PhD Program in Statistics, Optimization, and Applied Mathematics

Technical Efficiency Estimation Using Adaptive Constrained Enveloping Splines

Víctor Javier España Roch

Universidad Miguel Hernández of Elche

2024

Juan Aparicio Baeza

Supervisor

Universidad Miguel Hernández de Elche

Josep Xavier Barber Vallès

Co-supervisor

Universidad Miguel Hernández de Elche

QUALITY INDICATORS OF THE PhD THESIS

Published articles:

1. España, V. J., Aparicio, J., Barber, X., & Esteve, M. (2024). Estimating production functions through additive models based on regression splines. *European Journal of Operational Research, 312*(2), 684–699.
https://doi.org/10.1016/j.ejor.2023.06.035

2. España, V. J., Aparicio, J., & Barber, X. (2024). An adaptation of Random Forest to estimate convex non-parametric production technologies: an empirical illustration of efficiency measurement in education. *International Transactions in Operational Research.*
https://onlinelibrary.wiley.com/doi/full/10.1111/itor.13561

REPORT FROM THESIS SUPERVISOR AND CO-SUPERVISOR

Juan Aparicio Baeza, thesis supervisor, Full Professor in the Department of Statistics, Mathematics, and Computer Science at Miguel Hernández University of Elche and Director of the Operations Research Center.

Josep Xavier Barber Vallès, thesis co-supervisor, Full Professor in the Department of Statistics, Mathematics, and Computer Science at Miguel Hernández University of Elche.

HEREBY CERTIFY THAT:

Mr. Víctor Javier España Roch has conducted, under our supervision, the research entitled "**Technical Efficiency Estimation Using Adaptive Constrained Enveloping Splines**". This work has been carried out in accordance with the terms and conditions set out in his Research Plan and in compliance with the Code of Good Practice of Miguel Hernández University of Elche. The research has satisfactorily met the objectives and requirements to be defended publicly as a doctoral thesis.

Signed for all relevant purposes, in Elche, on the 29 day of September, 2025.

Thesis Supervisor
Prof. Dr. Juan Aparicio Baeza

Thesis Co-supervisor
Prof. Dr. Josep Xavier Barber Vallès

REPORT FROM THE ACADEMIC COMMITTEE OF THE DOCTORAL PROGRAM

Dr. Domingo Morales González, Coordinator of the Doctoral Program in Statistics, Optimization, and Applied Mathematics,

HEREBY CERTIFY THAT:

Mr. Víctor Javier España Roch has conducted, under the supervision of our Doctoral Program, the research work titled "**Technical Efficiency Estimation Using Adaptive Constrained Enveloping Splines**". This work has been carried out in accordance with the terms and conditions defined in his Research Plan and in compliance with the Code of Good Practice of Miguel Hernández University of Elche. The research has satisfactorily met the objectives and requirements for its public defense as a doctoral thesis.

Signed for all relevant purposes, in Elche, on the 29 day of September, 2025.

Director of the Doctoral Program in Statistics, Optimization, and Applied Mathematics

Prof. Dr. Domingo Morales González

FUNDING

- The ACIF/2021/135 grant, funded by the Conselleria d'Educació, Investigació, Cultura i Esports (Generalitat Valenciana) and European Social Fund (European Union).
- The CIBEFP/2023/041 grant, funded by the Conselleria d'Educació, Investigació, Cultura i Esports (Generalitat Valenciana).
- The R&D&I PID2019-105952GB-I00 project, funded by Ministerio de Ciencia e Innovación / Agencia Estatal de Investigación / 10.13039 / 501100011033.
- The R&D&I PID2022-136383NB-I00 project, funded by Ministerio de Ciencia, Innovación y Universidades / Agencia Estatal de Investigación / 10.13039 / 501100011033 and co-funded by European Regional Development Fund (European Union).

Agradecimientos

Nunca imaginé ser doctor. Probablemente ni siquiera sabía lo que eso significaba, hasta que un día me encontré en la línea de salida sin haberlo siquiera planeado. Solo un paso que llevó a otro, y cuando quise darme cuenta, ya estaba dentro. Nunca me vi iniciando una carrera investigadora, ni pensé que tendría que aprender a convivir con la sucesión infinita de tropiezos que suele traer consigo. Porque no, no soy ningún genio. Tampoco esperaba que el "esto no funciona" se convirtiera en una especie de banda sonora de mi día a día, ni imaginé lo complicado que sería avanzar cuando no tienes del todo claro hacia dónde te diriges. Resulta que investigar no es tanto tener grandes ideas, sino resistir mientras todo lo demás falla. Nunca pensé que pasaría tantas horas escribiendo, ejecutando, esperando, preparando clases, corrigiendo... y volviendo a empezar, una y otra vez, como si todo se tratara de un bucle mal programado.

Y sí, volvería a pasar por todo eso otra vez. No por la alegría de publicar unos cuantos artículos o algún que otro capítulo de libro —que, en cierto modo, la ha habido—. Tampoco por la docencia universitaria, que, en fin... tiene lo suyo. Lo volvería a hacer por la gente que me he cruzado por el camino. Granada, Badajoz, Madrid, León, Córdoba, Santiago, Copenhague, Málaga, Belfast, Bilbao, Lleida, Milán... y otros tantos que ya asoman en el calendario. Esta experiencia me ha permitido viajar, y en cada viaje conocer a gente buena, ni más ni menos. Gente con la que he compartido alguna que otra charla científica, pero también muchas cervezas y, sí, seguramente, alguna que otra mañana complicada. Por lo tanto, mi primer agradecimiento es para la comunidad de la SEIO. Porque, sin ellos, el doctorado habría sido mucho más aburrido.

También ha sido fundamental el tiempo compartido con toda la gente del CIO, con quienes he convivido durante estos últimos años. Como no podía ser de otra forma, empiezo por mis directores de tesis: Juan y Xavi. Os agradezco la oportunidad de iniciar este proyecto con vosotros, la confianza depositada en mí, el apoyo constante durante estos años y el impulso para seguir avanzando y mejorando día a día. Me alegra haber recorrido este camino en vuestro equipo. Y no quiero olvidarme de quienes, sin estar metidos de lleno en esto, me han echado un cable cuando más lo necesitaba: ya fuera en lo académico o, sobre todo, en lo personal. Gracias.

A mis compañeros del CIO: María G., María B., Ricardo, Lorena, Laura, Raul, Daniel, Esteban, Jesús y Juan Carlos. He escuchado muchas veces que hemos construido un grupo muy guay, y creo que es cierto. Gracias por el día a día: por las conversaciones que no llevaban a ningún sitio, por las comidas y cenas juntos, por las fiestas y las calçotadas, por los días de chill en casa, por la hora del descanso (a veces literal), por los ánimos cuando algo no salía y por los consejos, que no han sido pocos. Gracias por hacer que venir al despacho no fuera solo venir a trabajar. Por estar, sin más, y por conseguir que todo esto haya sido mucho más llevadero. También a vosotros, gracias.

A Sofía. Primero compañera de trabajo, luego compañera de todo. Si lidiar con una tesis es complicado, con dos a la vez podría haber sido imposible. Y, sin embargo, lo hemos hecho. Gracias por estar siempre ahí, sobre todo cuando no era sencillo. Por echarme una mano cuando tú necesitabas dos. Por recordarme una y otra vez que había que parar, aunque fuera solo un momento. Gracias por hacer tuyos mis problemas, por tu paciencia infinita y por acompañarme en todo lo importante, especialmente en esos días cotidianos que pasan desapercibidos, pero que al final lo son todo. Esta etapa habría sido muy diferente sin ti, y no una mejor.

Finalmente, a mi familia y a mis amigos. Por brindarme vuestro apoyo en todo momento y por facilitarme tanto el camino. Por la paciencia, por estar siempre ahí y por entender que, de algún modo u otro, siempre había algo que entregar pronto. A partir de ahora, dejaré de ser el último que llega y el primero que se va. Gracias por haberme hecho la vida más fácil.

Abstract

The accurate measurement of technical efficiency represents a central goal in both theoretical and applied economic analysis, as it allows organizations, industries, and policymakers to assess performance, allocate resources efficiently, and identify areas for improvement. Among the various tools developed for this purpose, Data Envelopment Analysis (DEA) has emerged as a predominant non-parametric methodology, widely adopted for its conceptual simplicity and operational flexibility. DEA constructs a production frontier enveloping observed data without requiring prior specification of a functional form, making it suitable for diverse applications across sectors such as education, healthcare or banking.

However, despite its widespread use, DEA presents several important methodological limitations that hinder its broader applicability and accuracy in practice. Chief among these are issues related to overfitting, particularly when the number of inputs and outputs is large relative to the sample size, leading to overly optimistic efficiency estimates. Additionally, DEA lacks a natural framework for statistical inference, preventing the derivation of confidence intervals or hypothesis testing without relying on complex and computationally intensive bootstrap procedures. A further critical limitation lies in the absence of systematic guidance for variable selection, which makes the analysis highly sensitive to the analyst's choices and susceptible to distortions from irrelevant or redundant variables. These challenges become even more pronounced in high-dimensional data environments, where relationships among variables tend to be complex and nonlinear, often exceeding the capacity of DEA to accurately capture the underlying production structure.

The thesis presents a unified family of techniques for estimating production frontiers, designed to address key limitations of traditional DEA, such as overfitting, limited robustness, and challenges in high-dimensional settings. This family includes three complementary methods: Adaptive Constrained Enveloping Splines (ACES), which offers a flexible estimator of technical efficiency; Random Forest-ACES (RF-ACES), which enhances robustness through ensemble learning; and Quick-ACES (Q-ACES), which focuses on computational efficiency for large-scale applications. Each method addresses different empirical needs, enabling researchers to select the most appropriate approach based on the characteristics of the data.

At the core of this framework lies ACES, a method built upon an adapted version of Multivariate Adaptive Regression Splines (MARS), specifically tailored for production frontier estimation. ACES integrates essential shape constraints—monotonicity and concavity—into a spline-based, non-parametric regression model, ensuring consistency with microeconomic theory. The estimation is formulated as a constrained optimization problem and follows a two-stage procedure: first, a rich set of candidate basis functions is generated through forward selection; then, the model is refined via backward elimination guided by generalized cross-validation. This process yields a flexible estimator capable of modeling complex, nonlinear input–output relationships while avoiding the overfitting commonly associated with DEA. A key strength of ACES is its ability to remain fully deterministic while achieving strong generalization beyond the observed sample—representing a major advance over traditional enveloping methods.

To improve robustness and mitigate the sensitivity of spline-based models to local data configurations, the thesis extends ACES into an ensemble version named RF-ACES. This method builds on the principles of bagging and random feature selection, inspired by the Random Forest algorithm. In RF-ACES, multiple ACES models are trained on bootstrap samples of the original dataset using randomly selected subsets of inputs at each iteration. The resulting estimators are aggregated to form a final predictor that is significantly more stable and less sensitive to random noise. An important strength of RF-ACES lies in its ability to provide internal variable importance measures, which can be used to guide dimensionality reduction and identify the most influential inputs. This makes it particularly effective in high-dimensional settings or when irrelevant variables are suspected to distort the estimation. Nevertheless, this increase in robustness comes at the cost of computational burden, since multiple constrained estimations must be performed and aggregated.

Finally, to address computational limitations and ensure scalability to large datasets or time-sensitive applications, the thesis proposes a third member of the family: Q-ACES. This accelerated variant introduces a set of heuristic strategies designed to reduce the computational burden of the estimation process without compromising the theoretical principles underlying ACES. These strategies include input pre-selection based on correlation analysis, reduction of knot sets through neighborhood analysis derived from DEA projections, and adaptive filtering of candidate basis functions during the forward selection phase. In addition to these mechanisms, Q-ACES incorporates a new automated procedure for variable selection, which serves as an alternative to the Random Forest-based relevance assessment implemented in RF-ACES. This procedure allows for efficient identification of the most influential inputs while preserving the model's accuracy. As a result, Q-ACES achieves substantial improvements in execution time and memory efficiency, enabling the application of shape-constrained frontier estimation in large-scale scenarios where the original ACES framework would be computationally impractical.

Together, these three methods—ACES, RF-ACES, and Q-ACES—constitute a flexible and modular toolkit for technical efficiency analysis. Analysts can select the most appropriate variant depending on the size and complexity of the dataset, the tolerance for approximation, and the need for robustness. This family-based approach allows practitioners to move beyond the static and sample-dependent nature of DEA, adopting frontier estimators that are not only theoretically grounded but also adaptive to modern data analysis challenges. The proposed methodologies have been extensively validated through hundreds of simulation experiments, covering a wide range of scenarios with varying dimensionality, noise levels, and production structures. Results consistently confirm the competitiveness of ACES and its variants against established techniques such as DEA, Corrected Concave Non-parametric Least Squares (CCNLS), Stochastic Non-Smooth Envelopment of Data (StoNED), and Bootstrap DEA, often yielding more accurate and stable estimates. Furthermore, the thesis offers practical guidance on how to configure and tune ACES in different empirical contexts, helping researchers make informed decisions to maximize performance and reliability in their applications.

Resumen

La medición de la eficiencia técnica representa un objetivo relevante tanto en el análisis económico teórico como en el aplicado, ya que permite a organizaciones, industrias y responsables políticos evaluar el desempeño, asignar recursos de forma eficiente e identificar áreas de mejora. Entre las diversas herramientas desarrolladas con este fin, el Análisis Envolvente de Datos (en inglés, *Data Envelopment Analysis*, DEA) ha surgido como una de las metodologías no paramétricas predominantes, ampliamente adoptada por su simplicidad conceptual y flexibilidad operativa. DEA construye una frontera de producción que envuelve los datos observados sin requerir la especificación previa de una forma funcional, lo que la hace adecuada para aplicaciones en sectores tan diversos como la educación, la sanidad o la banca.

Sin embargo, a pesar de su uso extendido, DEA presenta importantes limitaciones metodológicas que dificultan su aplicabilidad y precisión en contextos reales. Entre ellas destacan los problemas de sobreajuste, especialmente cuando el número de inputs y outputs es elevado en relación con el tamaño de la muestra, lo que conduce a estimaciones de eficiencia excesivamente optimistas. Además, DEA carece de un marco natural para la realización de inferencia estadística, lo que impide determinar intervalos de confianza o contrastes de hipótesis sin recurrir a procedimientos bootstrap complejos y computacionalmente costosos. Otra limitación crítica es la ausencia de un criterio sistemático para la selección de variables, lo que hace que los resultados dependan en exceso del juicio del analista y puedan verse distorsionados por variables irrelevantes o redundantes. Estos desafíos se agravan en entornos de datos de alta dimensión, donde las relaciones entre variables tienden a ser complejas y no lineales, superando con frecuencia la capacidad de DEA para capturar con precisión la estructura subyacente del proceso productivo.

La tesis presenta una familia unificada de técnicas para la estimación de fronteras de producción, diseñada para abordar estas limitaciones clave de DEA, tales como el sobreajuste, la falta de robustez y las dificultades en entornos de alta dimensión. Esta familia incluye tres métodos complementarios: Adaptive Constrained Enveloping Splines (ACES), que ofrece un estimador flexible de la eficiencia técnica; Random Forest-ACES (RF-ACES), que mejora la robustez mediante la agregación de modelos; y Quick-ACES (Q-ACES), que se centra en la eficiencia computacional para aplicaciones a gran escala. Cada uno de estos métodos responde a necesidades empíricas distintas, permitiendo al investigador seleccionar la alternativa más adecuada según las características del conjunto de datos.

En el núcleo de este marco se encuentra ACES, un método basado en una adaptación del algoritmo Multivariate Adaptive Regression Splines (MARS), ajustado específicamente para la estimación de fronteras de producción. ACES integra restricciones de forma esenciales —monotonía y concavidad— en un modelo de regresión no paramétrica con splines, garantizando así la coherencia con la teoría microeconómica. El procedimiento de estimación se formula como un problema de optimización con restricciones y sigue un enfoque en dos etapas: primero, se genera un conjunto amplio de funciones base mediante selección hacia adelante (forward selection); después, el modelo se depura con un paso de selección hacia atrás (backward elimination) guiado por validación cruzada generalizada. El resultado es un estimador flexible, capaz de capturar relaciones input–output complejas y no lineales, evitando al mismo tiempo el sobreajuste característico de DEA. Una de las principales fortalezas de ACES es su capacidad para mantenerse completamente determinista y, aun así, generalizar más allá de la muestra observada —lo que representa un avance significativo frente a los métodos envolventes tradicionales.

Para mejorar la robustez y reducir la sensibilidad de los modelos basados en splines a la configuración local de los datos, la tesis amplía ACES mediante una versión agregada denominada RF-ACES. Este método se inspira en los principios del bagging y la selección aleatoria de variables característicos del algoritmo Random Forest. En RF-ACES, múltiples modelos ACES se entrenan sobre muestras bootstrap del conjunto de datos original, utilizando subconjuntos aleatorios de inputs en cada iteración. Los estimadores resultantes se agregan para formar un predictor final más estable y menos sensible al ruido aleatorio. Una ventaja adicional de RF-ACES es su capacidad para proporcionar medidas internas de importancia de las variables, útiles para guiar la reducción dimensional e identificar los inputs más influyentes. Esto lo convierte en una opción especialmente eficaz en contextos de alta dimensionalidad o cuando se sospecha que hay variables que distorsionan la estimación. No obstante, esta mejora en robustez conlleva un mayor coste computacional, ya que implica realizar múltiples estimaciones con restricciones y combinarlas.

Para abordar las limitaciones computacionales y garantizar la escalabilidad del método en conjuntos de datos extensos o en aplicaciones que requieren tiempos de respuesta reducidos, la tesis propone un tercer miembro de la familia: Q-ACES. Esta variante acelerada introduce un conjunto de estrategias heurísticas diseñadas para reducir la carga computacional del proceso de estimación sin comprometer los principios teóricos que sustentan ACES. Entre estas estrategias se incluyen la preselección de variables basada en análisis de correlación, la reducción del conjunto de *knots* mediante análisis de vecindad a partir de proyecciones DEA, y el filtrado adaptativo de funciones base durante la fase de selección hacia adelante. Además, Q-ACES incorpora un nuevo procedimiento automático de selección de variables, que constituye una alternativa al mecanismo basado en Random Forest utilizado en RF-ACES. Este procedimiento permite identificar de manera eficiente los inputs más relevantes, preservando al mismo tiempo la precisión del modelo. Gracias a estas mejoras, Q-ACES logra reducciones sustanciales en el tiempo de ejecución y el uso de memoria, lo que permite aplicar estimadores de frontera con restricciones de forma en escenarios a gran escala donde el uso de ACES sería impracticable.

