.g., healthcare, education, and public administration). Counterfactual DEA integrates explainability into classical OR frameworks, aligning with recent trends in XAI and interpretable analytics [4, 3].

3.2.5 Applications in DEA and Counterfactual Benchmarking

Bilevel optimization has recently been applied within DEA to construct interpretable, actionable recommendations for inefficient DMUs [3]. In this framework:

- The upper-level problem minimizes a cost or distance function $D(x, \hat{x})$ representing the magnitude or sparsity of input/output changes.
- The lower-level problem enforces that the counterfactual DMU $\hat{x}$ lies on or above a required efficiency level (e.g., $E \geq E^*$).

This structure is exemplified by the CEDEA model proposed by Bogetoft et al. [3], which employs a bilevel formulation to identify least-cost counterfactual improvements that move a DMU to the efficient frontier.

3.2.6 Example Formulation for Counterfactual DEA

Let x_0 denote the input vector of an inefficient DMU, and $\hat{x}$ its counterfactual (target) input vector. The Bilevel counterfactual model can be expressed as:

Upper-level (Counterfactual optimization):

$$\min_{\hat{x}} \quad \|x_0 - \hat{x}\|_1 \quad \text{s.t.} \quad \hat{x} \in X \tag{3.14}$$

Lower-level (DEA efficiency):

$$\begin{aligned} &\text{find } \lambda \quad \text{s.t.} \\ &\quad Y\lambda \geq y_0 \\ &\quad X\lambda \leq \hat{x} \\ &\quad \lambda \geq 0 \\ &\quad \text{(Optional: } \sum_j \lambda_j = 1 \text{ for VRS)} \end{aligned} \tag{3.15}$$

The constraint ensuring that $\hat{x}$ achieves the desired DEA efficiency level is encoded through the feasibility of the lower-level problem. This formulation can be extended with alternative norms (e.g., ℓ_0, ℓ_2), fairness constraints, or sparsity preferences [3, 4]. Such models offer explainable and realistic guidance for performance improvement, supporting compliance with modern standards of fairness and interpretability in data-driven decision making. In summary, the theoretical concepts of DEA and Bilevel optimization discussed in this chapter provide the foundation for the proposed framework developed next. By integrating efficiency measurement with counterfactual reasoning, the forthcoming methodology translates these ideas into an operational model capable of generating interpretable and actionable performance targets. The next chapter formalizes this integration and presents the mathematical formulation of the counterfactual DEA model.

Chapter 4

Proposed Methodology

Building on the theoretical foundations established in Chapter 3, this chapter presents the proposed methodological framework for integrating DEA with Bilevel optimization to produce interpretable and prescriptive efficiency benchmarks. The method transforms raw ESG performance indicators into an optimization-ready dataset, applies a robust efficiency estimation model, and extends it through counterfactual reasoning to generate realistic improvement targets. The chapter is organized as follows: first, the data preprocessing pipeline is described in detail; second, the mathematical formulation of the counterfactual DEA model is presented; finally, the computational implementation and interpretative aspects of the approach are discussed.

4.1 Data and variables

The dataset used in this study is a proprietary panel of logistics companies from 2023, containing a mixture of environmental, social, and governance (ESG) indicators. The raw dataset consists of 1,494 rows and 7 features, including company names, ESG scores, and key operational indicators such as:

- Water Use to Revenues (USD million)
- Total CO_2 Equivalent Emissions to Revenues (USD million)
- Waste Recycled to Total Waste
- Women Managers (percentage)
- Board Gender Diversity (percentage)
- Renewable Energy Use Ratio

- Employee Health and Safety Training Hours

The ESG indicators used in this study were retrieved from the Refinitiv ESG database (formerly Refinitiv Eikon), accessed through the university's licensed connection to the London Stock Exchange Group (LSEG) Workspace (https://www.lseg.com/). Refinitiv/LSEG ESG data are commercially licensed and proprietary, requiring an institutional subscription for access. Consequently, the full reproducibility of the empirical results is limited to researchers who have access to LSEG's subscription-based ESG data services.

4.1.1 Data Preprocessing

Before applying DEA and counterfactual models, the raw dataset was cleaned and transformed to ensure data quality, consistency, and suitability for efficiency measurement.

Input and Output Specification

Following the conceptual framework introduced in Chapter 3, each ESG indicator was classified as an input or output based on managerial controllability and performance orientation. The selected variables are:

- Inputs: Water Use to Revenues, Total CO_2 Equivalent Emissions to Revenues, Waste Recycled to Total Waste, Women Managers, and Board Gender Diversity.
- Output: ESG Score.

Feature Selection Rationale

The selection of ESG indicators reflects the three primary sustainability pillars:

- Environmental: Water Use to Revenues and CO_2 Emissions to Revenues represent the firm's ecological efficiency, capturing resource intensity and carbon impact relative to economic output.
- Social: Women Managers reflects the degree of gender inclusiveness in managerial positions, serving as an indicator of social diversity within the organization.
- Governance: Board Gender Diversity captures gender balance at the board level, reflecting the firm's commitment to equitable representation and transparent decision-making structures.

This selection aligns with the current trend of evaluating corporate sustainability as a multidimensional production process, where firms transform environmental and social resources into perceived ESG performance.

Data Loading and Cleaning

The dataset was imported from Excel. To guarantee consistency:

- Column names were standardized by removing extra spaces and line breaks.
- Duplicate rows were removed to avoid bias from repeated company entries.
- Companies with missing values in the target variable (ESG Score) were excluded.

Linear Transformation of Undesirable Inputs

In the input-oriented DEA framework, inputs represent quantities that should ideally be minimized. However, certain variables in the dataset—such as social or governance indicators (e.g., Women Managers, Board Gender Diversity)—are desirable when large. To maintain the correct directional interpretation without violating linearity, these variables were transformed using a linear transformation rather than inversion.

The transformation is expressed as:

$$x_i' = 1 - x_i, \tag{4.1}$$

which preserves the relative differences among observations while ensuring that higher original values (desirable) correspond to lower transformed values in the DEA model. This approach maintains the linearity of the production possibility set and is therefore more appropriate for linear programming–based DEA models.

Unlike normalization or reciprocal inversion, the linear transformation retains the original data scale and avoids introducing nonlinearity into the optimization problem.

Justification of Modelling Choice

An alternative modelling strategy for desirable indicators—such as Women Managers and Board Gender Diversity—would be to classify them as outputs, on the basis that firms "produce" diversity and inclusiveness as part of their social performance. While this interpretation is valid, in this study these indicators were intentionally retained as inputs and transformed using the linear mapping $x_i' = 1 - x_i$. This choice preserves a fully input-oriented perspective, ensuring that

all ESG-related drivers of performance are expressed in a common decision space and can be directly compared.

A further motivation for this choice is that both social and governance indicators play a significant role in a firm's overall ESG score. By keeping them as (transformed) inputs within the DEA framework, we allow the counterfactual model to identify how changes in these variables contribute to ESG efficiency in practice. If they were moved to the output side, their effect on inefficiency would be partially absorbed into the aggregate ESG score, reducing the model's ability to show how improvements in diversity influence the firm's position relative to the frontier. Retaining them as (transformed) inputs therefore provides clearer, feature-specific insight into the behavioural adjustments that lead to higher ESG performance, while maintaining consistency with the input-oriented structure of the analysis.

Outlier Treatment

To mitigate the influence of extreme values, the Interquartile Range (IQR) rule was applied with bounds:

$$\mathrm{LB} = Q_1 - 1.5\,\mathrm{IQR}, \qquad \mathrm{UB} = Q_3 + 1.5\,\mathrm{IQR} \tag{4.2}$$

Values outside $[\,\mathrm{LB}, \mathrm{UB}\,]$ were clamped.

Logarithmic Transformation

Skewed environmental indicators were log-transformed using $\log(1 + x)$:

- Water Use To Revenues (USD in million)
- Total CO_2 Equivalent Emissions To Revenues (USD in million)

Missing Value Imputation

Remaining missing values in numerical columns were imputed as follows:

- Median for highly skewed variables.
- Mean for approximately symmetric variables.

Feature Selection and Scaling

The final features used for modeling are:

1. Water Use to Revenues (USD million)

2. Total CO_2 Equivalent Emissions to Revenues (USD million)
3. Waste Recycled to Total Waste
4. Women Managers (percentage)
5. Board Gender Diversity (percentage)

All variables were scaled to [0,1] using:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{4.3}$$

4.1.2 Final Dataset

The company name column was renamed to DMU, and the cleaned dataset was saved to an excel file for next steps. After preprocessing and filtering for missing data, the final dataset includes 312 Decision Making Units (DMUs), each representing a logistics firm observed in 2023.