Conjuntamente, estos tres métodos —ACES, RF-ACES y Q-ACES— conforman una caja de herramientas flexible y modular para el análisis de eficiencia técnica. El analista puede seleccionar la variante más adecuada en función del tamaño y la complejidad del conjunto de datos, el grado de tolerancia al uso de heurísticas y la necesidad de robustez. Este enfoque basado en una familia de métodos permite superar la naturaleza estática y dependiente de la muestra de DEA, y adoptar estimadores de frontera que no solo están fundamentados teóricamente, sino que también se adaptan a los desafíos contemporáneos del análisis de datos. Las metodologías propuestas han sido validadas mediante cientos de experimentos de simulación, que cubren un amplio rango de escenarios con diferentes niveles de dimensionalidad, estructura productiva y complejidad funcional. Los resultados confirman de forma consistente la competitividad de ACES y sus variantes frente a técnicas consolidadas como DEA, Corrected Concave Non-parametric Least Squares (CCNLS), Stochastic Non-Smooth Envelopment of Data (StoNED) o Bootstrap DEA, ofreciendo en muchos casos estimaciones más precisas y estables. Además, la tesis proporciona una guía práctica para configurar y ajustar ACES según el contexto empírico, ayudando a los investigadores a tomar decisiones fundamentadas para maximizar el rendimiento y la fiabilidad de sus análisis.

List of Contents

Chapter 1

Introduction

This chapter provides an overview of the study of efficiency analysis, establishing the conceptual framework that will guide the development of this doctoral thesis. It covers the fundamental pillars of the field, highlighting the importance of this area of study for the operational success of organizations. In addition, the historical context of current efficiency measurement techniques and their potential areas of application are introduced. A classification of the different approaches used for this purpose is also presented, pointing out certain competitive advantages of some methods over others. Special attention is given to Data Envelopment Analysis (DEA), the predominant non-parametric technique in efficiency analysis studies. Finally, recent research integrating statistical approaches for efficiency evaluation is reviewed, aiming to overcome some of the limitations associated with the traditional DEA method.

1.1 Literature review on efficiency analysis

The current economic framework is characterized by strong interconnectedness and increasing interdependence of global markets. In this context, organizations must be highly adaptable to changes in market dynamics. From a financial perspective, institutions need to analyze consumption trends, fiscal and monetary policies in the countries where they operate, or changes in labor markets, among other factors. These elements often fluctuate rapidly and unpredictably, adding further complexity to decision-making processes. Understanding the macroeconomic environment enables business entities to position themselves strategically in the face of challenges and opportunities, allowing them to make informed decisions regarding potential expansion, investment, and adaptation of their business models.

However, while this economic analysis provides invaluable insights into possible courses of action, organizational success equally depends on an entity's ability to execute internal operations efficiently. In this regard, achieving operational efficiency means maximizing the quantity of products and/or services delivered (outputs), assuming an acceptable quality, while simultaneously reducing the consumption of resources and operating costs (inputs).

The estimation of production frontiers and the measurement of the relative performance of institutions, known as Decision Making Units (DMUs), is based on the systematic comparison of entities employing similar production processes with the goal of identifying and correcting operational inefficiencies. It is important to note that inefficiency can stem from the excessive use of resources for a given level of output, from an insufficient level of output for a fixed amount of input, or from any combination that implies a reduction in input use and an increase in output. Take, for example, an automobile assembly plant, where workers, materials, and energy are evaluated as inputs, against the quantity and quality of vehicles produced as outputs. In this case, Factory A would be considered inefficient if it requires more employees and energy than Factory B to produce a similar number and quality of vehicles.

The applications of these evaluations range from internal analysis within a single organization to comparisons between multiple entities in the same sector. Internally, a company may conduct a comparative study of efficiency among its different divisions, branches, or departments to identify and replicate the most successful strategies. On the other hand, inter-organizational comparisons aim to assess how different institutions in the same sector perform in key aspects such as production efficiency, innovation, or sustainability. These evaluations are not limited to for-profit organizations—whose main goal is economic—but also include non-profit institutions such as schools, hospitals, or public administrations. Moreover, the scope of analysis may extend to entire industries, specific projects, individual agents, or, more generally, any group of homogeneous entities capable of transforming a common set of resources into a set of products or services.

From a historical perspective, the evolution of efficiency analysis has been shaped by the contributions of several key authors. The works of Koopmans (1951) and Debreu (1951) laid the foundational concepts for research in the field of efficiency measurement. Koopmans characterized the concept of technical efficiency based on the definition of technically efficient (non-dominated) input and output vectors, while Debreu was the first to propose a measure of technical efficiency through his "coefficient of resource utilization".

Later, Farrell (1957), who questioned the use of average performance to assess efficiency, suggested that institutional efficiency consists of two components: technical efficiency, which reflects the ability of a firm to maximize output from a given set of resources, and allocative efficiency, which reflects the ability to use inputs in optimal proportions, considering market prices and production technology.

These efficiency measures assume the existence of a production function underlying the observed data sample, which must be estimated. In this context, Farrell (1957) proposed two different approaches to achieve this objective: a piecewise linear, convex, non-parametric approach and, alternatively, the use of parametric production functions. Once the production function is estimated, Farrell's efficiency measure—inspired by the distance functions previously introduced by Shephard (1953)— is based on performing radial movements (i.e., proportional changes in inputs or outputs) from inefficient DMUs towards technically efficient units that lie on the estimated production frontier.

Subsequently, Aigner and Chu (1968) expanded on the work initiated by Debreu and Farrell by adopting a deterministic approach to assess technical efficiency. Their proposal required the estimation of a Cobb-Douglas-type parametric production function via linear and quadratic programming techniques, imposing the condition that residuals from this estimation remain positive. In this deterministic framework, any deviation of DMUs from the estimated production function is directly interpreted as a source of inefficiency. However, this approach faces limitations, as factors such as measurement errors and statistical noise—including the impact of unfavorable weather conditions, strikes, or other random events—could be mistakenly classified as indicators of technical inefficiency. To address this issue, Aigner et al. (1977) and Meeusen and van Den Broeck (1977) conducted foundational research aimed at identifying both random influences and sources of technical inefficiency among the possible causes of deviations from the production frontier, leading to the creation of the methodology known as Stochastic Frontier Analysis (SFA).

Additionally, it is important to note that deterministic approaches impose enveloping conditions on all units in the sample, unlike stochastic frontier approaches, where the presence of statistical noise allows for some units to lie above the estimated frontier. This difference results in comparatively lower robustness of deterministic approaches relative to stochastic ones, especially in the presence of outliers. On the other hand, stochastic methods offer the additional advantage of enabling statistical inference on the results, providing deeper insights of the nature of the observed inefficiency. For example, it becomes possible to calculate confidence intervals for the technical efficiency estimates.

The previous paragraphs classified efficiency analysis methods based on their deterministic or stochastic nature. Another distinction concerns whether the methods are parametric or non-parametric. Non-parametric approaches, which generally adopt a deterministic perspective, do not require a specific functional form for the underlying technology, such as a Cobb-Douglas production function. This independence provides these approaches with greater flexibility in modeling datasets with complex structures. In contrast, parametric methods—typically associated with a stochastic approach—require the specification of a concrete production function, which may compromise model fit. Given the context of our study and the particular characteristics of the methodologies proposed in this thesis, we focus on the application and analysis of non-parametric, deterministic methods.

Two decades after Farrell's pioneering work, and building on his principles, Charnes et al. (1978) and Banker et al. (1984) introduced the most widely recognized and extensively applied non-parametric methodology for efficiency estimation: DEA. This deterministic approach aimed to provide a satisfactory procedure for evaluating the relative efficiency of units whose production process involves multiple inputs and multiple outputs. The original idea behind DEA focused on identifying DMUs exhibiting optimal production practices, thus forming the efficient frontier (also known as the best-practice frontier). In this way, DEA emerged as an alternative to econometric methods, using a mathematical programming perspective to estimate production frontiers. Moreover, the efficiency estimates generated by DEA represent a natural extension of the classic output/input ratio, eliminating the need to assign specific weights to the variables involved. This characteristic gives DEA a competitive advantage over most parametric methods, such as SFA, which generally requires aggregating multiple outputs into a single production index to handle multi-output settings. Nevertheless, neither DEA nor SFA is entirely suitable for all applications. In this regard, research efforts have aimed to integrate stochastic elements into DEA, as in the work of Olesen and Petersen (2016), or to increase the flexibility of SFA, as illustrated in the studies by Griffin and Steel (2008).

Regarding areas of application, Liu et al. (2013) highlight five key sectors with significant use of efficiency analysis via DEA: banking, healthcare, agriculture and livestock, transportation, and education. Notable early applications within these sectors include Sherman and Gold (1985), who analyzed the operational efficiency of 14 savings bank branches by evaluating services provided and resources used; Nunamaker (1983), who examined nursing service efficiency across several hospitals in Wisconsin; Coelli (1995), who explored the application of production frontier estimation in agricultural economics; Schefczyk (1993), who measured operational performance among 15 international airlines; and Charnes et al. (1981), who assessed the impact of the Follow Through educational project in United States public schools.

More recently, research has further expanded DEA applications in these sectors, as illustrated by Tsolas et al. (2020) and Antunes et al. (2022) (banking), Md Hamzah et al. (2021) (healthcare), Kyrgiakos et al. (2023) (agriculture), Izadikhah et al. (2021) (transportation), and Aparicio et al. (2022b) (education). Additionally, various studies have compiled DEA applications across specialized fields, including actuarial science (Kaffash et al., 2020), energy and environmental analysis (Sueyoshi et al., 2017), sustainability (Zhou et al., 2018), manufacturing (Jain et al., 2011), public administration (Benito et al., 2019), sports management (Haas, 2003), and transportation (Barros, 2005).

For comprehensive reviews on efficiency estimation methods using DEA, SFA, and related approaches, readers are referred to Førsund et al. (1980), Färe et al. (1985, 1994), Schmidt (1985), Bauer (1990), Battese (1991), Lovell et al. (1993), and Greene et al. (1993).

1.2 On the limitations of Data Envelopment Analysis

Classical methodologies for estimating production frontiers, such as DEA[1], have been criticized for their non-statistical nature. Some researchers have even characterized these techniques as merely descriptive tools with limited inferential capacity (see, for example, Esteve et al. 2020; Tsionas 2022; Valero-Carreras et al. 2022; Aparicio and Esteve 2023). Standard enveloping methods position the efficient frontier as close as possible to the observed data by following the principle of minimal extrapolation. While such approaches yield precise efficiency measurements for the observations within the sample, they are prone to overfitting. In other words, these methods tend to conform too closely to the idiosyncrasies of the study's data (i.e., data in-sample), which results in reduced generalizability when faced with new data (i.e., data out-of-sample). In this context, the estimated frontier captures the specific characteristics of the analyzed units but may not faithfully represent the behavior of units not included in the analysis. A clear example arises in evaluating the efficiency of schools using results from PISA (Programme for International Student Assessment). Here, a DEA frontier estimated from the schools available in the sample might lead to erroneous interpretations about the educational system's efficiency at a regional or national level by ignoring the diversity of centers not initially included. Moreover, DEA results assume that the best practices are observed in the analyzed DMUs, meaning that the efficiency scores depend solely on the particular data sample used for estimation.

In this context, Korostelëv et al. (1995) demonstrated that when DEA is applied to a sample of independently and identically distributed observations drawn from a data generating process (DGP), the resulting frontier estimate is systematically lower than the true frontier defined by the DGP itself. This overfitting phenomenon significantly affects the results, leading many DMUs to be incorrectly classified as efficient. This suggests that caution is needed before considering a DMU truly efficient, even when DEA results indicate so. In this sense, DEA can be considered overly optimistic—a tendency that becomes more pronounced as the dimensionality of the analysis increases.

To address the problem of estimating technical efficiency at a theoretical level—rather than relying solely on specific samples—numerous researchers over the past decades have developed alternative methodologies to complement or even replace DEA analysis. For example, Simar and Wilson (1998, 2000) implemented bootstrapping techniques (Efron 1992) to construct confidence intervals around the efficiency scores obtained through DEA. Aragon et al. (2005) developed a non-parametric estimator for the efficiency frontier using conditional quantiles derived from the distribution of the production process. Building on this foundation, Daouia and Simar (2007) further explored this approach . Kuosmanen and Johnson (2010) presented the Corrected Concave Non-parametric Least Squares (CCNLS)

[1]Other non-parametric deterministic techniques, such as Free Disposal Hull, introduced by Deprins et al. (1984), exhibit limitations very similar to those found in DEA.

regression as a new interpretation of DEA, aiming to directly estimate the theoretical production function that generated the observed data. Parmeter and Racine (2013) introduced smooth estimators, both non-parametric and semi-parametric, to define production frontiers that comply with the theoretical axioms of production theory. Similarly, Daouia et al. (2016) proposed a estimation methodology to determine supporting frontiers, applying quadratic or cubic spline smoothing techniques. Finally, Ouenniche and Tone (2017) proposed a framework for out-of-sample evaluation in DEA, with applications in predicting institutional bankruptcies.

From a stochastic point of view, several authors have introduced methods incorporating both inefficiency and statistical noise into frontier estimation. An early contribution is the stochastic DEA formulation by Banker et al. (1988), who introduced a linear programming-based approach to estimate a frontier that accounts explicitly for statistical noise, positioning it within the data rather than strictly enveloping it. This model was further extended by Banker and Maindiratta (1992) into a semiparametric framework, incorporating maximum likelihood estimation and specific distributional assumptions about inefficiency and noise. More recently, Kuosmanen and Kortelainen (2012) introduced the Stochastic Non-Smooth Envelopment of Data (StoNED) method, which combines the nonparametric DEA approach with the SFA framework, aiming to estimate production frontiers while accounting for both inefficiency and statistical noise. Finally, Kuosmanen et al. (2015) and Kuosmanen and Johnson (2017) developed a consistent nonparametric estimator of the Directional Distance Function introduced by Chambers et al. (1998) using StoNED.

We present a review of relevant contributions in the literature on the adaptation of machine learning (ML) methods for production frontier and technical efficiency estimation. This reflects the growing interest within the DEA community in exploring the intersection between efficiency analysis, ML, Big Data (see, for example, Dellnitz, 2022 and Zhu, 2022), and frontier estimation. This research stream aims to address the overfitting issue often associated with traditional efficiency estimation approaches by providing accurate estimates of the DGP. Notably, this review excludes two-stage models, where DEA is first used to compute efficiency scores for the DMUs, and a ML model is then employed to predict technical efficiency based on inputs, outputs, and other exogenous variables (see, for example, Emrouznejad and Anouze, 2010 and Rebai et al., 2020). Instead, we focus on contributions that embed ML principles directly into the frontier estimation process, offering novel methodological frameworks that enhance the robustness and generalizability of efficiency analysis.

A first example can be found in the work of Olesen and Ruggiero (2018), who explored alternative approaches to the Afriat-Diewert-Parkan (Afriat, 1972; Diewert and Parkan, 1983) estimators by considering a weighted average of hinge functions (Breiman, 1993) with constraints on parameter estimation. In later studies, Olesen and Ruggiero (2022) used these hinge functions again to estimate S-shaped production functions, which allow for both increasing and decreasing returns to scale.

Esteve et al. (2020) introduced the Efficiency Analysis Trees (EAT) technique to estimate non-convex production frontiers in the style of Free Disposal Hull (FDH), using a modified version of the Classification and Regression Trees algorithm originally proposed by Breiman et al. (1984). Building on these ideas, Esteve et al. (2021), Esteve et al. (2023), and Guillen et al. (2023a) significantly improved the robustness and performance of the results produced by the EAT algorithm. First, Esteve et al. (2021) developed heuristic and backtracking algorithms to enhance the performance of EAT models in estimating production functions while reducing computational time. Later, Esteve et al. (2023) implemented a modified version of the Random Forest (RF) technique (Breiman, 2001), enabling the identification of super-efficiency. Finally, Guillen et al. (2023a,b) adapted the Gradient Tree Boosting method (Friedman, 2001) to the EAT framework, achieving substantial improvements in both mean squared error and bias when compared to the standard FDH methodology.

Valero-Carreras et al. (2021) adapted the Support Vector Regression (SVR) approach (Drucker et al., 1996) for production function estimation by combining DEA and FDH models within the structural risk minimization framework of SVR. Guerrero et al. (2022) introduced another SVR-based variant for efficiency estimation, aiming to mitigate the common issue of overfitting. Valero-Carreras et al. (2022) and Guerrero et al. (2024) further extended this line of research to the multi-output case. In a different direction, Moragues et al. (2023a,b) proposed a method combining DEA with One-Class Support Vector Machines (introduced by Schölkopf, 2001) to estimate technical efficiency in settings with multiple inputs and outputs, showing improvements over traditional DEA methods. However, the success of this approach largely depends on proper hyperparameter tuning, which may demand significant computational resources and expert knowledge.

Closely linked to the issue of overfitting, the selection of input and output variables remains one of the most critical methodological challenges in the DEA literature. Although DEA provides a flexible framework for evaluating efficiency, it offers no formal guidance on how to select the appropriate variables. Consequently, analysts often rely on subjective criteria—such as expert judgment, domain expertise, or data availability—which increases the risk of model misspecification. A substantial body of research has examined the impact of poor variable selection. For example, Sexton et al. (1986) and Smith (1997) showed that misspecified models can significantly alter efficiency estimates by reshaping the frontier. Dyson et al. (2001) noted that including irrelevant variables or omitting highly correlated ones may compromise result robustness, exposing the limits of correlation-based screening. From a statistical standpoint, Simar and Wilson (2001, 2013, 2015) showed that adding irrelevant variables can slow the convergence of DEA estimators as dimensionality increases, reducing the reliability of efficiency measures. In general, this problem is especially acute in high-dimensional contexts with limited samples, where irrelevant variables can distort the frontier, inflate efficiency scores, and weaken the model's ability to discriminate among units.

A number of studies have attempted to identify optimal ratios between model dimensionality and sample size to ensure reliable efficiency estimates. Specifically, a high ratio of variables to DMUs often results in overly optimistic evaluations, placing many units—incorrectly—on the efficient frontier. This issue, widely associated with the curse of dimensionality, limits the model's ability to distinguish between efficient and inefficient units. To address this, several "rules of thumb" have been proposed to relate sample size to model dimensionality. For instance, some authors suggest that the number of DMUs should be at least twice the number of input and output variables combined, while others recommend at least three times that total. An alternative rule indicates that the number of DMUs should be no less than either the product of the number of inputs and outputs or three times their sum—whichever is greater. Key contributions in this area include those by Nunamaker (1985), Golany and Roll (1989), Boussofiane et al. (1991), Friedman and Sinuany-Stern (1998), Homburg (2001), Dyson et al. (2001), Raab and Lichty (2002) and Cooper et al. (2007). However, as Wilson (2018) points out, these empirical rules, though widely cited, lack strong theoretical support.

To address issues related to the curse of dimensionality in DEA, various techniques have been developed to improve model specification and discriminatory power. These methods aim to select relevant variables while reducing dimensionality without compromising the accuracy of efficiency estimates. Broadly, most feature selection approaches can be categorized into two types: (i) methods that select a subset of the original variables and (ii) methods that rely on aggregate measures.

Aggregate measures are frequently used in dimensionality reduction by specifying a parametric aggregation function to combine inputs and outputs into indexes, replacing the original variables. These functions can be based on value judgments or statistical criteria, as proposed by Friedman and Sinuany-Stern (1998). Among these approaches, DEA-PCA—based on Principal Component Analysis (PCA) and developed independently by Ueda and Hoshiai (1997) and Adler and Golany (2001, 2002)—is particularly popular. This method reduces dimensionality by constructing uncorrelated linear combinations of inputs and outputs, effectively preserving most of the original information. Similarly, Daraio and Simar (2007) proposed an alternative eigensystem method that reduces dimensionality by focusing on principal components derived from moment matrices of inputs and outputs. Finally, Wilson (2018) emphasizes the benefits of dimensionality reduction in non-parametric models, particularly for improving the convergence rates of efficiency estimators. Through simulation experiments, Wilson suggests that reducing a multi-input, multi-output problem to a single-input, single-output framework by using the first principal components of the inputs and outputs can be advantageous in many scenarios.

On the other hand, feature selection methods in DEA are mainly rooted in statistical techniques including tests for marginal importance, regression-based methods, and stepwise procedures. These methods assess the contribution of each variable to the efficiency estimates. Examples include the marginal importance tests proposed by Banker (1996), the contribution efficiency measure introduced

by Pastor et al. (2002), and the bootstrap-based inference procedures developed by Simar and Wilson (2001). Regression-based approaches, such as the one proposed by Ruggiero (2005), identify relevant variables by regressing efficiency scores on candidate variables and analyzing the sign and significance of the resulting coefficients. Stepwise methods—like that of Wagner and Shimshak (2007)—sequentially add or remove variables to maximize or minimize the average change in efficiency scores. Another notable contribution is that of Jenkins and Anderson (2003), who proposed partial covariance analysis as a tool to identify variables that can be excluded without significantly affecting the model's explanatory power. While effective, these methods face two main limitations: their results are sensitive to the order in which variables are tested, and the computational burden increases rapidly as the number of inputs and outputs grows. Additionally, optimization-based frameworks have also been developed, such as those proposed by Charles et al. (2019), Benítez-Peña et al. (2020), and Peyrache et al. (2020).

A noteworthy methodology addressing feature selection and overfitting in DEA is the integration of Least Absolute Shrinkage and Selection Operator (LASSO) by Tibshirani (1996) and Elastic Net (EN) by Zou and Hastie (2005) into efficiency estimation models. Lee and Cai (2020) introduced the LASSO-CCNLS approach, embedding feature selection directly into the CCNLS framework developed by Kuosmanen and Johnson (2010). Building on this, Chen et al. (2021) extended this framework by incorporating EN and proposed a two-step method: first applying standard LASSO or EN for feature selection and, second, running DEA or CCNLS on the selected relevant variables. These approaches highlight the effective integration of ML techniques into DEA to mitigate the curse of dimensionality.

This thesis builds on recent advances in production frontier estimation by proposing a novel methodological framework that integrates the flexibility of ML with the foundational principles of production theory. We introduce two complementary approaches: one based on the Multivariate Adaptive Regression Splines (MARS) algorithm by Friedman (1991) and another that additionally uses the ensemble capabilities and robustness of RF by Breiman (2001). Both approaches accommodate shape constraints rooted in economic theory while maintaining the flexibility required to adapt to complex, real-world production environments. This enables the construction of DEA-type technologies that are both data-driven and theoretically grounded, allowing for meaningful efficiency measurement. Unlike traditional DEA models—highly sensitive to sample-specific characteristics and prone to overfitting in high-dimensional contexts—our framework mitigates these issues through a learning-based envelopment mechanism that avoids aligning the frontier too closely with observed data. It also incorporates an embedded feature selection procedure to retain only the most informative inputs, reducing dimensionality, enhancing interpretability, and improving the reliability of efficiency estimates. By combining non-parametric frontier modeling with ML techniques, the proposed framework offers a robust alternative to conventional deterministic and stochastic methods. The remainder of the thesis is structured as follows.

Chapter 2 details the objectives guiding the thesis, defining the central research questions and methodological aims. Chapter 3 introduces foundational concepts of DEA, MARS, and RF, presenting the simulation frameworks and evaluation metrics used to validate the proposed methodologies, alongside the computational infrastructure employed. Chapter 4 adapts MARS to estimate single-output production functions under monotonicity and concavity constraints, introduces modified forward and backward algorithms, and incorporates smoothing procedures using cubic and quintic basis functions. Chapter 5 presents Adaptive Constrained Enveloping Splines (ACES), extending the methodology to allow variable interactions and multi-output technologies, and includes extensive computational experiments to analyze performance. Chapter 6 introduces Random Forest-ACES (RF-ACES), enhancing robustness through ensemble learning and empirically validating the approach on real-world data from education and the food industry. Chapter 7 proposes Quick-ACES (Q-ACES), focusing on computational efficiency by implementing heuristic strategies and alternative automated variable selection methods, making frontier estimation scalable for larger datasets. Finally, Chapter 8 summarizes the main contributions, conclusions, and future research directions identified in the thesis.

Chapter 2

Objectives

Chapter 2 is devoted to defining the specific objectives that will guide the development of this thesis. The first chapter established the general framework of efficiency analysis, highlighting the relevance of Data Envelopment Analysis (DEA) as the predominant tool in the field, while also acknowledging its limitations. In this regard, the motivation behind this study lies in the search for effective solutions to the shortcomings identified in traditional production frontier estimation methodologies. In particular, the focus is on addressing two major challenges associated with DEA: the issue of overfitting—closely linked to its limited inferential capacity—and the difficulty of selecting the most relevant input and output variables.

In the context of production frontier estimation, DEA proves useful when the analysis is limited to a detailed description of the observed sample. However, the aim of this thesis is to move toward statistical generalization in efficiency analysis by embracing the principles of statistical learning theory. To this end, we propose the use of Multivariate Adaptive Regression Splines (MARS) and Random Forest (RF) for estimating production functions that are consistent with fundamental microeconomic principles, while ensuring that technical efficiency estimates can be extrapolated beyond the initial sample. These methods provide a flexible framework to capture complex relationships among variables, without compromising adherence to fundamental economic principles. This represents a shift from the purely descriptive nature of DEA toward a broader learning-based framework, seeking an optimal balance between precision on the (observed) training data and the ability to generalize to new observations without incurring significant errors. Accordingly, the proposed methodology formulates efficiency estimation as a regression problem, incorporating specific shape constraints into the predictor to maintain consistency with essential microeconomic axioms.

The first objective of this thesis is to develop a novel methodology—termed Adaptive Constrained Enveloping Splines (ACES)— for estimating production functions involving a single output. This approach is grounded in the adaptation of the MARS algorithm to the context of efficiency analysis. Specifically, it establishes a formal framework for incorporating the shape constraints required by production theory—monotonicity and concavity—into the MARS estimator. These constraints are imposed in a way that ensures coherence with microeconomic principles while retaining the flexibility of the non-parametric model. The resulting framework lays the foundation for using MARS as a reliable tool in the estimation of production frontiers.

Building on the promising performance of this initial method, the second objective is to expand the ACES framework through two complementary methodological extensions. First, we propose a general approach for estimating production technologies in multi-output settings, enabling the construction of enveloping representations that accommodate multiple outputs while preserving theoretical consistency. Second, we introduce a strategy to incorporate variable interactions into the estimation process, which allows the model to capture non-additive production surfaces and complex input relationships. To further improve robustness, we adapt an ensemble-based approach using RF, giving rise to a variant called Random Forest-Adaptive Constrained Enveloping Splines (RF-ACES). Although this refinement is formally developed in the context of the additive model, it remains fully applicable to the interaction-based framework, offering a more robust alternative in the presence of data noise or irregular patterns.

A third objective of this thesis is to address the challenge of variable selection in production frontier estimation. Both the standard MARS-based method and its RF variant offer mechanisms to assess the relative importance of each input in determining output levels. These insights make it possible to identify and retain only the most relevant inputs, effectively reducing dimensionality and improving model interpretability. While the RF approach provides a more robust assessment due to its ensemble nature, it also entails a higher computational cost. Building on these capabilities, we propose a methodology that estimates efficiency scores using only the most relevant variables, thereby excluding redundant or irrelevant inputs from the analysis. This contributes to more reliable efficiency measurement and helps mitigate the adverse effects of overfitting in high-dimensional settings.

The final objective of this thesis is to introduce a computationally efficient variant of the proposed methodology, referred to as Quick-Adaptive Constrained Enveloping Splines (Q-ACES). This version is designed to significantly reduce execution times while preserving the main structural features and estimation quality of the original model. Q-ACES incorporates heuristic strategies that prioritize variable selection and streamline the estimation process without compromising the underlying theoretical framework. As such, it provides a practical alternative for large-scale applications or contexts requiring rapid frontier estimation, making the methodology more accessible and scalable in real-world settings.

Chapter 3

Materials and methods

Chapter 3 provides a detailed introduction to the fundamental principles of Data Envelopment Analysis (DEA), Multivariate Adaptive Regression Splines (MARS), and Random Forests (RF), which serve as the core methodological pillars of the proposed approaches: Adaptive Constrained Enveloping Splines (ACES), Random Forest–ACES (RF-ACES), and Quick-ACES (Q-ACES). The chapter is organized into four sections: the first three offer a detailed overview of DEA, MARS and RF, while the fourth presents the experimental framework designed to assess the performance of the proposed methodologies. This includes the specification of simulation settings, the description of hardware and software used, and the definition of evaluation metrics. Additionally, the chapter introduces the notation adopted throughout the thesis to ensure conceptual clarity and internal consistency.

3.1 Data Envelopment Analysis

DEA has emerged as a widely accepted alternative to traditional parametric methods for estimating production frontiers and evaluating efficiency. Unlike parametric models, which require specifying a functional form—such as Cobb-Douglas or Translog—and fitting it to the data using statistical techniques, DEA employs a non-parametric approach based on mathematical programming. It constructs a piecewise linear frontier by solving a linear program for each Decision Making Unit (DMU) in the dataset. This frontier is defined by the Pareto-efficient DMUs—those for which no other unit can produce more output with the same or fewer inputs, or the same output with fewer inputs. These efficient units serve as benchmarks to assess the performance of the remaining ones, which are projected onto the frontier to determine their relative inefficiency.

Until the publication of the study by Aigner and Chu (1968), the prevailing approach among economists for estimating efficiency relied on analyzing average behaviors through traditional regression models, which assumed stochastic errors with zero mean. These models were designed to capture the central tendency of the data, rather than the boundary of production possibilities. However, in their work, Aigner and Chu introduced a methodology for estimating a parametric production function, such as Cobb-Douglas, using mathematical programming techniques that imposed positive residuals to ensure the estimated frontier enveloped the data, directly modeling the production boundary. In contrast, DEA is non-parametric, making no prior assumptions about the functional form of the input-output relationship or the distribution of errors, offering a flexible framework for efficiency estimation without the restrictions of parametric models.

Both the models introduced by Aigner and Chu (1968) and DEA are considered deterministic frontiers since any deviation of DMUs from the estimated production function is interpreted as a source of inefficiency. To address this limitation, Aigner et al. (1977) and Meeusen and van Den Broeck (1977) made pioneering contributions by distinguishing between random noise and technical inefficiency in the deviations of DMUs from the frontier, laying the groundwork for the development of the methodology known as Stochastic Frontier Analysis (SFA).

Figure 3.1 presents a comparison between DEA, SFA, and linear regression in estimating the production function $f(x) = \sqrt{x}$ for a set of 30 DMUs that use a single input to produce an output:

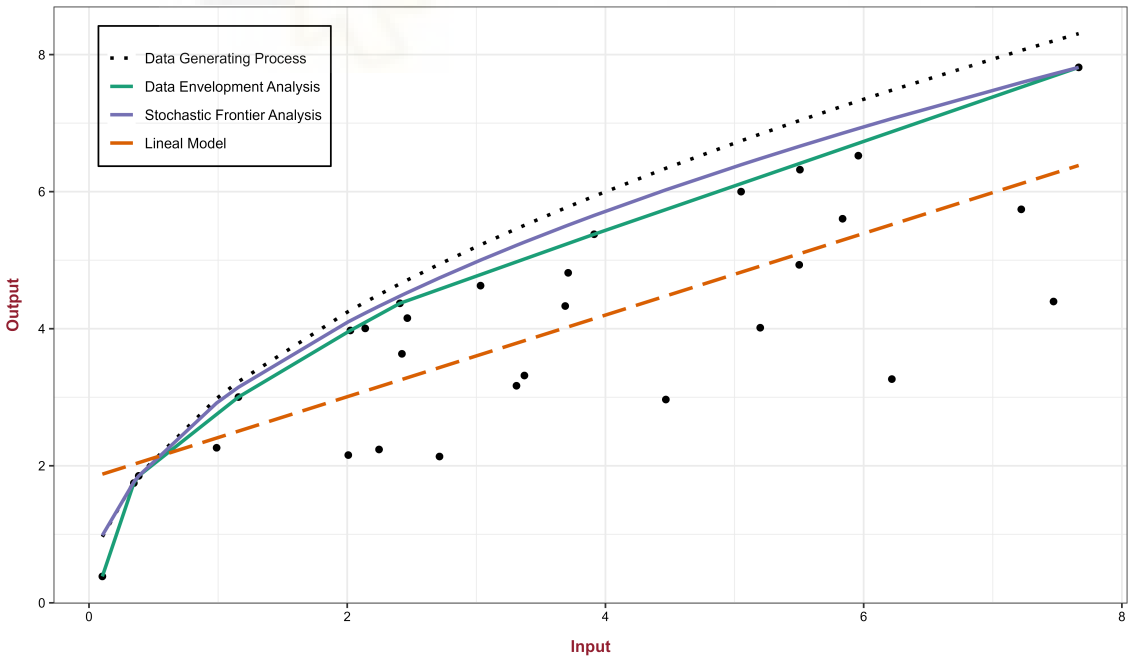

Figure 3.1: Comparison of production function estimates by DEA, SFA and linear regression.

First, it is worth noting that linear regression, which focuses on explaining average behaviors rather than capturing the extreme observations in the data, is inadequate for estimating production functions. In contrast, both DEA and SFA provide an enveloping approach that captures the set of evaluated DMUs. However, SFA is distinguished by allowing—although this situation is not observed in Figure 3.1—the presence of DMUs above the estimated frontier. Finally, DEA is characterized by identifying a set of reference units that support the frontier, a feature absent in SFA, which highlights a key difference in how the frontier is supported in both methods.

From this point on, we aim to evaluate the efficiency of a set of n DMUs. Specifically, each DMU_i, $i = 1, \ldots, n$, consumes $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ij}, \ldots, x_{im}) \in \mathbb{R}^m_+$ inputs to produce $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{ir}, \ldots, y_{is}) \in \mathbb{R}^s_+$ outputs. We denote the unit under evaluation as DMU_0, which can be described by the pair $(\boldsymbol{x}_0, \boldsymbol{y}_0)$. The following subsections describe basic DEA models introduced in the literature.

3.1.1 The CCR model (Charnes, Cooper and Rhodes)

Charnes et al. (1978) introduced a new conception of efficiency based on the ratio between multiple outputs and inputs, thus extending the traditional definition of efficiency, which is the ratio of a single output to a single input. Consequently, the following Fractional Programming problem, known as the input-oriented CCR model in its ratio form, provides the relative efficiency measure for the DMU_0:

$$\max_{\boldsymbol{\mu},\boldsymbol{\nu}} \quad \frac{\sum_{r=1}^{s} \mu_r y_{0r}}{\sum_{j=1}^{m} \nu_j x_{0j}} \tag{3.1}$$

subject to:

$$\begin{aligned} \frac{\sum_{r=1}^{s} \mu_r y_{ir}}{\sum_{j=1}^{m} \nu_j x_{ij}} &\leq 1, \quad i = 1, \ldots, n, \\ \mu_r &\geq 0, \quad r = 1, \ldots, s, \\ \nu_j &\geq 0, \quad j = 1, \ldots, m. \end{aligned}$$

The goal of Model (3.1) is to find a vector of non-negative weights for inputs and outputs to maximize the efficiency of DMU_0. However, the model is nonlinear, which makes it complex to solve. To address this, Charnes and Cooper (1963) proposed adding the constraint $\sum_{j=1}^{m} \nu_j x_{0j} = 1$, converting the initial Fractional Programming problem into an equivalent Linear Programming (LP) problem:

$$\max_{\boldsymbol{\mu},\boldsymbol{\nu}} \sum_{r=1}^{s} \mu_r y_{0r} \tag{3.2}$$

subject to:

$$\begin{aligned} \sum_{j=1}^{m} \nu_j x_{0j} &= 1, \\ -\sum_{j=1}^{m} \nu_j x_{ij} + \sum_{r=1}^{s} \mu_r y_{ir} &\leq 0, \quad i = 1, \ldots, n, \\ \mu_r &\geq 0, \quad r = 1, \ldots, s, \\ \nu_j &\geq 0, \quad j = 1, \ldots, m. \end{aligned}$$

Model (3.2) is known as the input-oriented CCR model in its multiplier formulation. The dual of this problem is the input-oriented CCR model in its envelopment formulation:

$$\min_{\theta,\boldsymbol{\lambda}} \ \theta \tag{3.3}$$

subject to:

$$\begin{aligned} \sum_{i=1}^{n} \lambda_i x_{ij} &\leq \theta x_{0j}, \quad j = 1, \ldots, m, \\ \sum_{i=1}^{n} \lambda_i y_{ir} &\geq y_{0r}, \quad r = 1, \ldots, s, \\ \lambda_i &\geq 0, \quad i = 1, \ldots, n. \end{aligned}$$

The optimal solution of Model (3.3) is given by $(\theta^*, \lambda_1^*, ..., \lambda_n^*)$, where it is easy to prove that the efficiency score lies within the interval $0 < \theta^* \leq 1$. Essentially, solving Model (3.3) implies performing the maximum equiproportional reduction in all inputs of DMU_0 using the coefficient θ^* until reaching a point in the efficient frontier, represented by $(\theta^* \boldsymbol{x}_0, \boldsymbol{y}_0)$.