Table 4.1: Variables used in the study and their modelling roles.

Variable (raw)	**Role**	**Notes / transformation**
Water Use To Revenues	Input	Skewed ⇒ log-transformed; then scaled.
Total CO_2 Emissions	Input	Skewed ⇒ log-transformed; then scaled.
Waste Recycled To Total	Input	Desirable-as-large ⇒ scaled, then linearly transformed.
Women Managers	Input	Desirable-as-large ⇒ scaled, then linearly transformed.
Board Gender Diversity	Input	Desirable-as-large ⇒ scaled, then linearly transformed.
ESG Score	Output	Scaled.

After all transformations, the cleaned dataset was exported to an Excel file for reproducibility. The final preprocessing step confirmed that:

- all records were unique and complete;
- all input and output variables were scaled, bounded, and nonnegative;
- no zero values remained without substitution by ε;
- transformed inputs were directionally consistent.

This finalized dataset was used as the empirical basis for model implementation.

4.2 Problem Statement and Model Rationale

The model is designed to address the following question: "What minimal changes to ESG-related inputs would allow an inefficient company to become efficient or reach a desired level of performance?"

To address this, a custom optimization routine was implemented in Python using Gurobi, capturing the principles of radial input contraction, norm-based deviation minimization, and DEA frontier approximation. The model balances sparsity, interpretability, and practicality in counterfactual target setting, and is tailored to ESG performance benchmarking within the logistics sector.

4.3 Modeling Framework

Let each DMU k be defined by an input vector $x_{0k} \in \mathbb{R}^m$ and output vector $y_{0k} \in \mathbb{R}^s$. The objective is to find a counterfactual input vector $\hat{x}_k$ such that the DMU would achieve at least a specified efficiency level E^*, relative to the empirical DEA frontier constructed from peer firms.

Objective Function: Multi-Norm Minimization

To promote both interpretability and realism in the counterfactual suggestions, the model minimizes a composite cost function consisting of ℓ_0, ℓ_1, and ℓ_2 norm components:

$$\min_{\hat{x}_k} \ \nu_0 \|\hat{x}_k - x_{0k}\|_0 \ + \ \nu_1 \|\hat{x}_k - x_{0k}\|_1 \ + \ \nu_2 \|\hat{x}_k - x_{0k}\|_2^2 \tag{4.4}$$

Each term plays a distinct role:

- ℓ_0 norm: encourages sparsity, i.e., changing as few input features as possible.
- ℓ_1 norm: controls the total absolute magnitude of changes, promoting interpretability.
- ℓ_2 norm: penalizes large deviations, ensuring smooth and moderate adjustments.

This formulation is inspired by the general norm-based approach proposed in [3], but is here fully adapted and implemented for ESG benchmarking in the logistics sector.

Efficiency Constraints (DEA BCC Model)

The feasibility of the counterfactual input vector $\hat{x}_k$ is enforced via the classical input-oriented BCC model (see Section 3.1.3):

Input constraint

$$\hat{x}_{k,i} \geq \sum_{f=1}^{n} \beta_f x_{f,i} \quad \forall i \tag{4.5}$$

Output constraint

$$\frac{1}{E^*} \cdot y_{0k,o} \leq \sum_{f=1}^{n} \beta_f y_{f,o} \quad \forall o \tag{4.6}$$

Convexity condition (VRS)

$$\sum_{f=1}^{n} \beta_f = 1 \tag{4.7}$$

The output vector y_{0k} remains fixed, while $\hat{x}_k$ is optimized to achieve the target efficiency.

Duality and KKT Conditions

To linearize the bilevel optimization problem, the lower-level DEA model is replaced by its Karush-Kuhn-Tucker (KKT) conditions (see Section 3.2.2):

Dual normalization

$$\sum_{o} \gamma_o y_{0k,o} + \chi = 1 \tag{4.8}$$

KKT inequality for all reference DMUs

$$\sum_{i} \gamma_i x_{f,i} - \sum_{o} \gamma_o y_{f,o} - \chi \geq 0 \quad \forall f \tag{4.9}$$

The model further includes binary slack variables (u, v, w) and large constant bounds (M) to approximate complementarity slackness.

Use of the Big M Method for Constraint Activation

The proposed model makes extensive use of the Big M method to linearize and activate logical constraints. In particular, binary variables are introduced to control the enforcement of slacks and complementary conditions.

For example, a typical constraint involving an input slack variable u_i is formulated as:

$$\gamma_i \leq M_1 \cdot u_i, \tag{4.10}$$

which ensures that γ_i is zero whenever $u_i = 0$, and allows γ_i to take a positive value only when $u_i = 1$. Similar constructs are used for output slacks (v_o) and frontier constraints (w_f), each with different upper bounds M_1, M_2, and M_4 respectively.

These M constants are chosen based on the scale of the problem to avoid overly loose bounds that could degrade numerical stability, while still ensuring feasibility. The use of Big M enables the model to encode piecewise and conditional logic within a mixed-integer linear programming (MILP) formulation, which is essential for implementing the bilevel-inspired structure and counterfactual generation mechanism.

Linearization of Norm Terms

- ℓ_0 norm is approximated by binary indicators ξ_i^0:

$$-M \cdot \xi_i^0 \leq \hat{x}_{k,i} - x_{0k,i} \leq M \cdot \xi_i^0$$

- ℓ_1 norm is modeled via auxiliary variables ξ_i^1:

$$\xi_i^1 \geq |\hat{x}_{k,i} - x_{0k,i}|$$

- ℓ_2 norm is added directly as:

$$\sum_i (\hat{x}_{k,i} - x_{0k,i})^2$$

Single-Level Reformulation via KKT Conditions

Although the proposed model is inspired by a bilevel optimization structure—with an upper-level objective seeking interpretability and a lower-level DEA feasibility problem—the formulation is ultimately solved as a single-level program. This transformation is made possible by leveraging the optimality conditions of the inner DEA problem.

In particular, following the approach proposed by Bogetoft et al. [3], the bilevel problem is replaced by a single-level reformulation where the lower-level problem is substituted by its Karush-Kuhn-Tucker (KKT) conditions. This is valid because the lower-level DEA problem is a convex linear program with continuous variables. As such, the KKT conditions are both necessary and sufficient for optimality.

As illustrated in Equation (5) of [3], the inner DEA problem is first expressed in its dual form, and then the complementary slackness, primal feasibility, and dual feasibility conditions are explicitly encoded into the model. These constraints are directly integrated into the upper-level formulation, thereby yielding a mixed-integer linear program (MILP) that captures both feasibility and optimality of the inner problem within a single optimization layer.

This KKT-based reformulation avoids the need for bilevel solvers and simplifies the computation while preserving the economic interpretation of the model. The binary variables introduced in the model (e.g., u_i, v_o, w_f) are used to linearize the complementarity constraints using the Big M method (see Section 4.3), resulting in a tractable formulation suitable for counterfactual analysis and sparse target setting.

4.3.1 Implementation Overview

The complete model is implemented in Python using the Gurobi solver. For each DMU k, the model solves the MILP problem defined above and returns:

- Counterfactual inputs $\hat{x}_k$
- Efficiency after adjustment (E_k)
- Peer weights β_f (defining the reference convex combination)
- Feature-wise input changes Δx
- Total deviation cost (objective value)

The analyst may control the trade-offs via the norm weights (ν_0, ν_1, ν_2) and the target efficiency level E^*. This framework provides a flexible tool for generating interpretable improvement targets in ESG performance assessment.

Chapter 5

Results Discussion, Conclusion and Future Work

This chapter illustrates the empirical implementation of the counterfactual-based DEA framework introduced in Chapter 4. The aim is to assess how the proposed model performs when applied to a real-world, high-dimensional ESG dataset from the logistics sector and to demonstrate its capability to produce interpretable and actionable efficiency improvements.

The analysis uses the panel of logistics firms described in Section 4.1. After the preprocessing procedures outlined in Chapter 4—including scaling, transformation of desirable inputs, and outlier adjustment—the final dataset contains 312 decision-making units (DMUs). Each DMU is characterized by five input indicators representing key environmental, governance, and social dimensions and by a single output variable, the ESG Score. All variables were scaled to the [0,1] range to ensure comparability and numerical stability during optimization.