Additionally, solving Model (3.3) identifies a set of benchmarks (referents) and targets for DMU_0. The referents, corresponding to DMUs with $\lambda_i^* > 0$, act as real models that the evaluated unit should emulate to optimize its performance if found inefficient. In contrast, the targets are the projection coordinates onto the efficient frontier, given by

$$\left(\sum_{i=1}^{n} \lambda_i^* x_{i1}, \cdots, \sum_{i=1}^{n} \lambda_i^* x_{im}, \sum_{i=1}^{n} \lambda_i^* y_{i1}, \cdots, \sum_{i=1}^{n} \lambda_i^* y_{is} \right), \tag{3.4}$$

indicating virtual input and output levels (not necessarily observed) that would make DMU_0 efficient.

Models (3.1), (3.2), and (3.3) can be analogously defined under output orientation. In this case, the relative efficiency measured by the CCR model arises from the maximum equiproportional expansion of the outputs of DMU_0, subject to its observed input level.

Definition 1 [Farrell-Debreu Efficiency]: DMU_0 is efficient if, and only if, $\theta^* = 1$.

By solving Model (3.3), DMU_0 might be deemed efficient according to the above definition yet may still have potential improvement in some inputs or outputs without adversely affecting other evaluated variables. Therefore, slack variables associated with "excesses" in input use ($\boldsymbol{s}_0^- \in \mathbb{R}_+^m$) and "shortfalls" in output attainment ($\boldsymbol{s}_0^+ \in \mathbb{R}_+^s$) are defined as follows:

$$s_{0j}^- = \theta x_{0j} - \sum_{i=1}^{n} \lambda_i x_{ij}, \quad j = 1, \ldots, m, \tag{3.5}$$

$$s_{0r}^+ = \sum_{i=1}^{n} \lambda_i y_{ir} - y_{0r}, \qquad r = 1, \ldots, s. \tag{3.6}$$

Typically, DEA determines efficiency scores and slacks through a two-phase procedure. The first phase solves Model (3.3), where θ^* coincides with Farrell-Debreu-type efficiency. Subsequently, a second phase incorporates θ^* into the following LP:

$$\max_{\boldsymbol{\lambda}, \boldsymbol{s}_0^-, \boldsymbol{s}_0^+} \sum_{j=1}^{m} s_{0j}^- + \sum_{r=1}^{s} s_{0r}^+ \tag{3.7}$$

subject to:

$$\begin{aligned}
\sum_{i=1}^{n} \lambda_i x_{ij} &= \theta^* x_{0j} - s_{0j}^-, \quad && j = 1, \ldots, m, \\
\sum_{i=1}^{n} \lambda_i y_{ir} &= y_{0r} + s_{0r}^+, && r = 1, \ldots, s, \\
\lambda_i &\geq 0, && i = 1, \ldots, n, \\
s_{0j}^- &\geq 0, && j = 1, \ldots, m, \\
s_{0r}^+ &\geq 0, && r = 1, \ldots, s.
\end{aligned}$$

The objective of Model (3.7) is to find a solution maximizing the sum of input and output slacks, i.e., maximizing the ℓ_1^{m+s} distance of point $(\theta^* \boldsymbol{x}_0, \boldsymbol{y}_0)$ to the efficient frontier.

Definition 2 [CCR Efficiency]: If an optimal solution $\left(\theta^*, \lambda_1^*, \ldots, \lambda_n^*, s_{01}^{-*}, \ldots, s_{0m}^{-*}, s_{01}^{+*}, \ldots, s_{0s}^{+*}\right)$ of Models (3.3) and (3.7) satisfies $\theta^* = 1$ (technical or radial efficiency), $\boldsymbol{s}_0^- = \boldsymbol{0}_m$ and $\boldsymbol{s}_0^+ = \boldsymbol{0}_s$, then the DMU_0 is said to be CCR-efficient.

If the evaluated DMU satisfies only the first condition, it is classified as "weakly efficient", whereas if it satisfies both conditions, it is classified as "strongly efficient".

3.1.2 The BCC model (Banker, Charnes and Cooper)

Nearly a decade later, Banker et al. (1984) proposed an axiomatic approach to efficiency evaluation, based on the characterization of the production technology or possibility set through fundamental axioms that define the essential properties of such a set:

$$\varphi = \left\{(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}_+^{m+s} : \boldsymbol{x} \text{ can produce } \boldsymbol{y}\right\}. \tag{3.8}$$

With respect to the set φ, Banker et al. (1984) assume the following postulates:

Postulate 0 [Determinism]: The observed entities $(\boldsymbol{x}_i, \boldsymbol{y}_i)$, $i = 1, \ldots, n$, belong to the technology φ.

Postulate 1 [Convexity]: If $(\boldsymbol{x}, \boldsymbol{y}) \in \varphi$ and $(\boldsymbol{x}', \boldsymbol{y}') \in \varphi$, then $\lambda(\boldsymbol{x}, \boldsymbol{y}) + (1 - \lambda)(\boldsymbol{x}', \boldsymbol{y}') \in \varphi$ for all $\lambda \in [0, 1]$.

Postulate 2 [Free disposability]: If $(\boldsymbol{x}, \boldsymbol{y}) \in \varphi$, then $(\boldsymbol{x}', \boldsymbol{y}') \in \varphi$ whenever $\boldsymbol{x}' \geq \boldsymbol{x}$ and $\boldsymbol{y}' \leq \boldsymbol{y}$ (component-wise).

Postulate 3 [Constant returns to scale]: If $(\boldsymbol{x}, \boldsymbol{y}) \in \varphi$, then $(a \cdot \boldsymbol{x}, a \cdot \boldsymbol{y}) \in \varphi$ for any $a \geq 0$.

Postulate 4 [Minimal extrapolation]: φ is the smallest subset of $\mathbb{R}_+^{m+s}$ satisfying the above postulates.

In this way, a technology that assumes constant returns to scale (CRS) is denoted as:

$$\hat{\varphi}_{\text{CRS}} = \left\{ (\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}_+^{m+s} \;\middle|\; \begin{array}{ll} x_j \geq \sum_{i=1}^{n} \lambda_i x_{ij}, & j = 1, \ldots, m, \\ y_r \leq \sum_{i=1}^{n} \lambda_i y_{ir}, & r = 1, \ldots, s, \\ \lambda_i \geq 0, & i = 1, \ldots, n \end{array} \right\}. \tag{3.9}$$

whereas if the technology assumes variable returns to scale (VRS), it is denoted as:

$$\hat{\varphi}_{\text{VRS}} = \left\{ (\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}_+^{m+s} \;\middle|\; \begin{array}{ll} x_j \geq \sum_{i=1}^{n} \lambda_i x_{ij}, & j = 1, \ldots, m, \\ y_r \leq \sum_{i=1}^{n} \lambda_i y_{ir}, & r = 1, \ldots, s, \\ \sum_{i=1}^{n} \lambda_i = 1, & \\ \lambda_i \geq 0, & i = 1, \ldots, n \end{array} \right\}. \tag{3.10}$$

Additionally, we define the set of points that constitute the efficient frontier in the technology as:

$$\partial(\varphi) = \{(\boldsymbol{x}, \boldsymbol{y}) \in \varphi : \hat{\boldsymbol{x}} < \boldsymbol{x},\ \hat{\boldsymbol{y}} > \boldsymbol{y} \Rightarrow (\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}) \notin \varphi\}\,. \tag{3.11}$$

Accordingly, if the goal is to determine the efficiency level of a given DMU_0 under input orientation and VRS (omitting the third postulate), the corresponding LP model is formulated as follows:

$$\min_{\theta,\,\boldsymbol{\lambda}}\ \theta \tag{3.12}$$

subject to:

$$\begin{aligned}
\sum_{i=1}^{n} \lambda_i x_{ij} &\leq \theta x_{0j}, \quad j = 1, \ldots, m,\\
\sum_{i=1}^{n} \lambda_i y_{ir} &\geq y_{0r}, \quad r = 1, \ldots, s,\\
\sum_{i=1}^{n} \lambda_i &= 1,\\
\lambda_i &\geq 0, \quad i = 1, \ldots, n.
\end{aligned}$$

Model (3.12) is referred to as the BCC model under input orientation in its envelopment form. The interpretation of its results is equivalent to that of model (3.3), providing both an efficiency score ($0 < \theta^* \leq 1$) and a set of reference units and targets (derived from variables $\boldsymbol{\lambda}$). Similarly, this model also allows the efficiency assessment to be performed through a two-stage process, using Definitions 1 and 2 to distinguish between weakly and strongly efficient units in the sample.

We now establish the BCC model under output orientation in its envelopment formulation:

$$\max_{\phi,\,\boldsymbol{\lambda}}\ \phi \tag{3.13}$$

subject to:

$$\begin{aligned}
\sum_{i=1}^{n} \lambda_i x_{ij} &\leq x_{0j}, \quad j = 1, \ldots, m,\\
\sum_{i=1}^{n} \lambda_i y_{ir} &\geq \phi y_{0r}, \quad r = 1, \ldots, s,\\
\sum_{i=1}^{n} \lambda_i &= 1,\\
\lambda_i &\geq 0, \quad i = 1, \ldots, n.
\end{aligned}$$

Here, the efficiency score $\phi^* \in [1, \infty)$ represents the maximum equiproportional expansion of the outputs required by DMU_0 to become efficient while keeping its input level constant. Consequently, the projection of DMU_0 onto the efficient frontier is given by the point $(\boldsymbol{x}_0, \phi^* \boldsymbol{y}_0)$.

Finally, Figure 3.2 presents a comparison between the efficient frontiers under CRS and VRS technologies, highlighting the fundamental differences in their structure.

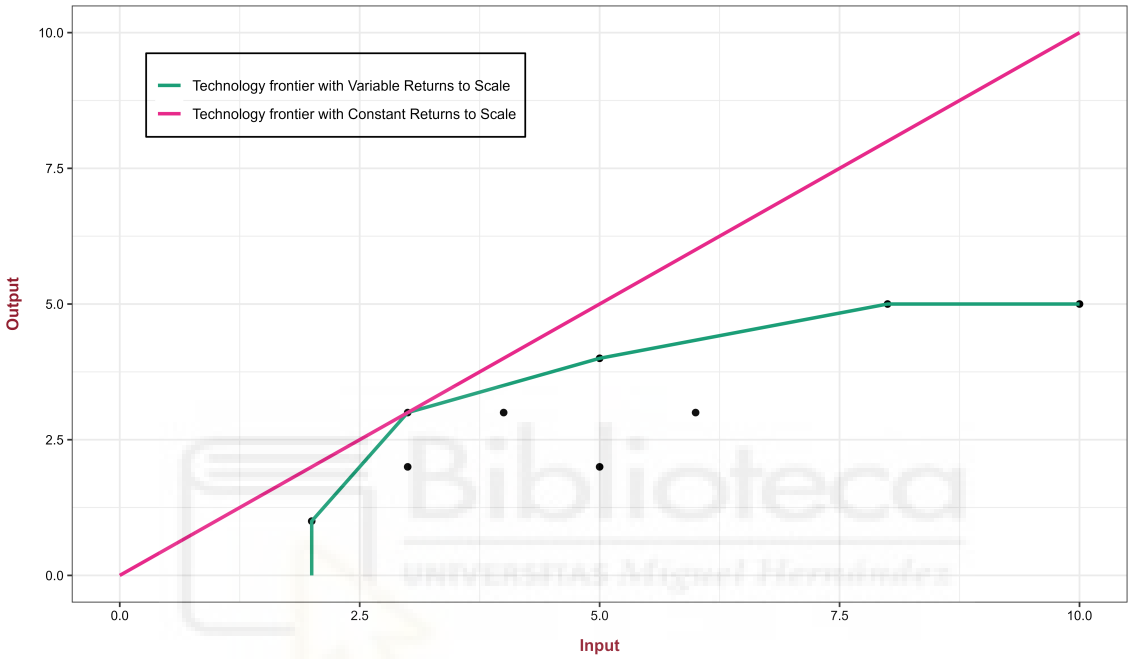

Figure 3.2: Comparison of DEA efficient frontiers under CRS and VRS assumptions.

Regarding Figure 3.2, it is important to note that the VRS technology is always contained within the CRS technology and, therefore, always yields more optimistic efficiency assessments. On the other hand, while the assumption of CRS simplifies the analysis, it often fails to accurately reflect the real conditions under which firms and organizations operate.

3.1.3 Other efficiency estimation models

The CCR and BCC models discussed above produce efficiency evaluations that are both oriented and radial. Specifically, these models focus on adjustments of either inputs or outputs (orientation-specific) and ensure that projected points on the frontier preserve the original input–output mix. In this subsection, we introduce alternative efficiency measures that are neither oriented nor radial. Two notable approaches in this context are the additive model proposed by Charnes et al. (1985) and the Directional Distance Function (DDF) introduced by Chambers et al. (1996a).

The additive model in its envelopment form under VRS technology is:

$$\max_{\boldsymbol{\lambda},\, \boldsymbol{s}_0^-,\, \boldsymbol{s}_0^+} \sum_{j=1}^{m} s_{0j}^- + \sum_{r=1}^{s} s_{0r}^+ \tag{3.14}$$

subject to:

$$\begin{aligned}
\sum_{i=1}^{n} \lambda_i x_{ij} &= x_{0j} - s_{0j}^-, \quad j = 1, \ldots, m, \\
\sum_{i=1}^{n} \lambda_i y_{ir} &= y_{0r} + s_{0r}^+, \quad r = 1, \ldots, s, \\
\sum_{i=1}^{n} \lambda_i &= 1, \\
\lambda_i &\geq 0, \qquad i = 1, \ldots, n, \\
s_{0j}^- &\geq 0, \qquad j = 1, \ldots, m, \\
s_{0r}^+ &\geq 0, \qquad r = 1, \ldots, s.
\end{aligned}$$

Upon solving model (3.14), the ℓ_1^{m+s} distance between DMU_0 and the efficient frontier is maximized. The evaluated DMU is deemed efficient if, and only if, the optimal value of (3.14) is zero—implying that no input or output can be improved without worsening another. Moreover, this measure also quantifies both the areas needing improvement and the degree of inefficiency in each input and output. However, the efficiency score is affected by the measurement units of the variables. To address this limitation, weighted additive approaches have been proposed, leading to the following LP model:

$$\max_{\boldsymbol{\lambda},\, \boldsymbol{s}_0^-,\, \boldsymbol{s}_0^+} \sum_{j=1}^{m} w_j^- \cdot s_{0j}^- + \sum_{r=1}^{s} w_r^+ \cdot s_{0r}^+ \tag{3.15}$$

subject to:

$$\begin{aligned}
\sum_{i=1}^{n} \lambda_i x_{ij} &= x_{0j} - s_{0j}^-, \quad j = 1, \ldots, m, \\
\sum_{i=1}^{n} \lambda_i y_{ir} &= y_{0r} + s_{0r}^+, \quad r = 1, \ldots, s, \\
\sum_{i=1}^{n} \lambda_i &= 1, \\
\lambda_i &\geq 0, \qquad i = 1, \ldots, n, \\
s_{0j}^- &\geq 0, \qquad j = 1, \ldots, m, \\
s_{0r}^+ &\geq 0, \qquad r = 1, \ldots, s.
\end{aligned}$$

The selection of slack-variable weights (w) gives rise to different additive models. The Measure of Inefficiency Proportions (MIP) proposed by Cooper et al. (1999) sets the weights proportional to the input and output values of the evaluated DMU, defining $w_j^- = 1/x_{0j}$ and $w_r^+ = 1/y_{0r}$. The Normalized Weighted Additive Model (WAM) by Lovell and Pastor (1995) instead normalizes the slacks using the standard deviation of each variable, assigning $w_j^- = 1/\sigma(x_j)$ and $w_r^+ = 1/\sigma(y_r)$. Alternatively, the Range Adjusted Measure (RAM), also proposed by Cooper et al. (1999), employs the observed ranges of input and output variables across all DMUs for normalization. Its slack weights are defined as $w_j^- = 1/\big[(m+s)(\max_{1\le i\le n}(x_{ij}) - \min_{1\le i\le n}(x_{ij}))\big]$ and $w_r^+ = 1/\big[(m+s)(\max_{1\le i\le n}(y_{ir}) - \min_{1\le i\le n}(y_{ir}))\big]$. Finally, the Bounded Adjusted Measure (BAM) introduced by Cooper et al. (2011) selects weights based on the evaluated DMU's proximity to boundary values (maximum outputs and minimum inputs), defined as $w_j^- = 1/\big[(m+s)(x_{0j} - \min_{1\le i\le n}(x_{ij}))\big]$ and $w_r^+ = 1/\big[(m+s)(\max_{1\le i\le n}(y_{ir}) - y_{0r})\big]$.

On the other hand, the DDF is defined as follows:

$$\max_{\omega,\,\boldsymbol{\lambda}} \ \omega \tag{3.16}$$

subject to:

$$\begin{aligned}
\sum_{i=1}^{n} \lambda_i x_{ij} &\le x_{0j} + \omega g_{x_j}, \quad j = 1,\dots,m,\\
\sum_{i=1}^{n} \lambda_i y_{ir} &\ge y_{0r} + \omega g_{y_r}, \quad r = 1,\dots,s,\\
\omega &\ge 0,\\
\lambda_i &\ge 0, \quad i = 1,\dots,n.
\end{aligned}$$

Here, the vector $\boldsymbol{g} = (\boldsymbol{g}_x, \boldsymbol{g}_y) = (g_{x_1},\dots,g_{x_m}, g_{y_1},\dots,g_{y_s})$ defines the direction in which the evaluated DMU_0 is projected onto the frontier. If $\omega = 0$, the DMU is considered efficient, since no directional improvement is possible. Common choices for the direction vector include: $(\boldsymbol{g}_x, \boldsymbol{g}_y) = (-\boldsymbol{x}_0, \boldsymbol{0}_s)$ for an input-oriented measurement; $(\boldsymbol{0}_m, \boldsymbol{y}_0)$ for output orientation; and $(-\boldsymbol{x}_0, \boldsymbol{y}_0)$ for the formulation introduced by Briec (1997). When measuring efficiency under a VRS technology, the additional constraint $\sum_{i=1}^{n} \lambda_i = 1$ must be imposed.