This chapter is structured as follows. Section 5.1 introduces the dataset and descriptive statistics of the normalized variables. Section 5.2 provides a detailed case study of one representative firm, Santos Brasil Participações SA, illustrating how the counterfactual framework prescribes interpretable improvement paths compared with the classical Farrell benchmark. Section 5.3 extends the analysis to the entire sample of logistics firms, comparing results across different efficiency targets ($E^* = 1.0$ and $E^* = 0.8$) and all combinations of cost-function configurations. Finally, Section 5.4 summarizes the key findings, discusses their managerial and methodological implications, and outlines directions for future research.

All optimization models were implemented in Python 3.11.5 using the gurobipy interface to the Gurobi Optimizer (version 11.0.3, build v11.0.3rc0, win64). Numerical experiments were performed on a laptop equipped with an AMD Ryzen 7 5800H processor (8 cores, 16 threads, base frequency 3.2 GHz) and 16 GB RAM, running

64-bit Windows 11. The main Python libraries employed in the implementation were NumPy 1.24.3 and pandas 2.0.3.

5.1 The Data

The dataset used in this study comprises 312 logistics firms observed in 2023, each characterized by five input indicators and one output variable corresponding to the overall ESG score. The selected input variables capture key environmental, governance and social dimensions of corporate sustainability, namely Water Use to Revenues, Total CO_2 Equivalent Emissions to Revenues, Waste Recycled to Total Waste, Women Managers, and Board Gender Diversity. All preprocessing steps—including normalization, transformation of desirable inputs, and outlier adjustment—were performed as described in Chapter 4. Consequently, all variables are scaled within the [0,1] interval, ensuring comparability and numerical stability in the optimization models. Table 5.1 reports the descriptive statistics of the normalized dataset used for the efficiency and counterfactual analyses.

Table 5.1: Descriptive statistics of the normalized ESG dataset (scaled to [0,1]).

Variable	**Mean**	**Min**	**Max**	**Std. dev.**
Water Use to Revenues	0.605,8	0.000,1	1.000,0	0.212,6
CO_2 Emissions to Revenues	0.642,0	0.000,1	1.000,0	0.217,6
Waste Recycled to Total Waste	0.471,0	0.000,0	0.999,9	0.202,8
Women Managers	0.618,2	0.000,1	0.999,9	0.154,4
Board Gender Diversity	0.695,6	0.000,1	0.999,9	0.202,2
ESG Score (output)	0.535,8	0.000,1	1.000,0	0.232,3

The normalized input and output values show moderate dispersion, with standard deviations ranging between 0.15 and 0.23, indicating a balanced spread across firms after scaling. The output variable (ESG Score) presents an average of 0.54, reflecting a mid-level overall performance for the sample. These values provide a consistent and well-conditioned input for the efficiency and counterfactual optimization models developed in this work.

After applying the input-oriented BCC DEA model, the efficiency distribution of firms was computed. Out of the 312 logistics companies analyzed, the majority exhibit substantial room for improvement relative to the estimated production frontier. Table 5.2 summarizes the frequency of firms across efficiency intervals before the application of counterfactual adjustments.

Table 5.2: Original DEA efficiency distribution.

Efficiency range (%)	**Number of firms**
<50%	206
50–60%	30
60–70%	24
70–80%	11
80–90%	11
90–95%	3
95–100%	27

As shown in Table 5.2, 206 firms (approximately 66% of the sample) record efficiency levels below 0.5, while only 27 firms are fully or nearly efficient ($E \geq 0.95$). This distribution indicates a highly skewed efficiency pattern within the sector, suggesting that a significant portion of firms operate far from the best-practice frontier. Such heterogeneity highlights the importance of counterfactual benchmarking to provide interpretable improvement targets and identify realistic pathways toward enhanced ESG performance.

5.2 Counterfactual analysis for one firm: Santos Brasil Participacoes SA

This section presents a detailed counterfactual analysis for Santos Brasil Participacoes SA, a representative firm in the logistics sector sample. In the classical input-oriented BCC model (see Section 3.1.5), the firm obtained an efficiency score of $\theta_k = 0.6307$, meaning that, under a purely radial contraction, its inputs could theoretically be reduced by approximately 36.9% while maintaining the same output level. While this Farrell measure provides a global efficiency benchmark, it does not indicate which specific variables drive inefficiency or how realistic the implied reductions are. The counterfactual DEA framework developed in this thesis addresses these limitations by producing interpretable, feature-level improvement paths that are both feasible and economically meaningful.

Baseline DEA and Farrell efficiency

The baseline DEA model follows the input-oriented BCC (VRS) formulation introduced by Banker et al. [8], which measures the Farrell efficiency of each

decision-making unit (DMU). For a given firm k, the efficiency score θ_k is obtained by solving

$$\min_{\theta,\lambda} \ \theta \quad \text{s.t.} \quad Y\lambda \geq y_k, \quad X\lambda \leq \theta x_k, \quad \lambda \geq 0, \quad \mathbf{1}^\top \lambda = 1. \tag{5.1}$$

The resulting $\theta_k \in (0,1]$ represents the proportion by which all inputs can be radially reduced while maintaining the same output levels, and its reciprocal $F_k = 1/\theta_k$ defines the corresponding Farrell frontier point. For each firm, the classical Farrell projection is then

$$x_k^{\text{Farrell}} = \theta_k \, x_k^{\text{original}},$$

which specifies the proportional input contraction required to reach the efficient frontier. For Santos Brasil Participacoes SA, this corresponds to

$$x^{\text{Farrell}} = 0.6307 \times x^{\text{original}},$$

serving as a conventional benchmark against which the counterfactual targets will be compared.

Counterfactual configurations and cost functions

The counterfactual DEA model extends this baseline by introducing a cost-of-adjustment function that minimizes the effort required for a firm to reach a desired efficiency target E^*. The general optimization problem is

$$\min_{x'_k} \ \nu_0 \, \|x'_k - x_k\|_0 + \nu_1 \, \|x'_k - x_k\|_1 + \nu_2 \, \|x'_k - x_k\|_2^2 \tag{5.2}$$

subject to the standard DEA feasibility constraints. Here, x'_k denotes the adjusted (counterfactual) input vector and x_k the original one. Each norm encodes a complementary notion of adjustment effort:

- $\|x'_k - x_k\|_0$ — number of variables that change (*sparsity*);
- $\|x'_k - x_k\|_1$ — total absolute deviation (*budgeted cost*);
- $\|x'_k - x_k\|_2^2$ — squared magnitude of deviation (*smooth proportional change*).

By choosing different combinations of the weights (ν_0, ν_1, ν_2), the model reproduces distinct managerial preferences: focusing effort on a few critical indicators, distributing small changes across all, or balancing both principles.

Table 5.3: Weight configurations (ν_0, ν_1, ν_2) used for counterfactual cost functions.

Cost Function	$\boldsymbol{\nu_0}$	$\boldsymbol{\nu_1}$	$\boldsymbol{\nu_2}$
ℓ_0	10^5	0	0
ℓ_2	0	0	10^5
$\ell_0 + \ell_2$	1	0	1

Table 5.3 summarizes the three cost-function specifications adopted in this study. The ℓ_0 configuration promotes parsimonious adjustments by penalizing the number of modified indicators; the ℓ_2 configuration favors smooth, distributed changes; and the combined ℓ_0+ℓ_2 setting balances interpretability and realism. These formulations are used consistently across all efficiency targets ($E^* = 0.8$ and $E^* = 1.0$), enabling a direct comparison of adjustment patterns under different regularization regimes.

Managerial interpretation

For Santos Brasil Participacoes SA, the counterfactual framework yields actionable, feature-level targets that complement the classical DEA projection. Each configuration reflects a different managerial philosophy: whether to concentrate resources on a few key inputs (ℓ_0), to pursue incremental system-wide improvements (ℓ_2), or to find a compromise between the two ($\ell_0 + \ell_2$). Together, these results illustrate how counterfactual DEA transforms traditional efficiency analysis from a purely diagnostic exercise into a prescriptive and interpretable decision-support tool.

Table 5.4: Signed input changes required under different targets and configurations.