Other notable efficiency measures—non-radial in nature—were introduced by Färe and Lovell (1978) and Färe et al. (1985), and are available in both oriented and non-oriented forms. In its oriented version, this measure is known as the Russell Graph Measure. Later, Pastor et al. (1999) and, independently, Tone (2001) refined this model, giving rise to the Enhanced Russell Graph Measure and the Slack-Based Measure, respectively. For further details on these measures, the aforementioned references are recommended.

Finally, Figure 3.3 provides a visual comparison of the projections obtained for a DMU, as computed by various efficiency measures discussed in this section (under a VRS technology). This figure offers a clear view of how these methods behave in practice.

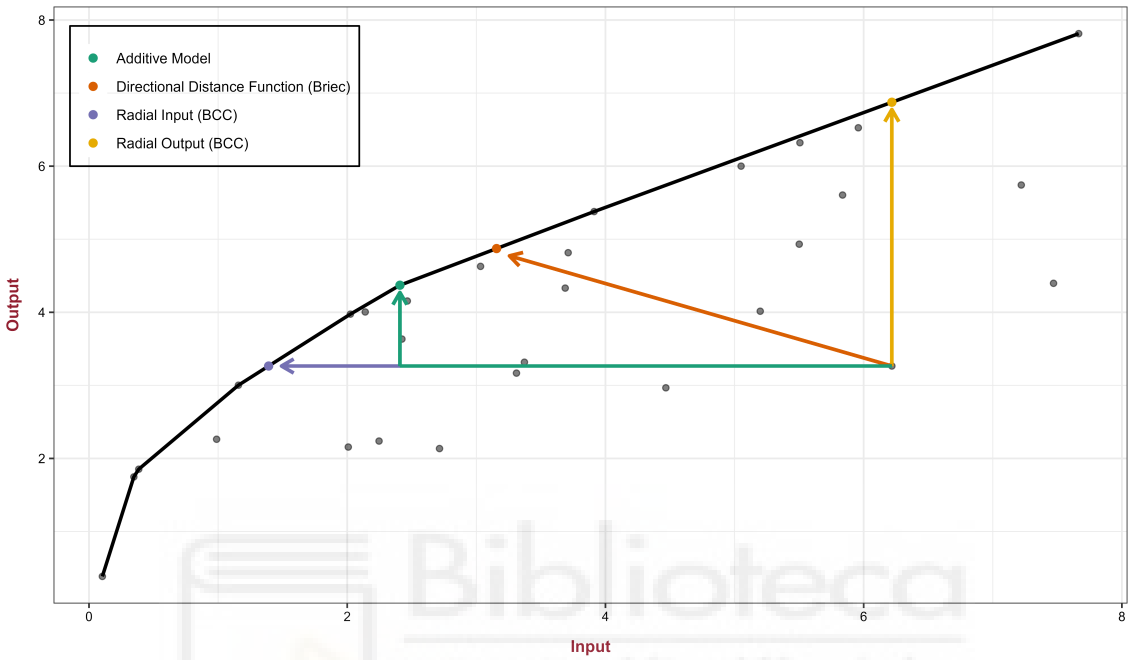

Figure 3.3: Comparative of different projections to the frontier.

As can be seen, radial measures are oriented, since they operate exclusively on either inputs or outputs. In contrast, the DDF and the additive model are non-oriented, as they simultaneously account for potential adjustments in both inputs and outputs.

3.1.4 Productivity indices

This section presents two widely used non-parametric indices for measuring productivity change over time: the Malmquist Productivity Index (MPI) and the Luenberger Productivity Indicator (LPI). Both rely on distance functions and are particularly suitable when reliable input and output price information is unavailable, as often occurs in public services or regulated sectors. The Malmquist index, introduced by Malmquist (1953) and later adapted by Caves et al. (1982) and Färe et al. (1994), assesses productivity through radial expansions or contractions relative to a reference technology. In contrast, the Luenberger indicator, developed by Chambers et al. (1996b) and Chambers and Pope (1996), uses a DDF (Chambers et al. 1996a) to capture non-proportional input and output adjustments. The following subsections present the formal definitions of both indices.

The Malmquist Productivity Index

The MPI is a widely used non-parametric tool for assessing total factor productivity (TFP) change over time. It was originally developed by Malmquist (1953) and later extended to the production context by Caves et al. (1982) and Färe et al. (1994). The index uses output distance functions, as originally formalized by Shephard (1953), to evaluate the performance of a single DMU across two time periods (t_1 and t_2).

Let the output-oriented Shephard distance function at time t be defined as:

$$D^t(\boldsymbol{x}_0, \boldsymbol{y}_0) = \inf \left\{ \phi > 0 : \left(\boldsymbol{x}_0, \frac{\boldsymbol{y}_0}{\phi} \right) \in \varphi^t \right\}, \tag{3.17}$$

where $\boldsymbol{x}_0 \in \mathbb{R}_+^m$ and $\boldsymbol{y}_0 \in \mathbb{R}_+^s$ represent the input and output vectors of the DMU_0 under evaluation, and φ^t denotes the production technology set in period t, defined by (3.9) under CRS (D^t_{CRS}) or (3.10) under VRS (D^t_{VRS}). This function captures the maximum proportional expansion of outputs that can be achieved with the observed inputs, while remaining within the feasible production set at time t. A value of $D^t(\boldsymbol{x}_0, \boldsymbol{y}_0) \leq 1$ indicates technical inefficiency, whereas $D^t(\boldsymbol{x}_0, \boldsymbol{y}_0) = 1$ signifies that the DMU is operating on the efficient frontier. Importantly, the reciprocal of this distance function, $1/D^t(\boldsymbol{x}_0, \boldsymbol{y}_0)$, coincides with the optimal value ϕ^* obtained from the output-oriented radial DEA model defined in Equation (3.13) under VRS. Under CRS, the same model applies but without the convexity constraint $\sum_{i=1}^n \lambda_i = 1$. This establishes a direct link between Shephard's distance and the standard DEA formulations.

Based on this distance function, the output-oriented MPI under CRS between two time periods t_1 and t_2 is defined as:

$$M^0_{t_1,t_2} = \sqrt{\frac{D^{t_1}_{\text{CRS}}(\boldsymbol{x}_0^{t_2}, \boldsymbol{y}_0^{t_2})}{D^{t_1}_{\text{CRS}}(\boldsymbol{x}_0^{t_1}, \boldsymbol{y}_0^{t_1})} \cdot \frac{D^{t_2}_{\text{CRS}}(\boldsymbol{x}_0^{t_2}, \boldsymbol{y}_0^{t_2})}{D^{t_2}_{\text{CRS}}(\boldsymbol{x}_0^{t_1}, \boldsymbol{y}_0^{t_1})}}, \tag{3.18}$$

In Equation (3.18), each distance function $D^t_{\text{CRS}}(\cdot)$ measures the maximal proportional expansion of the output vector relative to the CRS technology (φ^t_{CRS}) defined in Equation (3.9) for the corresponding period t. The MPI requires the computation of four Shephard output distance functions—or equivalently, the inverse of the optimal values obtained from output-oriented radial DEA models—each corresponding to a specific combination of data and technology:

- $D^{t_1}_{\text{CRS}}(\boldsymbol{x}_0^{t_1}, \boldsymbol{y}_0^{t_1})$ evaluates DMU_0 using its own period t_1 data with respect to the period t_1 technology (i.e., its technical efficiency at time t_1).

- $D^{t_2}_{\text{CRS}}(\boldsymbol{x}_0^{t_2}, \boldsymbol{y}_0^{t_2})$ evaluates DMU_0 using its own period t_2 data with respect to the period t_2 technology (i.e., its technical efficiency at time t_2).

- $D^{t_1}_{\text{CRS}}(\boldsymbol{x}_0^{t_2}, \boldsymbol{y}_0^{t_2})$ evaluates the period t_2 performance of DMU_0 with respect to the earlier technology from period t_1.

- $D^{t_2}_{\text{CRS}}(\boldsymbol{x}_0^{t_1}, \boldsymbol{y}_0^{t_1})$ evaluates the period t_1 performance of DMU_0 with respect to the later technology from period t_2.

The first two terms capture the DMU's efficiency relative to its contemporaneous frontiers, while the last two reflect how its performance would be assessed under the alternative period's technology. This setup enables the decomposition of productivity change into two main components: efficiency change (i.e., movement towards or away from the frontier) and technological change (i.e., shifts in the frontier itself). In this way, values of $M^0_{t_1,t_2} > 1$ indicate productivity improvement, whereas values below one suggest productivity deterioration.

In most practical situations, assuming CRS or VRS when computing distance functions leads to different Malmquist index values. Nonetheless, various authors have noted that the traditional formulation of the Malmquist index does not explicitly account for the type of returns to scale (RTS) that best represents the underlying production technology. For example, Grifell-Tatjé and Lovell (1995) present a two-dimensional case showing that, under VRS, the index may fail to accurately capture true productivity changes. Along similar lines, Ray and Desli (1997) argue that the MPI is correctly measured by the ratio of CRS distance functions even when the technology exhibits VRS. Conversely, Balk (2001) contends that when the index is built from the definitions of its individual components, the resulting measure naturally aligns with one computed under CRS assumptions, regardless of the actual RTS in each period. This position is further reinforced by Lovell (2003), who supports using CRS-based distance functions in the construction of the index. Therefore, in line with this stream of the literature, we adopt the CRS assumption throughout when computing the distance functions in Equation (3.18).

The MPI can be decomposed into two fundamental components (Färe et al., 1994) when maintaining a CRS assumption. This classical decomposition is expressed as:

$$M^0_{t_1,t_2} = \sqrt{\frac{D^{t_1}_{\text{CRS}}(\boldsymbol{x}_0^{t_2}, \boldsymbol{y}_0^{t_2})}{D^{t_1}_{\text{CRS}}(\boldsymbol{x}_0^{t_1}, \boldsymbol{y}_0^{t_1})} \cdot \frac{D^{t_2}_{\text{CRS}}(\boldsymbol{x}_0^{t_2}, \boldsymbol{y}_0^{t_2})}{D^{t_2}_{\text{CRS}}(\boldsymbol{x}_0^{t_1}, \boldsymbol{y}_0^{t_1})}} = \text{EC} \cdot \text{TC}, \tag{3.19}$$

where the two components are defined as:

$$\text{EC} = \frac{D^{t_2}_{\text{CRS}}(\boldsymbol{x}_0^{t_2}, \boldsymbol{y}_0^{t_2})}{D^{t_1}_{\text{CRS}}(\boldsymbol{x}_0^{t_1}, \boldsymbol{y}_0^{t_1})}, \tag{3.20}$$

$$\text{TC} = \sqrt{\frac{D^{t_1}_{\text{CRS}}(\boldsymbol{x}_0^{t_2}, \boldsymbol{y}_0^{t_2})}{D^{t_2}_{\text{CRS}}(\boldsymbol{x}_0^{t_2}, \boldsymbol{y}_0^{t_2})} \cdot \frac{D^{t_1}_{\text{CRS}}(\boldsymbol{x}_0^{t_1}, \boldsymbol{y}_0^{t_1})}{D^{t_2}_{\text{CRS}}(\boldsymbol{x}_0^{t_1}, \boldsymbol{y}_0^{t_1})}}. \tag{3.21}$$

This decomposition allows the TFP change to be interpreted in terms of changes in technical efficiency (EC) and shifts in the production frontier (TC), both evaluated under a consistent CRS specification.

To gain further insight into the sources of productivity change, Ray and Desli (1997) proposed an extended decomposition that introduces VRS into the analysis. This triple decomposition separates the overall productivity change into three effects: efficiency change (EC), technical change (TC), and scale efficiency change (SEC). The resulting expression is:

$$M^0_{t_1,t_2} = \sqrt{\frac{D^{t_1}_{\text{CRS}}(\boldsymbol{x}^{t_2}_0, \boldsymbol{y}^{t_2}_0)}{D^{t_1}_{\text{CRS}}(\boldsymbol{x}^{t_1}_0, \boldsymbol{y}^{t_1}_0)} \cdot \frac{D^{t_2}_{\text{CRS}}(\boldsymbol{x}^{t_2}_0, \boldsymbol{y}^{t_2}_0)}{D^{t_2}_{\text{CRS}}(\boldsymbol{x}^{t_1}_0, \boldsymbol{y}^{t_1}_0)}} = \text{EC} \cdot \text{TC} \cdot \text{SEC}, \tag{3.22}$$

with components defined as:

$$\text{EC} = \frac{D^{t_2}_{\text{VRS}}(\boldsymbol{x}^{t_2}_0, \boldsymbol{y}^{t_2}_0)}{D^{t_1}_{\text{VRS}}(\boldsymbol{x}^{t_1}_0, \boldsymbol{y}^{t_1}_0)}, \tag{3.23}$$

$$\text{TC} = \sqrt{\frac{D^{t_1}_{\text{VRS}}(\boldsymbol{x}^{t_2}_0, \boldsymbol{y}^{t_2}_0)}{D^{t_2}_{\text{VRS}}(\boldsymbol{x}^{t_2}_0, \boldsymbol{y}^{t_2}_0)} \cdot \frac{D^{t_1}_{\text{VRS}}(\boldsymbol{x}^{t_1}_0, \boldsymbol{y}^{t_1}_0)}{D^{t_2}_{\text{VRS}}(\boldsymbol{x}^{t_1}_0, \boldsymbol{y}^{t_1}_0)}}, \tag{3.24}$$

$$\text{SEC} = \sqrt{\frac{\text{SE}^{t_2}(\boldsymbol{x}^{t_2}_0, \boldsymbol{y}^{t_2}_0)}{\text{SE}^{t_1}(\boldsymbol{x}^{t_1}_0, \boldsymbol{y}^{t_1}_0)} \cdot \frac{\text{SE}^{t_2}(\boldsymbol{x}^{t_1}_0, \boldsymbol{y}^{t_1}_0)}{\text{SE}^{t_1}(\boldsymbol{x}^{t_2}_0, \boldsymbol{y}^{t_2}_0)}}, \tag{3.25}$$

where the scale efficiency (SE) at time t is given by:

$$\text{SE}^t(\boldsymbol{x}_0, \boldsymbol{y}_0) = \frac{D^t_{\text{CRS}}(\boldsymbol{x}_0, \boldsymbol{y}_0)}{D^t_{\text{VRS}}(\boldsymbol{x}_0, \boldsymbol{y}_0)}. \tag{3.26}$$

The Luenberger Productivity Indicator

The LPI, originally proposed by Chambers et al. (1996b) and Chambers and Pope (1996), emerges as an (additive) alternative framework to the MPI for measuring productivity changes over time. While both indicators rely on distance functions, the LPI is specifically formulated using the concept of DDFs, which provide a flexible approach to modeling simultaneous input contractions and output expansions

To compute the directional distance, a vector $\boldsymbol{g} = (\boldsymbol{g}_x, \boldsymbol{g}_y) \in \mathbb{R}^m \times \mathbb{R}^s$ is chosen, specifying the intended direction of improvement in the input-output space. Then, the DDF at period t under VRS is defined as:

$$\vec{DDF}^t(\boldsymbol{x}_0, \boldsymbol{y}_0; \boldsymbol{g}) = \sup\left\{\omega : (\boldsymbol{x}_0 - \omega \boldsymbol{g}_x, \boldsymbol{y}_0 + \omega \boldsymbol{g}_y) \in \varphi^t_{\text{VRS}}\right\}, \tag{3.27}$$

where φ^t_{VRS} denotes the VRS technology set at time t, as defined in Equation (3.10). This formulation corresponds to the model described in (3.16).

Based on this formulation, the LPI between two periods t_1 and t_2 for a given observation $(\boldsymbol{x}_0, \boldsymbol{y}_0)$ is calculated as:

$$L^0_{t_1,t_2} = \frac{1}{2}\left[\left(\vec{DDF}^{t_2}(\boldsymbol{x}_0^{t_1}, \boldsymbol{y}_0^{t_1}; \boldsymbol{g}) - \vec{DDF}^{t_2}(\boldsymbol{x}_0^{t_2}, \boldsymbol{y}_0^{t_2}; \boldsymbol{g})\right) + \left(\vec{DDF}^{t_1}(\boldsymbol{x}_0^{t_1}, \boldsymbol{y}_0^{t_1}; \boldsymbol{g}) - \vec{DDF}^{t_1}(\boldsymbol{x}_0^{t_2}, \boldsymbol{y}_0^{t_2}; \boldsymbol{g})\right)\right]. \quad (3.28)$$

A positive value of $L^0_{t_1,t_2}$ indicates productivity growth, while a negative value reflects productivity deterioration. Similar to the MPI, the LPI can be decomposed into two components: efficiency change (ECH) and technical change (TCH). These components are defined as follows:

$$\text{ECH} = \vec{DDF}^{t_1}(\boldsymbol{x}_0^{t_1}, \boldsymbol{y}_0^{t_1}; \boldsymbol{g}) - \vec{DDF}^{t_2}(\boldsymbol{x}_0^{t_2}, \boldsymbol{y}_0^{t_2}; \boldsymbol{g}) \quad (3.29)$$

$$\text{TCH} = \frac{1}{2}\left[\left(\vec{DDF}^{t_2}(\boldsymbol{x}_0^{t_2}, \boldsymbol{y}_0^{t_2}; \boldsymbol{g}) - \vec{DDF}^{t_1}(\boldsymbol{x}_0^{t_2}, \boldsymbol{y}_0^{t_1}; \boldsymbol{g})\right) + \left(\vec{DDF}^{t_2}(\boldsymbol{x}_0^{t_1}, \boldsymbol{y}_0^{t_1}; \boldsymbol{g}) - \vec{DDF}^{t_1}(\boldsymbol{x}_0^{t_1}, \boldsymbol{y}_0^{t_1}; \boldsymbol{g})\right)\right]. \quad (3.30)$$

3.2 Multivariate Adaptive Regression Splines

Unlike production frontier estimation, which focuses on studying extreme behaviors in the data, traditional regression techniques—in statistics—aim to explain or predict average behaviors, referring implicitly to the general trend observed in the available information. This type of analysis focuses on modeling the relationship between a response or dependent variable y and a set of independent variables, predictors, or covariates $(x_1, \ldots, x_m)$, using a training (or learning) dataset $\mathcal{L} = \{\boldsymbol{x}_i, y_i\}_{i=1}^n$. The data generating process (DGP) can be described by the following expression:

$$y = f(x_1, \ldots, x_m) + \varepsilon, \quad (3.31)$$

where the function $f(x_1, \ldots, x_m)$ captures the relationship between the response variable y and the set of selected predictors $(x_1, \ldots, x_m)$ over a certain domain $\mathcal{D} \subset \mathbb{R}^m$ that contains the observed data. On the other hand, the stochastic term ε, whose expected value is assumed to be zero, reflects the variability in y that cannot be explained by the set of selected predictors. The objective is therefore to estimate a mathematical expression $\hat{f}(x_1, \ldots, x_m)$ that approximates the true function $f(x_1, \ldots, x_m)$ as closely as possible. The different methods applied for this purpose can be classified as either parametric or nonparametric.