Target (E^*)	**Configuration**	**Water Use to Revenues**	**CO_2 Emissions to Revenues**	**Waste Recycled to Total Waste**	**Women Managers**	**Board Gender Diversity**	**Number of Changes**
0.8	ℓ_2	−0.07845	−0.01315	+0.00000	+0.05386	+0.03538	4
	ℓ_0	−0.00000	−0.00000	+0.28445	+0.00000	+0.00000	1
	$\ell_0 + \ell_2$	−0.13221	−0.00000	+0.00000	+0.00000	+0.00000	1
1.0	ℓ_2	−0.00000	−0.23487	+0.14357	+0.00000	+0.00000	2
	ℓ_0	−0.00000	−0.32264	+0.00000	+0.00000	+0.00000	1
	$\ell_0 + \ell_2$	−0.00000	−0.32264	+0.00000	+0.00000	+0.00000	1

Numerical targets

Tables 5.5 and 5.6 report the counterfactual target values for Santos Brasil Participacoes SA at efficiency levels $E^* = 0.8$ and $E^* = 1.0$. Each configuration is

Table 5.5: Counterfactual targets for Santos Brasil Participacoes SA at $E^* = 0.8$.

Configuration	Water Use to Revenues	Total CO_2 to Revenues	Waste Recycled to Total Waste	Women Managers	Board Gender Diversity
ℓ_2	0.4561	0.5104	0.7154	0.4804	0.2946
ℓ_0	0.5346	0.5236	0.9999	0.4266	0.2593
$\ell_0 + \ell_2$	0.4023	0.5236	0.7154	0.4266	0.2593
Farrell Benchmark	0.3371	0.3302	0.9796	0.5840	0.3550
Original Values	0.5346	0.5236	0.7154	0.4266	0.2593

Table 5.6: Counterfactual targets for Santos Brasil Participacoes SA at $E^* = 1.0$.

Configuration	Water Use to Revenues	Total CO_2 to Revenues	Waste Recycled to Total Waste	Women Managers	Board Gender Diversity
ℓ_2	0.5346	0.2887	0.8590	0.4266	0.2593
ℓ_0	0.5346	0.2010	0.7154	0.4266	0.2593
$\ell_0 + \ell_2$	0.5346	0.2010	0.7154	0.4266	0.2593
Farrell Benchmark	0.3371	0.3302	0.9796	0.5840	0.3550
Original Values	0.5346	0.5236	0.7154	0.4266	0.2593

compared with the firm's Original Values and the Farrell Benchmark. The latter was computed as a radial contraction aligned with the average DEA efficiency across firms. For the social and governance indicators—Waste Recycled, Women Managers, and Board Gender Diversity—higher values correspond to better performance.

To facilitate interpretation, Figures 5.1 and 5.2 display grouped bar charts for $E^* = 0.8$ and $E^* = 1.0$. Each feature cluster includes the Original Values, the Farrell Benchmark, and the three counterfactual configurations (ℓ_2, ℓ_0, ℓ_0+ℓ_2). This format makes the trade-offs visible across environmental Water Use, CO_2 Emissions) and social/governance (Waste Recycled, Women Managers, Board Diversity) dimensions.

Analysis and discussion

Table 5.4 quantifies how each configuration modifies the firm's ESG profile. Negative values indicate required reductions (in environmental burdens), while positive values denote desirable increases (in social or governance performance). To enhance interpretability, percentage changes have been computed relative to the firm's original values reported in Tables 5.5 and 5.6.

At $E^* = 0.8$, the ℓ_2 model applies four moderate yet coordinated adjustments: Water Use decreases from 0.5346 to 0.4561, corresponding to a 14.7% reduction; CO_2 Emissions fall slightly from 0.5236 to 0.5104 (2.5% decrease); while Women Managers increase from 0.4266 to 0.4804 (12.6% improvement) and Board Gender

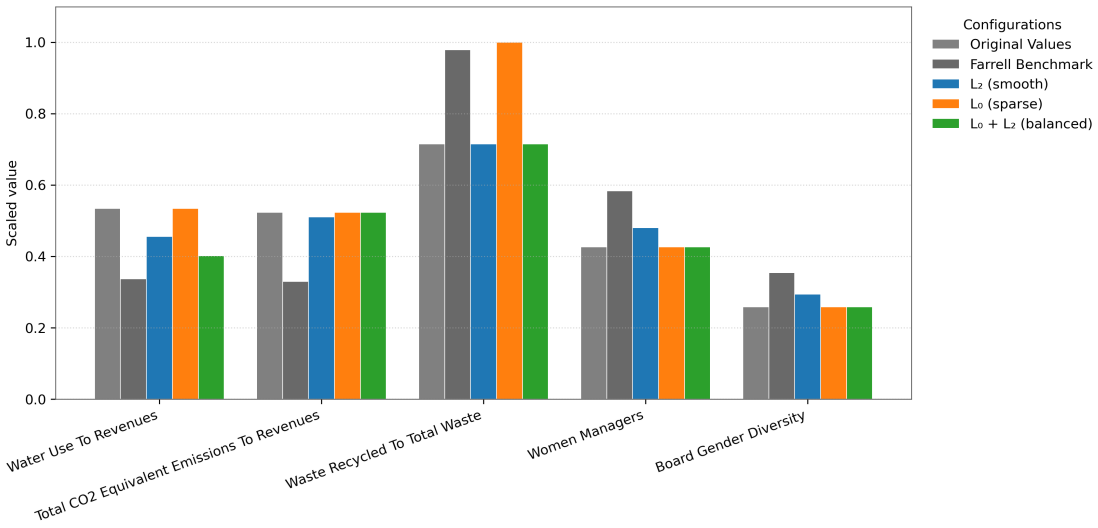


Figure 5.1: Grouped bar chart at $E^* = 0.8$: Original, Farrell, and counterfactual targets.

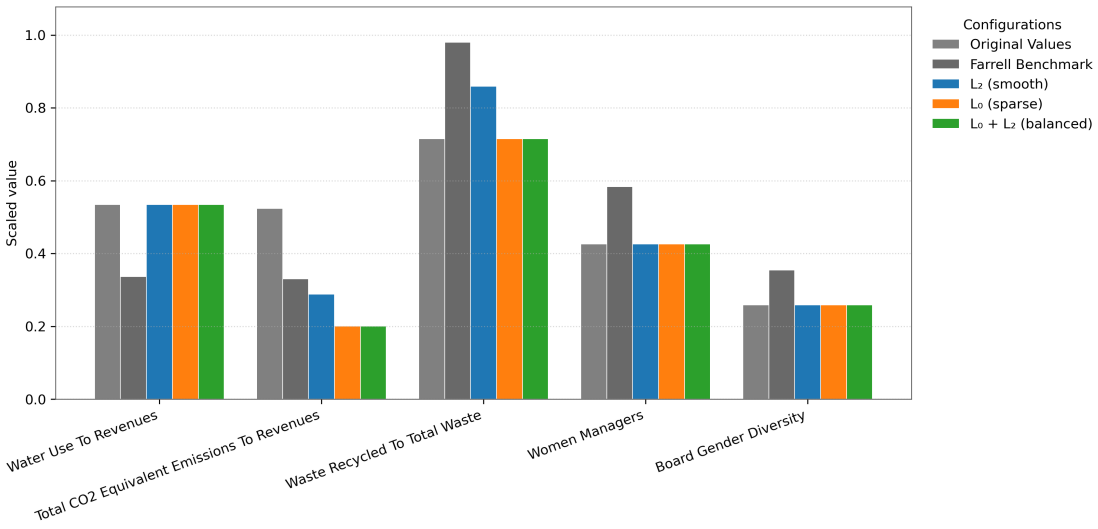


Figure 5.2: Grouped bar chart at $E^* = 1.0$: Original, Farrell, and counterfactual targets.

Diversity rises from 0.2593 to 0.2946 (13.6% increase). These small, distributed changes describe a balanced transition—modest environmental reductions combined with tangible gains in social representation.

In contrast, the sparse ℓ_0 configuration focuses entirely on one dimension: Waste Recycled rises sharply from 0.7154 to 0.9999, an extraordinary 39.8% increase, while all other indicators remain fixed. This pattern reflects a selective, one-dimensional improvement strategy centered on waste management, demonstrating that achieving moderate efficiency ($E^* = 0.8$) can be realized through targeted, high-impact actions. The mixed $\ell_0 + \ell_2$ configuration instead acts on Water Use alone, reducing it from 0.5346 to 0.4023 (24.7% decrease), while all other variables remain constant. This highlights that a single, substantial environmental improvement can be sufficient for reaching an intermediate efficiency target.