Linear regression stands out as the predominant technique in the parametric domain, which assumes a linear dependence between the set of covariates and the response variable: $f(x_1, \ldots, x_m) = \sum_{j=1}^m \gamma_j x_j$. This methodology is especially popular in the social sciences due to the simplicity of its approach and the direct interpretability of its results. However, this simplicity stems from restrictive assumptions about the structure of the data, such as linear dependence among variables,

homoscedasticity, and normality in the distribution of the random error, among others. As a result, these assumptions may lead to models that do not adequately reflect the reality of the observed data.

In response to these limitations, more advanced parametric methods have been developed to provide greater flexibility and accuracy in modeling. Some of these methods are outlined below. Polynomial regression, for instance, allows for capturing nonlinear relationships among variables. Ridge regression, introduced by Hoerl and Kennard (1970), adds a penalty term to address potential multicollinearity issues. Generalized Linear Models, proposed by Nelder and Wedderburn (1972), extend the applicability of linear models to dependent variables that follow a variety of probability distributions, not limited exclusively to the normal distribution. Similarly, following the approach of Ridge regression, Tibshirani (1996) developed the Least Absolute Shrinkage and Selection Operator (LASSO) method, which facilitates variable selection by imposing penalties that can shrink some coefficients to zero. Finally, the Elastic Net regression, introduced by Zou and Hastie (2005), combines the advantages of Ridge and LASSO by balancing variance reduction with variable selection.

On the other hand, non-parametric techniques offer an appealing alternative by avoiding assumptions about the relationship between variables or the distribution of errors, allowing greater flexibility in estimating the predictive function. In particular, spline-based methods stand out for their ability to approximate the underlying function using piecewise linear polynomials. These methods divide the domain $\mathcal{D}$ into contiguous intervals by placing points called knots, and fit a polynomial to the data within each interval. These knots typically identify trend changes within the data. However, the effectiveness of spline-based methods can be affected by the complexity of predefining both the location and number of knots, especially in high-dimensional settings (Friedman, 2001).

To address these limitations, several methods have been developed to optimize the empirical selection of knots through recursive partitioning of the input space, such as the Classification and Regression Trees (CART) algorithm by Breiman et al. (1984), and the MARS algorithm by Friedman (1991). Among the key differences between these two techniques, particular attention is paid to the modeling approach. While CART produces stepwise models by assigning constant estimates of the response variable within each region defined by the selected knots, MARS adjusts the slope of the estimator within those regions, thereby avoiding potential discontinuities. A detailed comparison of these methods can be found in the work of Zhang and Singer (2010).

In particular, MARS is a nonparametric regression technique specifically designed to handle high-dimensional scenarios with nonlinear relationships and complex interactions in the data. The resulting model is a linear basis expansion, where the basis functions (BFs) are derived from an exhaustive search for the optimal locations of the knots, using a recursive partitioning approach of the input space. Friedman (1991) suggests using piecewise linear truncated univariate splines of degree 1 as BFs:

$$B^{+}(\boldsymbol{x}) = (x_j - k)_+ = \max\{0, x_j - k\}, \tag{3.32}$$
$$B^{-}(\boldsymbol{x}) = (k - x_j)_+ = \max\{0, k - x_j\}$$

Each of these functions is piecewise linear, with a knot at point k. Based on the previous expression, we can define the concept of a reflected pair as a pair of BFs that share a common knot location (k). Specifically, a reflected pair consists of a right-sided spline and a left-sided spline. The right-sided spline captures the relationship between predictor x_j and the response variable on the right side of the knot, while the left-sided spline does the same on the left side of the knot. When a spline is removed from a reflected pair, resulting in the loss of its counterpart, we refer to it as an unpaired BF. This terminology allows us to distinguish between splines that remain part of a reflected pair and those that become unpaired during the model fitting process.

Figure 3.4 illustrates the different statuses of a BF. BFs with knot locations at $x_1 = 1.88$ (red), $x_1 = 4.58$ (blue), and $x_1 = 8.16$ (purple) are paired (discontinuous and continuous line types). On the other hand, the BF with a knot location at $x_1 = 3.88$ (green, discontinuous line type) is an unpaired left-sided BF, whereas the BF with a knot at $x_1 = 5.98$ (yellow, continuous line type) is an unpaired right-sided BF.

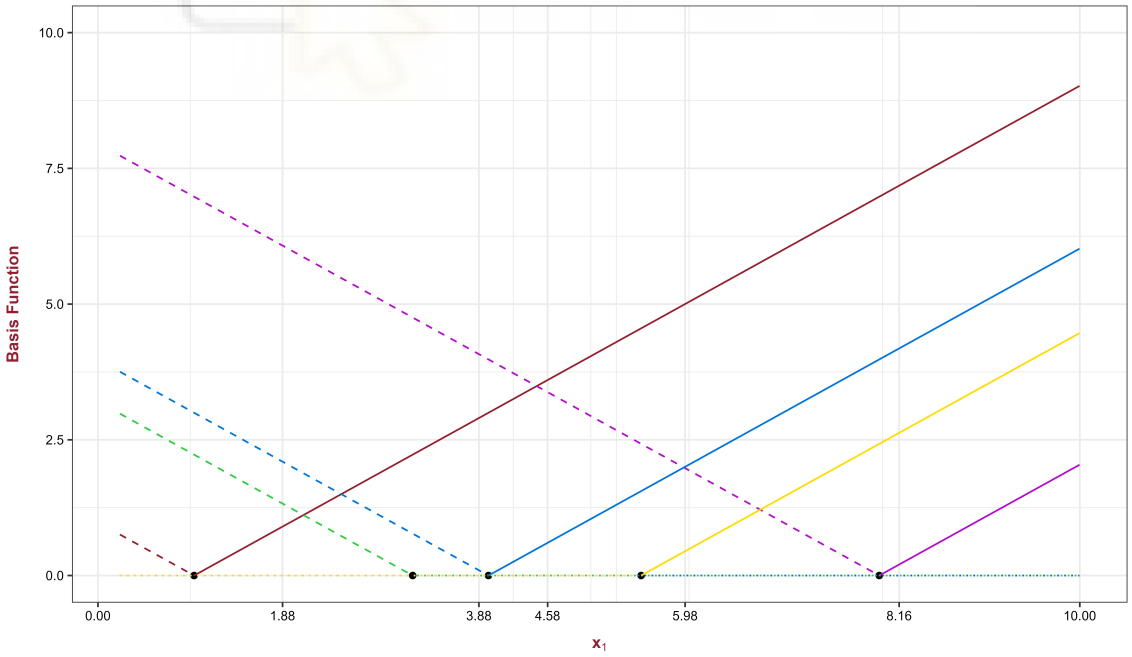

Figure 3.4: Example of different basis functions.

3.2.1 The forward algorithm

The fitting process of the MARS model consists of two procedures: a forward selection algorithm and a backward elimination algorithm. During the forward selection phase, the input space is recursively divided into new (overlapping) subregions through an intensive search for optimal knot locations across the range of the predictors. These knots are used to create reflected pairs that transform the original variables of the problem into additional features. The set of BFs used to create reflected pairs during the forward phase is as follows:

$$\mathcal{C} = \{(x_j - k)_+, (k - x_j)_+\}_{k \in \{x_{1j}, x_{2j}, \ldots, x_{nj}\}, \quad j=1,\ldots,m} \tag{3.33}$$

Although each of these functions depends on a single variable (x_j), they are treated as functions defined on the full input space $\mathbb{R}^m$. The model-building strategy is similar to a stepwise linear regression, but instead of using the original inputs, it uses the functions defined in set (3.33) and their products. Therefore, the model takes the following form:

$$\hat{f}^{\text{MARS}}(\boldsymbol{x}) = b_0 + \sum_{l=1}^{L} b_l \cdot B_l^{(q)}(\boldsymbol{x}), \tag{3.34}$$

where $B_l^{(q)}(\boldsymbol{x})$ represents the l-th BF of degree q, and L is the total number of BFs selected in the final model.. When $q = 1$, $B_l^{(q)}(\boldsymbol{x})$ is referred to as a univariate BF and is selected directly from set (3.33). In contrast, if $q > 1$, the BF results from the product of q functions from set (3.33), thus forming a multivariate BF that incorporates variable interaction information. Specifically, multivariate BFs are formed by taking the product of a new univariate spline, as defined in Equation (3.32), and a BF already included in the model, whether univariate or multivariate:

$$\begin{aligned} &B_l^{(q)}(\boldsymbol{x}) \cdot (x_j - k)_+ \,, \\ &B_l^{(q)}(\boldsymbol{x}) \cdot (k - x_j)_+ \,. \end{aligned} \tag{3.35}$$

Below we outline some important considerations regarding the generation of multivariate BFs. First, each factor in the BF must involve a different predictor variable to avoid high-degree individual dependencies, which may be highly sensitive to extreme values. Second, a hierarchical structure is maintained, meaning that a BF of degree q cannot be included in the model until a function of degree $q - 1$ is already present. Finally, these BFs are only activated in the region of the input space where all splines involved in the multivariate BF take non-zero values.

For example, consider the following BF that incorporates the interaction of three predictor variables:

$$B_l^{(3)}(\boldsymbol{x}) = B_l^{(2)}(\boldsymbol{x}) \cdot (x_1 - 5)_+ \quad \text{where} \quad B_l^{(2)}(\boldsymbol{x}) = (x_3 - 9)_+ \cdot (7 - x_2)_+ \,.$$

The terms $(x_1 - 5)_+$, $(x_3 - 9)_+$, and $(7 - x_2)_+$ are univariate splines and their product constitutes the multivariate BF. Finally, it is worth noting that the maximum interaction degree among variables is typically bounded by a user-defined hyperparameter ($q_{\max}$), generally set to 2 or 3.

The algorithm is initialized by adding the constant function $B_0^{(0)}(\boldsymbol{x}) = 1$ to the model to establish the initial region across the entire domain. Then, a new reflected pair from set (3.33) is selected and multiplied by another BF previously introduced in the model, referred to as the parent term. Only if the parent term corresponds to $B_0^{(0)}(\boldsymbol{x})$, the new BF is considered univariate. Based on this process, the coefficient vector

$$\boldsymbol{b} = (b_0, b_1, b_2, \ldots, b_{L-1}, b_L)$$

is estimated by minimizing the sum of squared residuals:

$$\boldsymbol{b} = \arg\min_{\boldsymbol{b}} \sum_{i=1}^{n} \left(y_i - \left(b_0 + \sum_{l=1}^{L} b_l \cdot B_l^{(q)}(\boldsymbol{x}) \right) \right)^2 \tag{3.36}$$

Finally, the BF—either univariate or multivariate—that achieves the greatest reduction in the lack-of-fit is selected and added to the model. The forward phase of the MARS algorithm, while powerful, is computationally intensive. At each iteration, the algorithm must evaluate approximately $n \cdot m$ different models. This complexity stems from the exhaustive search for optimal knot locations and BFs combinations. Naturally, once a knot has been incorporated into the model, it is not reconsidered. Nevertheless, the process remains demanding, especially in high-dimensional settings or with large datasets.

Beyond its computational cost, the forward phase is also susceptible to numerical instability caused by multicollinearity. Every new knot generates a reflected pair of BFs. When two knots are positioned too closely together, their associated BFs often become nearly collinear. This near collinearity undermines the stability of the least squares estimation, resulting in inflated coefficient variances, overfitting to local noise, and reduced generalizability of the model.

To mitigate these issues, several safeguards have been proposed. Notably, Friedman (1991) and Zhang (1994) proposed constraints that enforce a minimum separation between successive knots. This separation is governed by two parameters that limit knot placement based on the distribution of the data: the minimum span (minspan) and the end span (endspan) constraints. These parameters are formally defined below.

The minspan constraint establishes a lower bound on the number of observations that must lie between any two interior knots. Following the approach proposed by Friedman (1991), the minimum span can be computed as:

$$\text{minspan} = -\frac{\log_2\left[-\frac{1}{mB_n}\cdot\ln(0.95)\right]}{2.5}. \tag{3.37}$$

In this expression, m denotes the total number of predictor variables in the model, and B_n represents the number of observations contained within the active region of the parent BF under consideration. In the case of an additive MARS model—where no interactions are present and, in consequence, the parent BF corresponds to the constant function—this region spans the entire input space, and thus $B_n = n$.

On the other hand, the endspan constraint complements the minspan rule by ensuring adequate spacing between an extreme knot and the boundary of the domain. Following Friedman (1991), the minimum number of observations required near the domain boundaries is computed as:

$$\text{endspan} = 3 - \log_2\left(\frac{0.05}{m}\right). \tag{3.38}$$

The expressions for minspan and endspan in Equations (3.37) and (3.38) assume that the data are equally spaced. However, this is rarely the case in practice. When data are unevenly distributed, Zhang and Singer (2010) suggest replacing the count of observations with a distance-based criterion. For predictor j this involves rescaling by the average spacing $\frac{(x_{nj}-x_{1j})}{n-1}$, where the values $x_{1j} < x_{2j} < ...x_{nj}$ denote the sorted observations along that variable. The resulting robust constraint ensures a minimum average distance between adjacent knots and is defined as:

$$\text{minspan}_j^{\star} = \max\left\{\frac{x_{nj}-x_{1j}}{2.5(n-1)}\cdot\log_2\left[-\frac{1}{n}\ln(0.95)\right], \frac{1}{n}\sum_{i=1}^{3}(x_{n-i+1j}-x_{ij})\right\} \tag{3.39}$$

3.2.2 The backward algorithm

The forward algorithm generates BFs while iteratively reducing the lack-of-fit criterion, stopping either when a predefined maximum number of terms (η) is reached, or when the training error no longer decreases significantly—according to a user-specified threshold (ξ). At the end of the forward stage, the estimator tends to overfit the training data. In this phase, the model's precision is relatively high due to its close fit to the observed sample, resulting in low bias. However, the estimator heavily depends on the training data, leading to high variance and reduced generalization capacity when applied to new data. To address this issue, a backward elimination algorithm is applied with the objective of removing less relevant BFs, relying on the generalized cross-validation (GCV) metric, initially proposed by Golub et al. (1979):

$$\mathrm{GCV}(\mathcal{B}) = \frac{\frac{1}{n}\sum_{i=1}^{n}\left[y_i - \hat{f}_{\mathcal{B}}^{\mathrm{MARS}}(\boldsymbol{x}_i)\right]^2}{\left[1 - \frac{C(\mathcal{B}) + d \cdot \chi(\mathcal{B})}{n}\right]^2}. \tag{3.40}$$

In this context, $C(\mathcal{B})$ denotes the number of parameters estimated in the model $\hat{f}_{\mathcal{B}}^{\mathrm{MARS}}$, where $\mathcal{B}$ is a specific set of BFs. This includes the intercept and one coefficient per active BF. The hyperparameter d penalizes model complexity by adjusting the contribution of the number of knots to the overall fit. The term $\chi(\mathcal{B})$ represents the total number of knots used to construct $\mathcal{B}$, that is, the number of hinge points involved in generating the BFs.

Therefore, the backward elimination algorithm simplifies the model obtained during the forward stage by iteratively removing one BF at a time[1]. At each iteration, it evaluates the effect of removing each individual BF from the current model $\mathcal{B}$ and identifies the removal that leads to the greatest reduction in the GCV value. The corresponding reduced model is retained, and the procedure is repeated with this new model as the current one. This process continues until only the intercept term $B_0^{(0)}(\boldsymbol{x}) = 1$ remains, which cannot be removed. As a result, a sequence of nested models is generated—each containing one fewer BF than the previous one—culminating in a total of $|\mathcal{B}|$ candidate models, including the original forward-selected model. The final model is then selected as the one within this sequence that achieves the minimum GCV value.

Additionally, Milborrow (2014) proposed a standardized version of the GCV metric:

$$1 - \frac{\mathrm{GCV}(\mathcal{B})}{\mathrm{GCV}_0(\mathcal{B})}, \tag{3.41}$$

where $\mathrm{GCV}_0(\mathcal{B})$ denotes the GCV value for the model containing only the intercept. Both the raw and standardized GCV values assess the model's generalization capacity, offering a measure of its expected performance on unseen data. A negative standardized GCV suggests overparameterization and poor out-of-sample performance. Following this approach, Milborrow (2014) recommends using a threshold of -10 as an additional stopping criterion during the forward selection stage of the MARS algorithm.

Figure 3.5 illustrates the estimator obtained using the MARS algorithm to estimate the average behavior of a set of 100 DMUs that use a single input to produce a single output. The figure highlights the differences between the estimators resulting from the forward stage (gold) and the backward stage (green). Notably, the forward estimator captures all the random noise present in the data, whereas the backward estimator successfully isolates the underlying trend.

[1]It is worth noting that removing a BF from a reflected pair reduces $C(\mathcal{B})$ by one—since one fewer coefficient is estimated—but does not necessarily affect $\chi(\mathcal{B})$, as the knot associated with that pair could still exists in the remaining spline component.