At $E^* = 1.0$, where the efficiency requirement is stricter, the magnitude and selectivity of adjustments increase. Under the smooth ℓ_2 configuration, CO_2 Emissions drop from 0.5236 to 0.2887, equivalent to a substantial 44.9% reduction, and Waste Recycled improves from 0.7154 to 0.8590 (20.1% increase). Water Use, Women Managers, and Board Diversity remain unchanged, indicating that the path to full efficiency primarily depends on environmental performance. Both the ℓ_0 and $\ell_0 + \ell_2$ configurations focus exclusively on CO_2 Emissions, reducing them from 0.5236 to 0.2010—an impressive 61.6% decrease. This result confirms that, at higher efficiency levels, environmental efficiency—particularly emission control—becomes the dominant driver of improvement.

Feature-level insight

Across both efficiency levels, Water Use and CO_2 Emissions act as the main "pull" variables (requiring reductions), while Waste Recycled, Women Managers, and Board Gender Diversity serve as "push" variables (increasing when feasible). At $E^* = 0.8$, the ℓ_2 model promotes a balanced multi-dimensional change: about 15% reduction in water intensity, minor 2–3% CO_2 cut, and 12–14% increases in social and governance indicators. At $E^* = 1.0$, however, the adjustment strategy becomes more concentrated—driven almost entirely by large-scale emission cuts (45–62%) and moderate recycling improvements (20%). This transition from distributed adjustments to concentrated environmental improvements illustrates the nonlinear trade-off between interpretability and ambition: as the efficiency goal becomes more demanding, firms must focus their resources on the most elastic, high-impact variables.

Sparsity versus magnitude

The final column in Table 5.4 confirms this trade-off between sparsity and magnitude. At $E^* = 0.8$, the ℓ_2 model modifies four inputs with relatively small individual magnitudes (between 2–15%), whereas the sparse ℓ_0 and $\ell_0 + \ell_2$ configurations each alter only one input but with much greater amplitude—25–40% changes in a single feature. At $E^* = 1.0$, the pattern becomes even sharper: ℓ_2 coordinates two variables (with 20–45% adjustments), while the sparse variants perform one very large shift (over 60% reduction in CO_2 intensity). Thus, as sparsity increases, the model compensates by requiring more drastic changes in individual inputs, reflecting the inherent trade-off between interpretability (fewer changes) and realism (smaller steps).

Comparison with the Farrell benchmark

The classical Farrell benchmark, derived through uniform proportional contraction, reduces all inputs proportionally (to around 63% of their original levels) and increases desirable inputs uniformly. This approach lacks differentiation and interpretability, as it assumes identical reduction ratios across all dimensions. In contrast, counterfactual DEA identifies precisely which dimensions matter most and by how much they should change. For instance, it shows that realistic efficiency improvements can be achieved through targeted actions—such as reducing CO_2 emissions by up to 60%, lowering water use by 25%, or raising waste recycling by 40%—rather than applying blanket proportional cuts. By quantifying improvement paths in percentage terms and aligning them with managerial dimensions, the counterfactual framework transforms DEA from a diagnostic evaluation tool into a prescriptive decision-support system, capable of offering actionable and data-grounded guidance for sustainability transitions.

5.3 Aggregate Counterfactual Results across All Firms

After analyzing one representative firm, the counterfactual DEA framework is now applied to the entire sample of 312 logistics companies. This section presents the experimental design adopted for the aggregate analysis, which includes all combinations of cost-function parameters and two efficiency targets, $E^* = 1.0$ and $E^* = 0.8$. Based on the original efficiency scores reported in Table 5.2, 27 firms in the sample were already classified as efficient at the full efficiency level $E^* = 1.0$. When the target was relaxed to $E^* = 0.8$, this number increased to 41 firms, reflecting the expected expansion of the feasible set under a less stringent

performance requirement. .Each configuration was solved independently for every firm in the dataset.

Table 5.7: Combinations of cost-function parameters (ν_0, ν_1, ν_2) used for aggregate counterfactual analysis.

Cost Function	$\boldsymbol{\nu_0}$	$\boldsymbol{\nu_1}$	$\boldsymbol{\nu_2}$
ℓ_0	10^5	0	0
$\ell_0 + \ell_2$	1	0	1
ℓ_2	0	0	10^5
ℓ_1	0	10^5	0
$\ell_0 + \ell_1$	1	1	0
$\ell_1 + \ell_2$	0	1	1
$\ell_0 + \ell_1 + \ell_2$	1	1	1

The configurations in Table 5.7 represent all tested combinations of the three norm-based cost components. Each setup was evaluated for the two efficiency targets to examine how the choice of regularization and ambition level jointly influence the feasibility and magnitude of counterfactual adjustments across the full sample of firms.

The subsequent analysis compares the firms reaching the desired efficiency under each configuration and quantifies the distribution of required input changes.

Aggregate counterfactual results for $E^* = 1.0$

We now summarize the statistics of the counterfactual adjustments obtained for the full sample when the desired efficiency is set to $E^* = 1.0$. Following the target–benchmarking structure, Table 5.8 reports the percentage of firms for which each input changes; Table 5.9 shows the average absolute change in each input; and Table 5.10 provides the average number of inputs that change per firm. Heatmaps in Figures 5.3a–5.3c illustrate these patterns visually for the three main principal cost functions.

Table 5.8: Share of firms (%) for which each input changes when desired efficiency is $E^* = 1.0$.

Input	ℓ_0	ℓ_0+ℓ_2	ℓ_2	ℓ_1	ℓ_0+ℓ_1	ℓ_1+ℓ_2	ℓ_0+ℓ_1+ℓ_2
Women Managers	44.9	11.9	41.4	16.0	11.9	16.4	12.2
CO_2 Emissions to Revenues	24.0	43.3	48.1	43.3	43.6	42.6	43.6
Waste Recycled to Total Waste	15.7	11.5	57.4	19.6	11.5	23.1	11.9
Water Use to Revenues	9.3	14.7	25.6	15.1	14.7	16.7	14.4
Board Gender Diversity	9.3	21.8	33.3	18.9	21.5	19.6	21.2

Table 5.9: Average absolute change by input when desired efficiency is $E^* = 1.0$.

Input	ℓ_0	ℓ_0+ℓ_2	ℓ_2	ℓ_1	ℓ_0+ℓ_1	ℓ_1+ℓ_2	ℓ_0+ℓ_1+ℓ_2
Women Managers	0.237	0.046	0.083	0.063	0.046	0.065	0.047
CO_2 Emissions to Revenues	0.140	0.209	0.174	0.205	0.211	0.193	0.211
Waste Recycled to Total Waste	0.080	0.024	0.066	0.028	0.024	0.034	0.024
Water Use to Revenues	0.054	0.052	0.077	0.049	0.050	0.060	0.050
Board Gender Diversity	0.056	0.100	0.061	0.072	0.100	0.068	0.097

Table 5.10: Average number of inputs that change per firm at $E^* = 1.0$.

Cost Function	**Mean number of inputs changed**
ℓ_0	2.57
ℓ_0+ℓ_2	2.53
ℓ_2	3.18
ℓ_1	1.74
ℓ_0+ℓ_1	2.41
ℓ_1+ℓ_2	2.02
ℓ_0+ℓ_1+ℓ_2	2.44

For $E^* = 1.0$, the most frequently adjusted variable across all cost functions is CO_2 Emissions to Revenues, changing in about 43–48% of firms (except for ℓ_0). Under the ℓ_2 configuration, Waste Recycled to Total Waste also shows a high adjustment frequency (57%), consistent with its role as a flexible compensating input. In contrast, Water Use and Board Gender Diversity remain among the least frequently modified indicators, with changes observed in roughly 15–25% of firms.

Average magnitudes confirm these patterns. CO_2 reductions are the most substantial (average change of approximately 0.20), whereas adjustments in recycling and water use are modest (below 0.08). Under ℓ_0, where few features are altered,

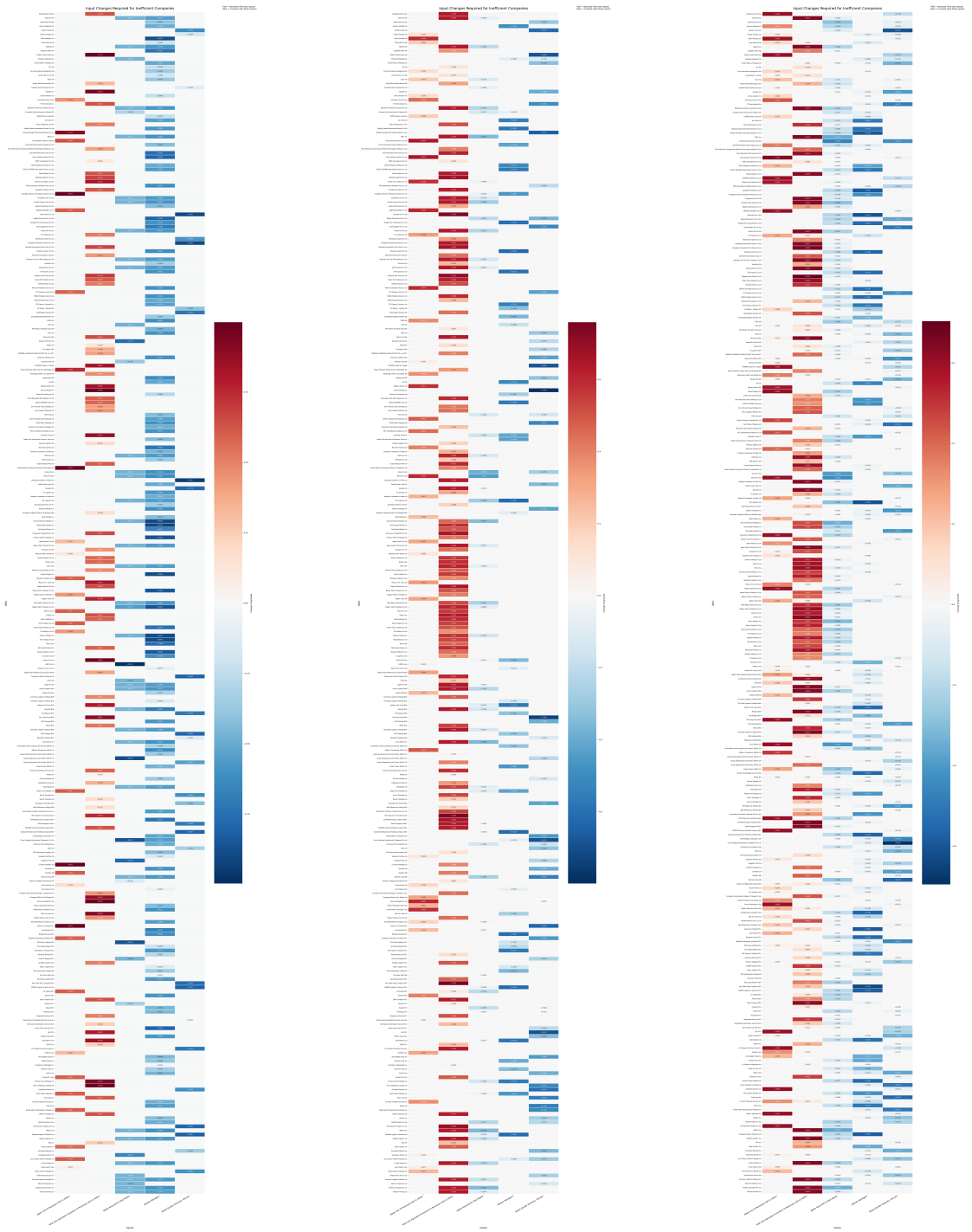
Women Managers displays the largest mean shift (0.24), emphasizing that sparse configurations require larger single-variable moves.

Finally, Table 5.10 shows that firms typically modify between two and three inputs to achieve full efficiency, with the smallest adjustment sets observed for ℓ_1(1.7 inputs) and the largest for ℓ_2 (3.2 inputs). These results align with the visual heatmaps: ℓ_2 (Figure 5.3c) produces dense modification patterns, while ℓ_0 (Figure 5.3a) and ℓ_1 Figures 5.3 and 5.4 provide a visual overview of the magnitude and direction of input adjustments for all firms across the three cost-function configurations (ℓ_0, ℓ_1, and ℓ_2) and the two target efficiency levels ($E^* = 1.0$ and $E^* = 0.8$). Each row corresponds to a firm, and each column represents one of the input indicators. The color intensity encodes the size of the adjustment, with blue tones indicating reductions (improvements for environmental indicators such as Water Use and CO_2 Emissions) and red tones indicating increases (improvements for social and governance indicators such as Women Managers and Board Gender Diversity). Darker shades correspond to larger absolute changes, allowing for a clear visual comparison of adjustment magnitude and distribution.

At the full-efficiency target ($E^* = 1.0$), the heatmaps reveal a denser and more intense pattern of modifications, particularly for the ℓ_2 configuration, which applies small but widespread adjustments across most firms and variables. In contrast, the ℓ_0 configuration remains highly sparse, showing isolated blue or red patches that correspond to substantial changes in only one or two dimensions per firm—consistent with its emphasis on minimal, high-impact modifications. The ℓ_1 variant lies between these two extremes, balancing coverage and intensity by allowing moderate changes across a broader range of inputs.

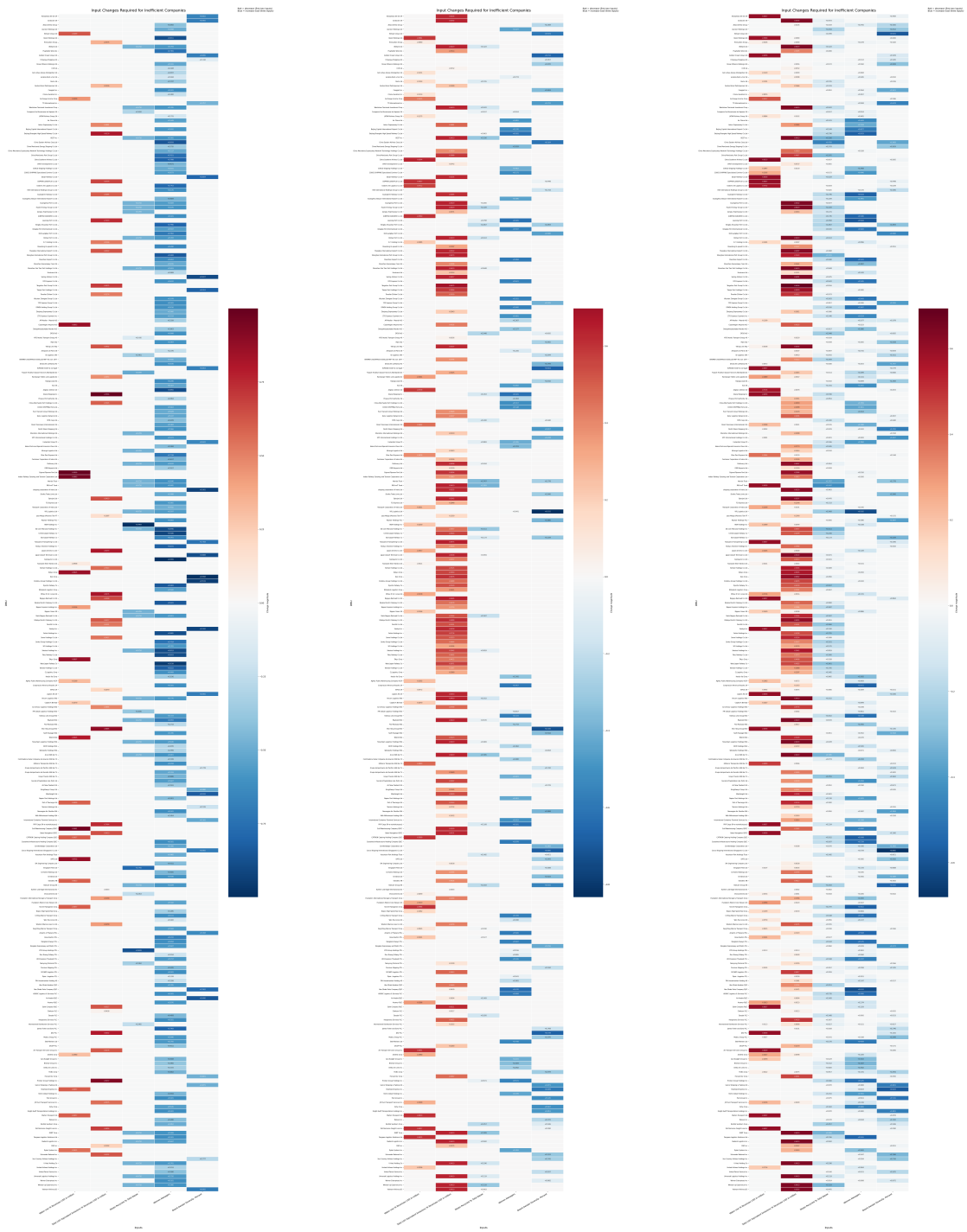
When the target is relaxed to $E^* = 0.8$, the overall color intensity visibly diminishes, reflecting the smaller magnitude of adjustments required to reach the efficiency goal. The ℓ_2 configuration still exhibits a distributed pattern of mild changes, whereas ℓ_0 and ℓ_1 become even sparser, indicating that fewer variables need to be modified at this lower efficiency level. This contrast between the two efficiency targets confirms the quantitative findings: higher performance ambitions ($E^* = 1.0$) demand more extensive and intense input adjustments, while moderate targets ($E^* = 0.8$) can be achieved through fewer, less disruptive changes.