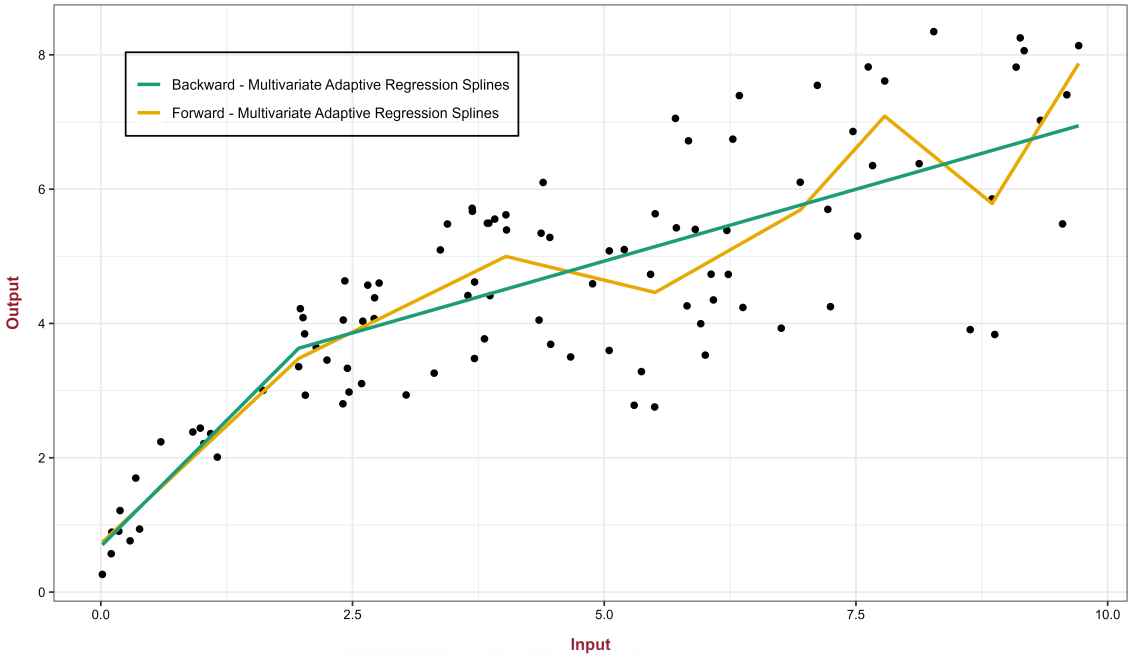

Figure 3.5: Comparative of MARS estimators: forward and backward

3.2.3 The smoothing procedure

In this section, we introduce two smoothed versions of the MARS algorithm that employ cubic BFs (Friedman, 1991) and quintic BFs (Chen et al., 1999). These approaches enable the construction of models with continuous first- and second-order derivatives, respectively. It is important to note that such techniques should only be applied when the underlying function is assumed to be smooth, as is often the case in production theory—for example, with Cobb-Douglas-type production functions.

Friedman (1991) and Chen et al. (1999) suggest smoothing the model resulting from the backward elimination stage, which is characterized by the lowest associated GCV value. This model consists of a set of BFs and the corresponding knots that define them. The following describes the standard procedure for creating a new set of knots needed to smooth the model, focusing specifically on the case where only univariate BFs are selected[2].

The first step in either of the smoothing approaches is to expand the collection of surviving knots from the previous stage. From the backward stage, we can define the set of (surviving) selected central knots as:

[2]The extension of these smoothing procedures to include multivariate BFs, although straightforward, is beyond the scope of this thesis. For a more detailed exploration of the subject, see Friedman (1991) and Chen et al. (1999).

$$\mathcal{K} = \{K_1, \ldots, K_m\} = \left\{ \left\{ k_{1_1}, \ldots, k_{1_{|K_1|}} \right\}, \ldots, \left\{ k_{1_m}, \ldots, k_{1_{|K_m|}} \right\} \right\}, \quad k_{j_1} < \cdots < k_{j_{|K_j|}}, \tag{3.42}$$

where k_{j_c}, $c = 1, \ldots, |K_j|$, is the c-th selected knot (in ascending order) in the j-th dimension. However, in the smoothing phase, both cubic and quintic BFs are constructed using three knots: a central knot k_{j_c}, and two additional side knots: $k_{j_c}^+$ and $k_{j_c}^-$. Regarding their placement, the side knots are positioned at the midpoints between the central knot and its adjacent central knots in the same dimension. In the case of central knots located at the boundaries of the data range $\left(k_{j_1},\ k_{j_{|K_j|}}\right)$, the corresponding side knots are placed at the midpoints between these central knots and the extreme values of the data. Therefore, for each central knot obtained in the backward phase, a new configuration is defined to support smoothing.

Figure 3.6 shows a diagram illustrating the construction of the knot triplets associated with three central knots obtained after the backward phase:

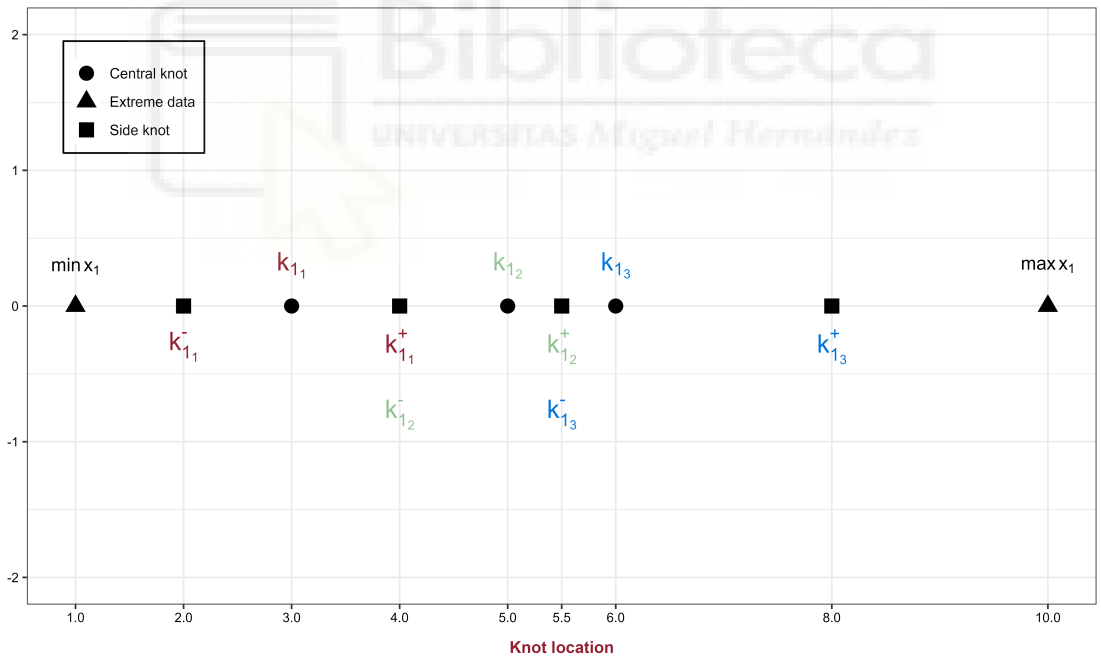

Figure 3.6: Diagram illustrating the construction of the knot triplets.

Once the new set of knots has been established, the surviving BFs from the backward stage must be replaced—ensuring that the central knots match—by the cubic BFs proposed by Friedman (1991) in order to obtain a model with continuous first derivatives:

$$B_C^+(x;k,k^+,k^-) = \begin{cases} 0, & x \le k^-, \\ \frac{2k^+ + k^- - 3k}{(k^+ - k^-)^2}(x-k^-)^2 + \frac{2k - k^+ - k^-}{(k^+ - k^-)^3}(x-k^-)^3, & k^- < x < k^+, \\ x - k, & x \ge k^+, \end{cases} \tag{3.43}$$

$$B_C^-(x;k,k^+,k^-) = \begin{cases} k - x, & x \le k^-, \\ \frac{3k - 2k^- - k^+}{(k^- - k^+)^2}(x-k^+)^2 + \frac{k^- + k^+ - 2k}{(k^- - k^+)^3}(x-k^+)^3, & k^- < x < k^+, \\ 0, & x \ge k^+. \end{cases} \tag{3.44}$$

On the other hand, to obtain a model with continuous second derivatives, the same procedure used in the cubic version is followed, but using the new set of quintic functions developed by Chen et al. (1999):

$$B_Q^+(x;k,k^+,k^-) = \begin{cases} 0, & x \le k^-, \\ \Lambda_1^+ \cdot (x-k^-)^3 + \Lambda_2^+ \cdot (x-k^-)^4 + \Lambda_3^+ \cdot (x-k^-)^5, & k^- < x < k^+, \\ x - k, & x \ge k^+, \end{cases} \tag{3.45}$$

where $\Lambda_1^+ = \frac{6(k^+ - k) - 4(k - k^-)}{(k^+ - k^-)^3}$, $\Lambda_2^+ = \frac{-8(k^+ - k) + 7(k - k^-)}{(k^+ - k^-)^4}$ and $\Lambda_3^+ = \frac{3(k^+ - k) - 3(k - k^-)}{(k^+ - k^-)^5}$.

$$B_Q^-(x;k,k^+,k^-) = \begin{cases} k - x, & x \le k^-, \\ \Lambda_1^-(x-k^+)^3 + \Lambda_2^-(x-k^+)^4 + \Lambda_3^-(x-k^+)^5, & k^- < x < k^+, \\ 0, & x \ge k^+, \end{cases} \tag{3.46}$$

where $\Lambda_1^- = \frac{4(k^+ - k) - 6(k - k^-)}{(k^+ - k^-)^3}$, $\Lambda_2^- = \frac{7(k^+ - k) - 8(k - k^-)}{(k^+ - k^-)^4}$ and $\Lambda_3^- = \frac{3(k^+ - k) - 3(k - k^-)}{(k^+ - k^-)^5}$.

In expressions (3.43), (3.44), (3.45), and (3.46), the superscript "+" indicates that the BF is right-sided, while the superscript "–" indicates that the BF is left-sided. The symbol k denotes a knot that has survived the backward elimination stage, and k^+ and k^- refer to its associated side knots, whose locations are computed following the procedure described above.

3.2.4 Further extensions of MARS

Since its introduction by Friedman, MARS has established itself as a flexible and effective tool for nonparametric regression. Nonetheless, the method exhibits several limitations that have prompted numerous methodological enhancements. Key concerns highlighted in the literature include the potential instability of parameter estimates, sensitivity to outliers, and the computational burden associated with the two-stage (forward and backward) selection procedure. In response, various extensions have been proposed to improve numerical stability, enhance robustness, streamline the model selection process, and incorporate regularization techniques.

Chen et al. (1999) introduced a quintic BF to smooth the estimator, thereby obtaining a MARS model with continuous second derivatives. Bakin et al. (2000) developed a new version of MARS, named BMARS, which uses second-order B-splines instead of truncated linear functions to improve the numerical stability of parameter estimation. Tsai and Chen (2005) proposed two new MARS variants: first, by implementing automatic stopping rules based on the (adjusted) coefficient of determination, instead of allowing the forward algorithm to run until reaching the maximum number of BF (thus removing the backward stage); and second, by developing a robust version to reduce the order of interaction terms. These strategies allowed Tsai and Chen (2005) to lower the computational cost of MARS and improve its performance in the presence of outliers. Taylan et al. (2010) provided parameter estimates for partially linear generalized models using B-splines and conic quadratic programming, which may serve as a foundation for future MARS developments. Weber et al. (2012) proposed a new approach called CMARS, in which the backward stepwise algorithm is modified using a penalized residual sum of squares formulation, cast as a Tikhonov regularization problem solvable via conic quadratic programming. Later, Özmen et al. (2011) and Özmen and Weber (2014) improved CMARS (RCMARS) and MARS (RMARS), respectively, using robust optimization techniques to manage data uncertainty (see also Özmen et al., 2017). Yazıcı et al. (2015) presented an enhancement of MARS over the CMARS formulation, incorporating a bootstrap procedure (BCMARS) to obtain the empirical distribution of fitted parameters and assess their significance. Koc and Bozdogan (2015) introduced another alternative to the conventional backward algorithm by employing the theoretical measure of information complexity (ICOMP) for model selection. Martinez et al. (2015) proposed a convex version of MARS by modifying how interaction terms are introduced and by imposing coefficient constraints to eliminate the method's inherent non-convexity. In addition, Zhang (1994) and Koc and Iyigun (2014) modified the forward algorithm by implementing new knot selection procedures. Finally, Murat (2023) proposed a strategy to detect outliers during the variable selection process in MARS by constructing a design matrix that adds as many dummy variables as potential outliers identified in the observed data.

3.3 Random Forest

In the field of machine learning, it is widely acknowledged that models based on a single predictive function often exhibit low bias but high variance (Berk et al., 2008). One way to address this issue is through bagging or bootstrap aggregation, which involves generating multiple predictors from bootstrap samples and aggregating their outputs. This methodology, introduced by LeBlanc and Tibshirani (1996), combines predictions from several models to produce a single estimate of the response variable. Specifically, in tree-based models, the RF technique (Breiman, 2001) represents a significant evolution of bagging.

RF builds upon decision trees of the CART type (Breiman et al., 1984). A CART tree is constructed through a recursive binary partitioning of the input space using a greedy algorithm. At each node, all possible splits are exhaustively evaluated by testing every predictor variable and potential threshold, and the split that yields the greatest reduction in a predefined impurity measure is selected. This process is repeated recursively: starting from the root node, the data are split into two child nodes, and the partitioning continues until no further improvement can be achieved or a stopping criterion is met. In each terminal node, the prediction corresponds to the average value of the response variable among the data points in that node. Once the tree is fully grown, a pruning phase is typically applied to remove splits that do not provide a significant gain in predictive accuracy. This post-processing step reduces model complexity and helps avoid overfitting by selecting a simpler sub-tree that balances predictive performance and interpretability.

The training phase of RF consists of several key steps. Initially, the user specifies the desired number of CART-type trees to be built. Typically, it is sufficient to consider between 500 and 1000 trees. Then, two randomization strategies are implemented: one for the training set and another for the feature selection. In the first case, $T_{\max}$ bootstrap samples are generated, denoted by $\mathcal{L}_t^*$, for $t = 1, ..., T_{\max}$, where each sample $\mathcal{L}_t^*$ serves as the basis to create an individual decision tree $T^{\text{CART}}(\mathcal{L}_t^*)$. In the second case, the tree-building procedure randomly selects a reduced set of features to consider at each possible split during the learning process. Unlike the CART method, which uses cross-validation (CV) for individual pruning, RF eliminates the need for individual pruning thanks to its randomization strategy, which significantly reduces the computational complexity of the process. Finally, the predictions of all trees are aggregated using an average to obtain the final estimate. In this way, by integrating multiple randomized models, the performance of a single non-randomized model is surpassed (Berk et al., 2008).

The RF estimation can be defined through the following random variable:

$$Y^{(F)} = \frac{1}{T_{\max}} \sum_{t=1}^{T_{\max}} Y_t^{(T)}, \tag{3.47}$$

where $Y_t^{(T)}$ is the random variable that determines the estimation of the t-th tree $T^{\text{CART}}(L_t)$ in the model, with a total of $T_{\max}$ trees. The variables $Y_1^{(T)}, \ldots, Y_{T_{\max}}^{(T)}$ are identically distributed with mean μ, variance σ^2, and pairwise correlation ρ. Therefore, $\mathbb{E}[Y^{(F)}] = \mathbb{E}[Y^{(T)}]$, and consequently, both estimators are equivalent in terms of bias.

On the other hand, the variance of the RF estimator is defined as:

$$\text{Var}\left[Y^{(F)}\right] = \text{Var}\left[\frac{1}{T_{\max}} \sum_{t=1}^{T_{\max}} Y_t^{(T)}\right]. \tag{3.48}$$

Since independence cannot be assumed among $Y_1^{(T)}, \ldots, Y_{T_{\max}}^{(T)}$, it follows that:

$$\begin{aligned}
\text{Var}\left[Y^{(F)}\right] &= \frac{1}{T_{\max}^2} \sum_{t_1=1}^{T_{\max}} \sum_{t_2=1}^{T_{\max}} \text{Cov}(Y_{t_1}^{(T)}, Y_{t_2}^{(T)}) \\
&= \frac{1}{T_{\max}^2} \left(\sum_{t=1}^{T_{\max}} \text{Var}(Y_t^{(T)}) + \sum_{t_1=1}^{T_{\max}-1} \sum_{t_2=t_1+1}^{T_{\max}} \text{Cov}(Y_{t_1}^{(T)}, Y_{t_2}^{(T)}) \right) \\
&= \frac{1}{T_{\max}^2} \left(\sum_{t=1}^{T_{\max}} \text{Var}(Y_t^{(T)}) + 2 \sum_{t_1=1}^{T_{\max}-1} \sum_{t_2=t_1+1}^{T_{\max}} \rho_{t_1 t_2} \sqrt{\text{Var}(Y_{t_1}^{(T)}) \cdot \text{Var}(Y_{t_2}^{(T)})} \right) \\
&= \frac{1}{T_{\max}^2} \left(T_{\max}\sigma^2 + 2 \cdot \frac{T_{\max}(T_{\max}-1)}{2} \cdot \rho\sigma^2 \right) \\
&= \rho\sigma^2 + \frac{(1-\rho)}{T_{\max}}\sigma^2.
\end{aligned}$$

From the previous expression, it follows that as $T_{\max}$ increases, the second term tends to vanish, while the first term remains constant. RF aims to minimize $\text{Var}\left[Y^{(F)}\right]$ by reducing the tree correlation (ρ) without significantly increasing their individual variance (σ^2). This is achieved through the randomization in the selection of the (reduced) set of features to be considered at each possible split of the individual trees. By varying the predictors used at each split, subsets strongly associated with the response are less likely to dominate all trees, thereby reducing correlation. Typically, the size of the reduced feature set is one third of the total number of features in regression problems.

Finally, the predictor of the RF algorithm for regression can be expressed as:

$$\hat{f}^{\text{RF}}(\boldsymbol{x}_i) = \frac{1}{T_{\max}} \sum_{t=1}^{T_{\max}} \hat{f}_t^{\text{CART}}(\boldsymbol{x}_i), \tag{3.49}$$

where $\hat{f}_t^{\text{CART}}$ is the predictor of the t-th tree $T^{\text{CART}}(L_t)$ in RF.

An attractive feature of using ensemble methods such as RF is the use of out-of-sample data $\mathcal{L} \setminus \mathcal{L}_t^*$ to evaluate the generalization error inherent to statistical learning processes. In particular, RF benefits from the concept of out-of-bag (OOB) estimation. For a given observation $(\boldsymbol{x}_i, y_i)$, the OOB estimate is obtained through the aggregation of predictions extracted from the base models $T^{\text{CART}}(\mathcal{L}_t^*)$ such that $(\boldsymbol{x}_i, y_i) \notin \mathcal{L}_t^*$. In this way, the generalization error is defined as the average of the OOB estimates computed over all samples in the learning set $\mathcal{L}$:

$$\text{err}_{\text{OOB}}(\mathcal{L}) = \frac{1}{n} \sum_{(\boldsymbol{x}_i, y_i) \in \mathcal{L}} \left(y_i - \hat{f}_{\text{OOB}}(\boldsymbol{x}_i) \right)^2, \tag{3.50}$$

where

$$\hat{f}_{\text{OOB}}(\boldsymbol{x}_i) = \frac{1}{|\{t : (\boldsymbol{x}_i, y_i) \notin \mathcal{L}_t^*\}|} \sum_{t:(\boldsymbol{x}_i, y_i) \notin \mathcal{L}_t^*} \hat{f}_t^{\text{CART}}(\boldsymbol{x}_i),$$

and $|.|$ denotes the cardinality of a set. As we can see, the sum is taken over all trees for which the pair $(\boldsymbol{x}_i, y_i)$ was not included in the corresponding bootstrap sample $\mathcal{L}_t^*$.