Overall, the two sets of heatmaps provide an intuitive visual summary of how adjustment strategies evolve across configurations and efficiency targets. They illustrate the core trade-off between sparsity and magnitude: ℓ_0 concentrates effort on a few key variables, ℓ_1 distributes moderate changes across several, and ℓ_2 achieves smooth, system-wide improvements. Together, these figures complement the numerical analysis by showing at a glance how counterfactual adjustments are distributed across the ESG input space and how their intensity scales with the desired efficiency level.



(a) ℓ_0 **(b)** ℓ_1 **(c)** ℓ_2

Figure 5.3: Inputs that change for individual firms under $E^* = 1.0$ across the three cost functions.



(a) ℓ_0 **(b)** ℓ_1 **(c)** ℓ_2

Figure 5.4: Inputs that change for individual firms under $E^* = 0.8$ across the three cost functions.

Aggregate counterfactual results for $E^* = 0.8$

We now report the aggregate results for the relaxed target efficiency $E^* = 0.8$. As expected, when the target is less stringent, more firms can reach the efficiency threshold with smaller or fewer changes in their inputs. Table 5.11 shows the percentage of firms in which each input changes under different cost functions, while Table 5.12 reports the average absolute size of those changes. The overall number of efficient firms rises from 27 at $E^* = 1.0$ to 41 (see Section 5.3), confirming that a lower target increases feasibility.

Table 5.11: Share of firms (%) for which each input changes when desired efficiency is $E^* = 0.8$.

Input	ℓ_0	ℓ_0+ℓ_2	ℓ_2	ℓ_1	ℓ_0+ℓ_1	ℓ_1+ℓ_2	ℓ_0+ℓ_1+ℓ_2
Women Managers	52.24	16.35	43.91	18.59	16.35	17.31	16.35
CO_2 Emissions to Revenues	13.14	37.50	50.00	37.50	37.82	37.18	37.50
Waste Recycled to Total Waste	12.50	8.33	54.49	14.42	8.01	17.31	8.33
Water Use to Revenues	7.05	12.50	22.12	13.14	12.18	16.03	12.50
Board Gender Diversity	9.62	19.87	32.69	18.59	20.19	19.23	19.87

Table 5.12: Average absolute change by input when desired efficiency is $E^* = 0.8$.

Input	ℓ_0	ℓ_0+ℓ_2	ℓ_2	ℓ_1	ℓ_0+ℓ_1	ℓ_1+ℓ_2	ℓ_0+ℓ_1+ℓ_2
Women Managers	0.288	0.066	0.090	0.076	0.064	0.071	0.066
CO_2 Emissions to Revenues	0.075	0.174	0.149	0.170	0.176	0.166	0.174
Waste Recycled to Total Waste	0.059	0.016	0.060	0.021	0.015	0.024	0.044
Water Use to Revenues	0.044	0.044	0.066	0.045	0.042	0.060	0.016
Board Gender Diversity	0.063	0.081	0.057	0.064	0.084	0.055	0.081

Table 5.13: Average number of inputs that change per firm at $E^* = 0.8$.

Cost Function	**Mean number of inputs changed**
ℓ_0	2.21
ℓ_0+ℓ_2	2.15
ℓ_2	2.84
ℓ_1	1.52
ℓ_0+ℓ_1	2.09
ℓ_1+ℓ_2	1.88
ℓ_0+ℓ_1+ℓ_2	2.03

As expected, relaxing the target from full efficiency to $E^* = 0.8$ reduces the overall frequency and magnitude of input changes across all cost functions. For instance, the percentage of firms requiring CO_2 adjustments under ℓ_2 decreases from 48.1% to 50%, while the average absolute CO_2 change drops from 0.19–0.21 to approximately 0.17. In contrast, Women Managers becomes more prominent under the sparse specification (ℓ_0), increasing from 44.9% of firms at $E^* = 1.0$ to 52.2%. This indicates that at moderate efficiency levels, managerial diversity improvements serve as a common, lower-cost path to partial efficiency. Waste Recycled remains the most frequently modified input under the ℓ_2 configuration (54.5%), confirming that environmental factors still dominate the adjustment space even under relaxed targets.

Overall, both the number of changing inputs (Table 5.13) and their average magnitudes decline slightly compared with $E^* = 1.0$, consistent with the intuition that smaller efficiency goals can be met through incremental rather than structural adjustments.

5.4 Conclusion

This chapter presented the empirical validation of the counterfactual-based DEA framework proposed in this thesis, demonstrating how the model provides interpretable and feasible improvement paths for ESG performance in the logistics sector. Unlike classical DEA formulations, which measure inefficiency through proportional input contractions, the counterfactual approach identifies the minimal and most meaningful adjustments that would make each firm attain a desired efficiency level. By integrating norm-based cost functions—ℓ_0, ℓ_1, and ℓ_2—the framework explicitly encodes managerial preferences between sparsity, proportionality, and smoothness, enabling a richer and more realistic interpretation of efficiency enhancement strategies.

From the empirical standpoint, the application to 312 logistics firms highlights several central insights. At full efficiency ($E^* = 1.0$), the model prescribes primarily environmental adjustments, with significant reductions in CO_2 emissions and improved waste recycling ratios being the most frequent and impactful levers for improvement. These patterns are consistent across smooth (ℓ_2) and mixed configurations, indicating that environmental efficiency remains the structural bottleneck in the sector. When the target is relaxed to $E^* = 0.8$, smaller and more diversified changes become sufficient: moderate improvements in social and governance dimensions, particularly in the proportion of women managers and board gender diversity, increasingly contribute to reaching the efficiency goal. This transition reveals that the model adapts naturally to the ambition of the target, shifting from large-scale structural interventions to incremental organizational

adjustments as the required efficiency level decreases.

Beyond its numerical results, the analysis demonstrates that counterfactual reasoning enhances the explanatory and prescriptive capacity of DEA. Traditional DEA often yields scalar efficiency scores with limited operational guidance; by contrast, counterfactual DEA specifies what to change, by how much, and at what implicit cost. This feature turns DEA into a genuine decision-support instrument capable of informing ESG policy design and managerial action. Furthermore, because the counterfactual solutions are grounded in optimization rather than statistical approximation, they provide concrete, data-consistent recommendations, bridging the gap between quantitative benchmarking and actionable strategy.

An additional contribution of this research is the potential to scale the methodology from individual firms to groups or sectors. Group counterfactual analysis extends the current firm-level formulation by identifying shared improvement trajectories that minimize the total cost of adjustment across several DMUs simultaneously. Such a model could, for instance, determine coordinated ESG targets for an entire supply chain, balancing individual firm flexibility with collective progress toward sustainability benchmarks. This perspective aligns with the growing emphasis on systemic efficiency, where the objective is not only to make each firm efficient in isolation but to guide the whole network toward environmentally and socially optimal performance.

5.5 Future Work

Several promising research directions arise from this study. First, integrating the counterfactual DEA framework with modern machine learning techniques—such as Random Forests and Logistic Regression—could enhance both predictive power and interpretability. Machine learning models could help identify which ESG indicators most strongly influence efficiency and predict which firms are most likely to achieve specific targets, providing a data-driven complement to optimization-based benchmarking. This hybridization of DEA and machine learning could yield a new class of models capable of both explaining and forecasting performance outcomes.

Second, future research could develop group counterfactual formulations that capture shared adjustment paths across firms or industrial clusters. Such an extension would allow identifying common leverage points that yield the greatest collective efficiency gains with minimal aggregate adjustment cost, supporting collaborative sustainability initiatives and policy design.

Third, the adoption of alternative efficiency measures—including additive, directional distance, or slack-based models—could provide additional flexibility in characterizing inefficiency and exploring how different definitions of “distance to

the frontier" influence counterfactual recommendations. Fourth, accounting for uncertainty and data variability, for example through stochastic programming or robust optimization, would increase the credibility of counterfactual suggestions when ESG indicators are subject to measurement error or reporting noise. Finally, dynamic counterfactual models could track firms over time, enabling longitudinal assessment of ESG convergence and quantifying the persistence of improvement effects across multiple periods.

Overall, this thesis contributes a rigorous and interpretable framework that advances the methodological boundaries of DEA and its application to sustainability analysis. By merging optimization-based efficiency evaluation with counterfactual reasoning, it establishes a bridge between abstract frontier modeling and the tangible realities of managerial decision-making. The framework not only measures how far firms are from best practice but also reveals how and through which variables they can realistically reach it. In doing so, it provides a foundation for transparent, data-driven, and strategically grounded ESG benchmarking. The insights obtained from both firm-level and aggregate analyses underscore the broader significance of counterfactual DEA: it transforms efficiency analysis into a language of actionable improvement—a necessary step toward embedding analytical rigor into the pursuit of sustainable development.

Bibliography

[1] A. Charnes, W. W. Cooper, and E. Rhodes. «Measuring the efficiency of decision making units». In: *European Journal of Operational Research* 2.6 (1978), pp. 429–444. DOI: 10.1016/0377-2217(78)90138-8 (cit. on pp. 1, 6, 10–13, 15).

[2] William W. Cooper, Lawrence M. Seiford, and Joe Zhu. «Data Envelopment Analysis: History, Models, and Interpretations». In: *Handbook on Data Envelopment Analysis*. Vol. 164. International Series in Operations Research & Management Science. Springer, 2011, pp. 1–39. DOI: 10.1007/978-1-4419-6151-8_1 (cit. on pp. 1, 6).

[3] Peter Bogetoft, Jasone Ramírez-Ayerbe, and Dolores Romero Morales. «Counterfactual Analysis and Target Setting in Benchmarking». In: *European Journal of Operational Research* (2024). Forthcoming. DOI: 10.1016/j.ejor.2024.01.005 (cit. on pp. 1, 2, 8, 16, 18–20, 26, 28).

[4] Riccardo Guidotti. «Counterfactual explanations and how to find them: literature review and benchmarking». In: *Data Mining and Knowledge Discovery* 38.8 (2022), pp. 2770–2824. DOI: 10.1007/s10618-022-00831-6 (cit. on pp. 2, 3, 8, 18–20).

[5] Sahil Verma, Varich Boonsanong, Minh Hoang, Keegan Hines, John Dickerson, and Chirag Shah. «Counterfactual Explanations and Algorithmic Recourses for Machine Learning: A Review». In: *ACM Computing Surveys* 56.12 (2024), pp. 1–42. DOI: 10.1145/3677119 (cit. on pp. 2, 3, 8).

[6] Elena Carrizosa, Dolores Romero Morales, and M. Teresa Rodríguez-Madroñero. «Counterfactual Explanations and Target Setting: A Mathematical Optimization Perspective». In: *Omega* (2024). Forthcoming. DOI: 10.1016/j.omega.2023.103123 (cit. on pp. 2, 8).

[7] M. J. Farrell. «The Measurement of Productive Efficiency». In: *Journal of the Royal Statistical Society: Series A (General)* 120.3 (1957), pp. 253–290. DOI: 10.2307/2343100 (cit. on pp. 6, 10, 11, 15).

[8] R. D. Banker, A. Charnes, and W. W. Cooper. «Some models for estimating technical and scale inefficiencies in Data Envelopment Analysis». In: *Management Science* 30.9 (1984), pp. 1078–1092. DOI: 10.1287/mnsc.30.9.1078 (cit. on pp. 6, 10, 11, 13, 32).

[9] William W. Cooper, Lawrence M. Seiford, and Kaoru Tone. *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software*. 2nd. New York, NY: Springer, 2007. ISBN: 978-0387452814 (cit. on pp. 6, 10, 12–15, 19).

[10] Beata Gavurova, Lubos Vagner, and Eva Huculova. «Fifty Years of Data Envelopment Analysis: A Bibliometric Review of Trends and Frontiers». In: *European Journal of Operational Research* (2024). DOI: 10.1016/j.ejor.2024.03.045 (cit. on p. 7).

[11] J. Aparicio, J. L. Ruiz, and I. Sirvent. «Closest targets and minimum distance to the Pareto-efficient frontier in DEA». In: *Journal of Productivity Analysis* 28 (2007), pp. 209–218. DOI: 10.1007/s11123-007-0041-1 (cit. on p. 7).

[12] Kaoru Tone and Makoto Tsutsui. «Development and Evolution of Slacks-Based Measure in Data Envelopment Analysis: A Review». In: *Journal of Economic Surveys* (2023). DOI: 10.1111/joes.12682 (cit. on p. 7).

[13] Jonathan F. Bard. *Practical Bilevel Optimization: Algorithms and Applications*. Dordrecht: Springer, 1998. DOI: 10.1007/978-1-4615-5661-7 (cit. on pp. 7, 16, 17).

[14] Stephan Dempe. *Foundations of Bilevel Programming*. Berlin, Heidelberg: Springer, 2002. DOI: 10.1007/978-1-4757-3614-2 (cit. on pp. 7, 16–18).

[15] Ankur Sinha, Pekka Malo, and Kalyanmoy Deb. «A Review on Bilevel Optimization: From Classical to Evolutionary Approaches and Applications». In: *IEEE Transactions on Evolutionary Computation* 22.2 (2018), pp. 276–295. DOI: 10.1109/TEVC.2017.2712906 (cit. on pp. 7, 16–18).

[16] Eoin Delaney and Derek Greene. «Robust Counterfactual Explanations in Machine Learning: A Survey». In: *arXiv preprint* (2024). eprint: 2402.01928. URL: https://arxiv.org/abs/2402.01928 (cit. on p. 8).

[17] Leif Hancox-Li and Zachary C. Lipton. «Counterfactuals and Causability in Explainable Artificial Intelligence». In: *arXiv preprint* (2021). eprint: 2103.04244. URL: https://arxiv.org/abs/2103.04244 (cit. on p. 8).

[18] Matt Kusner, Joshua R. Loftus, and Chris Russell. «The Use and Misuse of Counterfactuals in Ethical Machine Learning». In: *arXiv preprint* (2021). eprint: 2102.05085. URL: https://arxiv.org/abs/2102.05085 (cit. on p. 8).

[19] Rafael Poyiadzi, Shubham Joshi, and Robert Turner. «Evaluating the Practicality of Counterfactual Explanations». In: *Proceedings of the Workshop on Human-Centric Explainable AI*. 2023. URL: https://ceur-ws.org/Vol-3277/paper3.pdf (cit. on p. 8).

[20] Chiang Kao. «Network Data Envelopment Analysis and Its Applications (2017–2022)». In: *Mathematics* 11.9 (2023), p. 2141. DOI: 10.3390/math11092141 (cit. on p. 9).

[21] Edward Elson Kosasih, Emmanuel Papadakis, George Baryannis, and Alexandra Melike Brintrup. «Explainable Artificial Intelligence in Supply Chain Management: A Systematic Review of Neurosymbolic Approaches». In: *International Journal of Production Research* (2023). DOI: 10.1080/00207543.2023.2281663 (cit. on p. 9).

[22] Kaoru Tone. «A Slack-Based Measure of Efficiency in Data Envelopment Analysis». In: *European Journal of Operational Research* 130.3 (2001), pp. 498–509. DOI: 10.1016/S0377-2217(99)00407-5 (cit. on p. 14).

[23] Robert G. Chambers, Yangho Chung, and Rolf Färe. «Benefit and Distance Functions». In: *Journal of Economic Theory* 70.2 (1996), pp. 407–419. DOI: 10.1006/jeth.1996.0096 (cit. on p. 15).

[24] Per Andersen and Niels Christian Petersen. «A Procedure for Ranking Efficient Units in Data Envelopment Analysis». In: *Management Science* 39.10 (1993), pp. 1261–1264. DOI: 10.1287/mnsc.39.10.1261 (cit. on p. 15).

[25] Douglas W. Caves, Laurits R. Christensen, and W. Erwin Diewert. «The Economic Theory of Index Numbers and the Measurement of Input, Output, and Productivity». In: *Econometrica* 50.6 (1982), pp. 1393–1414. DOI: 10.2307/1913388 (cit. on p. 15).

[26] Ole B. Olesen, Niels C. Petersen, and Victor V. Podinovski. «Efficiency Analysis with Inaccurate Data: DEA Models». In: *European Journal of Operational Research* 247.1 (2015), pp. 183–193. DOI: 10.1016/j.ejor.2015.05.056 (cit. on p. 15).