The concept of generalization can also be used, for instance, to establish a measure of the importance of each feature, which can in turn be used to rank the variables by relevance. In particular, the importance of the j-th input variable is assessed through the following four steps:

Step 1 [Create a permuted dataset]: Generate a new dataset, denoted by $\mathcal{L}^{(-j)}$, which is similar to the original dataset $\mathcal{L}$ but with the j-th variable randomly permuted.

Step 2 [Train a RF model]: Fit a RF model on the new "virtual" training set $\mathcal{L}^{(-j)}$ with the aim of assessing how much the model performance degrades when the j-th variable is permuted.

Step 3 [Compute generalization error]: Calculate the generalization error associated with the RF model trained in Step 2, denoted by $\text{err}_{\text{OOB}}(\mathcal{L}^{(-j)})$.

Step 4 [Compute relative increase in generalization error]: Compute the percentage increase in the model's generalization error when the j-th variable is permuted:

$$\%\text{Inc}(x_j) = 100 \cdot \left(\frac{\text{err}_{\text{OOB}}(\mathcal{L}^{(-j)}) - \text{err}_{\text{OOB}}(\mathcal{L})}{\text{err}_{\text{OOB}}(\mathcal{L})} \right). \tag{3.51}$$

The value of the above expression quantifies the impact of permuting the j-th variable on the model's generalization capacity with respect to unseen data. Higher values indicate greater expected importance of the feature. Conversely, negative values imply that the model performs better when the variable is randomly permuted, suggesting its contribution is purely random. Additionally, to improve robustness, the permutation procedure can be repeated several times, and the final value of $\%\text{Inc}(x_j)$ is computed as the average across model replications.

Lately, Figure 3.7 shows a (toy) predictor obtained using the RF algorithm, composed of five CART-type regression trees. Additionally, the predictions of each individual tree are represented by dashed lines.

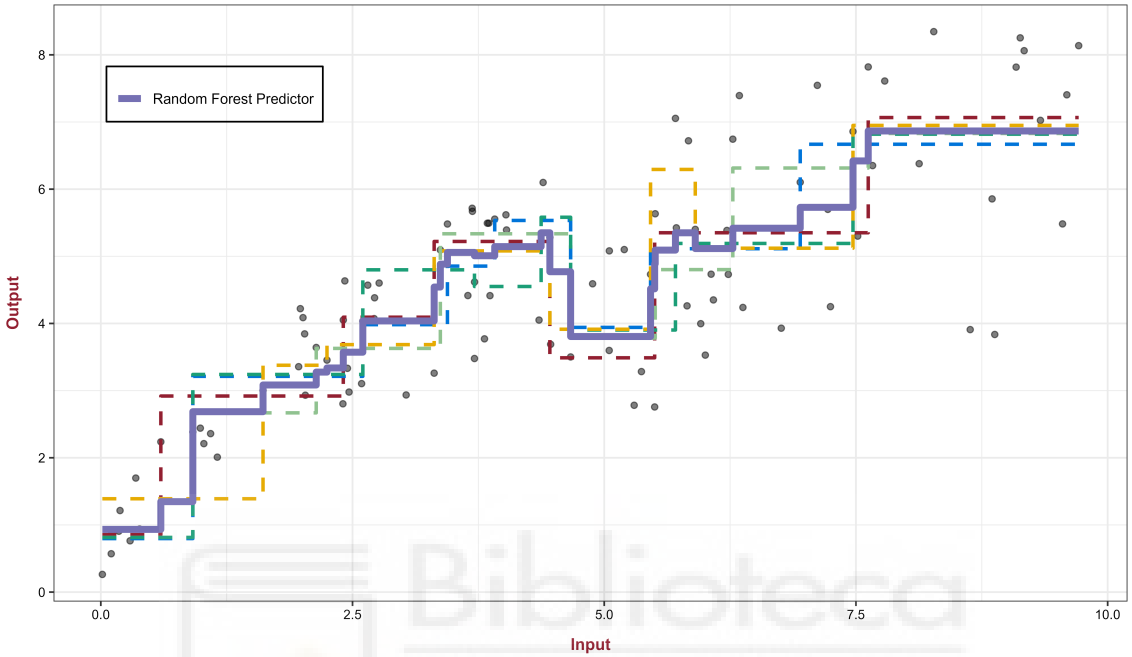

Figure 3.7: A Random Forest predictor composed of five individual regression trees.

3.4 Performance evaluation methods

This section details the simulation procedures employed throughout the thesis to assess the performance of the proposed methodologies: ACES, RF-ACES, and Q-ACES. These simulations serve a dual purpose: first, to calibrate and validate the methods under controlled scenarios; and second, to provide a systematic framework for evaluating predictive accuracy, robustness, and generalizability under diverse data-generating processes. The experimental settings and performance metrics are described in detail to ensure both reproducibility and comparability.

To facilitate exposition, the content is organized into three parts. Subsection 3.4.1 presents the artificial scenarios used to simulate production environments and benchmark the proposed methods against established techniques. Each design targets specific aspects of performance, such as sensitivity to noise, handling of multiple outputs, RTS, or the presence of irrelevant variables. Subsection 3.4.2 introduces the criteria used to assess estimation quality, while Subsection 3.4.3 describes the hardware and software environment employed to conduct the experiments.

3.4.1 Experimental Designs

This subsection presents the DGP used to evaluate the performance of the proposed methodologies against well-established approaches such as DEA by Banker et al. (1984), Corrected Concave Non-parametric Least Squares (CCNLS) by Kuosmanen and Johnson (2010), Stochastic Non-Smooth Envelopment of Data (StoNED) by Kuosmanen and Kortelainen (2012) and Kuosmanen and Johnson (2017), Convexified Efficiency Analysis Trees (CEAT) by Esteve et al. (2020), and Bootstrap DEA (BDEA) by Simar and Wilson (1998, 2000). Each experimental design replicates a specific production environment to test how the models deal with diverse structural features, including single- and multi-output technologies, functional forms with varying flexibility, input–output interactions, and the presence of noise or irrelevant inputs. To ground the simulations in economic theory, we draw on canonical production functions such as the Cobb–Douglas (CD) form, which assumes unitary elasticity of substitution and exhibits CRS or VRS depending on the sum of the output elasticities; the Translog specification, a flexible second-order approximation that captures non-linear substitution patterns; and the Constant Elasticity of Transformation (CET) function, which is a multi-output analogue of the Constant Elasticity of Substitution form and allows a constant, but arbitrary, elasticity of transformation between outputs.

Each simulation generates input variables x_j, frontier (optimal) outputs y_r^*, and observed outputs y_r. Inefficiency terms u are simulated to represent deviations from the production frontier, while stochastic noise components v are included only in selected designs to account for random disturbances. In the presence of such noise, it is possible for some units to exhibit observed outputs y that exceed their corresponding optimal outputs y^*, despite being inefficient, due to favorable random shocks.

Next, we describe the four different frameworks employed throughout the study.

Experimental Design 1: Single-output production functions

The first experimental design is based on the simulation framework proposed by Kuosmanen and Johnson (2010), which aims to evaluate model performance under single-output settings. Table 3.1 summarizes the six simulated scenarios, varying in the number of inputs and the functional form of the target production function $y^* = f(\boldsymbol{x})$. Scenarios 1 and 2 correspond to single-input settings, while Scenarios 3 to 6 introduce interactions between two or three inputs, generating different levels of curvature in the production frontier.

Input variables are independently drawn from a uniform distribution $x_j \sim \mathcal{U}(1, 10)$ for each firm. Inefficiency is introduced by generating a random term $u \sim|\mathcal{N}(0, 0.4)|$, and the observed output is computed as:

$$y = f(\boldsymbol{x}) - u. \tag{3.52}$$

Table 3.1: Simulated scenarios from Kuosmanen and Johnson (2010).

Scenario	# Inputs	Target function $f(\boldsymbol{x})$
1	1	$\ln(x_1) + 3$
2	1	$3 + x_1^{0.5} + \ln(x_1)$
3	2	$0.1x_1 + 0.1x_2 + 0.3(x_1x_2)^{0.5}$
4	3	$0.1x_1 + 0.1x_2 + 0.1x_3 + 0.3(x_1x_2x_3)^{1/3}$
5	2	$0.1x_1 + 0.1x_2 + 0.3(x_1x_2)^{1/3}$
6	3	$0.1x_1 + 0.1x_2 + 0.1x_3 + 0.3(x_1x_2x_3)^{1/4}$

Experimental Design 2: Multi-output Translog production frontier

This second experimental design follows the simulation framework proposed by Perelman and Santín (2009) to model a two-input, two-output production process. Input variables x_1 and x_2 are independently drawn from a uniform distribution $\mathcal{U}(5, 50)$, and the outputs on the production frontier are generated using a flexible Translog functional form. The frontier output values y_1^* and y_2^* are implicitly defined by the following equations:

$$\begin{aligned}
-\ln(y_1^*) = & -1 + 0.5 \cdot \ln\left(\frac{y_2}{y_1}\right) + 0.25 \cdot \ln^2\left(\frac{y_2}{y_1}\right) - 1.5 \cdot \ln(x_1) - 0.6 \cdot \ln(x_2) \\
& + 0.2 \cdot \ln^2(x_1) + 0.05 \cdot \ln^2(x_2) - 0.1 \cdot \ln(x_1) \cdot \ln(x_2) \\
& + 0.05 \cdot \ln(x_1) \cdot \ln\left(\frac{y_2}{y_1}\right) - 0.05 \cdot \ln(x_2) \cdot \ln\left(\frac{y_2}{y_1}\right), \\
-\ln(y_2^*) = & -\ln(y_1^*) - \ln\left(\frac{y_2}{y_1}\right),
\end{aligned} \tag{3.53}$$

where the log-ratio of outputs $\ln\left(\frac{y_2}{y_1}\right)$ is sampled from a uniform distribution $\mathcal{U}(-1.5, 1.5)$.

To model technical inefficiency, we generate a half-normal random variable $u \sim |\mathcal{N}(0, \sqrt{0.3})|$. The observed (inefficient) outputs are then computed as:

$$y_1 = y_1^* \cdot e^{-u}, \quad y_2 = y_2^* \cdot e^{-u}. \tag{3.54}$$

To account for stochastic noise, independent random components are added using normal distributions $v_1, v_2 \sim \mathcal{N}(0, \sqrt{0.01})$. The final observed outputs, incorporating both inefficiency and noise, are calculated as:

$$y_1 = y_1^* \cdot e^{-u} \cdot e^{-v_1}, \quad y_2 = y_2^* \cdot e^{-u} \cdot e^{-v_2}. \tag{3.55}$$

Experimental Design 3: CET-CD multi-output production frontier

The third experimental design is adapted from Färe et al. (1994) and reflects a stylized production setting where technologies exhibit increasing, constant, and decreasing RTS. These simulations are constructed in a two-input, two-output setting. In this framework, the input side of the production function is modeled using a log-linear CD specification, while the output side follows a CET form. The input-output combinations are simulated to satisfy the identity:

$$f(\boldsymbol{x}) = h(\boldsymbol{y}), \tag{3.56}$$

where

$$f(\boldsymbol{x}) = (\sqrt{x_1} \cdot \sqrt{x_2})^{\varpi}, \tag{3.57}$$

and

$$h(\boldsymbol{y}) = \sqrt{\frac{1}{2}y_1^2 + \frac{1}{2}y_2^2}. \tag{3.58}$$

From Equations (3.57) and (3.58), we define four different groups of producers. Once the sample size is fixed, each observation is randomly assigned to one of these groups—small, medium I, medium II, or large—with equal probability (i.e., each group receives close to 25% of the total observations). This approach ensures a balanced representation across different production technologies and supports a realistic simulation of varying RTS. For each group, the first input (x_1) and the first optimal output (y_1^*) are independently drawn from uniform distributions, specifically $x_1 \sim \mathcal{U}(a^x_{\min}, b^x_{\max})$ and $y_1^* \sim \mathcal{U}(a^y_{\min}, b^y_{\max})$, where $(a^x_{\min}, b^x_{\max})$ and $(a^y_{\min}, b^y_{\max})$ represent the lower and upper bounds for the input and output variables, respectively. These bounds vary across groups to reflect differences in scale and production capacity.

Therefore, each group is characterized by a common frontier identity $f(\boldsymbol{x}) = h(\boldsymbol{y})$, a group-specific scale elasticity ϖ used in Equation (3.57), and the specific data generation intervals for x_1 and y_1^*. Table 3.2 summarizes the configuration of each group:

Table 3.2: Configuration of producer groups by scale level and data generation parameters.

Producer size	$f(\boldsymbol{x}) = h(\boldsymbol{y})$	ϖ	x_1	y_1^*
Small	25	0.898	[20, 60]	[10, 35]
Medium I	50	1.000	[30, 80]	[15, 70]
Medium II	75	1.000	[50, 100]	[25, 100]
Large	100	0.927	[90, 230]	[45, 135]

Once the values of x_1 and y_1^* are generated, the second input (x_2) and the second optimal output (y_2^*) are computed to ensure that each observation satisfies the pre-assigned values of $f(\boldsymbol{x})$ and $h(\boldsymbol{y})$. Regarding inefficiency and random noise, these are introduced following the same procedure described in the Experimental Design 2 by Perelman and Santín (2009).

Experimental Design 4: CD production functions with irrelevant inputs

The fourth experimental design introduces two simulation frameworks differentiated by the dimensionality of the input space. The first design replicates the simulations proposed by Nataraja and Johnson (2011). The simulation setting employs a CD production function to generate the optimal output y^* from three input variables $x_1, x_2, x_3 \sim \mathcal{U}(10, 20)$, as defined by the following expression:

$$y^* = x_1^{\varpi_1} \cdot x_2^{\varpi_2} \cdot x_3^{\varpi_3}, \tag{3.59}$$

where the sum of the exponents defines the RTS and each individual value indicates the contribution of each corresponding input to the generation of the output.

To account for inefficiency, we define the observed output y as:

$$y = y^* \cdot e^{-u}, \tag{3.60}$$

where $u \sim \mathcal{N}(0, 0.4)$ represents the inefficiency term[3].

To introduce correlation among the inputs, we use the equation adapted from Wang and Schmidt (2002):

$$x_{j_2} = \rho x_{j_1} + \Psi\sqrt{1-\rho^2}, \tag{3.61}$$

where x_{j_1} is independently generated, and the correlation between x_{j_1} and x_{j_2} is controlled by ρ. A random variable $\Psi \sim \mathcal{U}(10, 20)$ is included to ensure variability. Additionally, an irrelevant variable x_4 is independently generated from a uniform distribution over $[10, 20]$. This variable can be correlated with relevant inputs but does not contribute to the output, serving as an irrelevant variable for the model.

In addition to the base scenario with three inputs and a half-normal inefficiency term, three additional scenarios are explored: one with four relevant inputs (scenario 12), one with only two relevant inputs (scenario 13), and one with an exponential inefficiency term (scenario 14). These scenarios are summarized in Table 3.3:

[3] Nataraja and Johnson (2011) does not provide a fixed value for the standard deviation of the inefficiency term u.

Table 3.3: List of experiments from Nataraja and Johnson (2011).

set	Correlation	Sample size	ϖ_1	ϖ_2	ϖ_3	ϖ_4	Inefficiency
1	Independently generated	100	1/3	1/3	1/3		Half-Normal
2	$\rho_{x_1x_2} = 0.8$, $\rho_{x_1x_3} = 0.2$	100	1/3	1/3	1/3		Half-Normal
3	$\rho_{x_1x_2} = 0.8$, $\rho_{x_1x_3} = 0.8$	100	1/3	1/3	1/3		Half-Normal
4	Independently generated	100	1/3	4/9	2/9		Half-Normal
5	$\rho_{x_1x_2} = 0.8$, $\rho_{x_1x_3} = 0.2$	100	1/3	4/9	2/9		Half-Normal
6	$\rho_{x_1x_2} = 0.8$, $\rho_{x_1x_3} = 0.2$	100	1/3	2/9	4/9		Half-Normal
7	$\rho_{x_1x_4} = 0.8$	100	1/3	1/3	1/3		Half-Normal
8	$\rho_{x_1x_2} = 0.8$	100	1/3	1/3	1/3		Half-Normal
9	Independently generated	25	1/3	1/3	1/3		Half-Normal
10	Independently generated	300	1/3	1/3	1/3		Half-Normal
11	Independently generated	100	1/3	1/3	1/3		Half-Normal
12	Independently generated	100	1/4	1/4	1/4	1/4	Half-Normal
13	Independently generated	100	1/2	1/2			Half-Normal
14	Independently generated	100	1/3	1/3	1/3		Exponential

The second experimental design is developed specifically for this study to assess the performance of the proposed methodologies in scenarios contaminated with irrelevant input variables at varying proportions. The scenarios are systematically constructed by varying three key dimensions: the number of relevant inputs (4 or 8), the ratio of irrelevant to relevant inputs (0.5 or 1.5), and the level of correlation—ranging from low to medium and medium to high—both among the relevant inputs and between relevant and irrelevant ones. All simulations are carried out under the assumption of decreasing returns to scale.

The simulation framework employs a CD production function to generate the theoretical output y^* using two different sets of relevant inputs:

$$y^* = x_1^{0.35} \cdot x_2^{0.25} \cdot x_3^{0.15} \cdot x_4^{0.15}, \tag{3.62}$$

and

$$y^* = x_1^{0.25} \cdot x_2^{0.15} \cdot x_3^{0.15} \cdot x_4^{0.10} \cdot x_5^{0.10} \cdot x_6^{0.05} \cdot x_7^{0.05} \cdot x_8^{0.05}, \tag{3.63}$$

where each relevant input x_j is independently drawn from a uniform distribution $\mathcal{U}(10, 20)$. Additionally, irrelevant inputs $\tilde{x}_j$ are also independently generated from the same uniform distribution. Importantly, these irrelevant inputs have no direct relationship with y^*, but may be strongly correlated with the relevant inputs.

The degree of linear correlation between relevant inputs is randomly generated by using Equation (3.61) and a uniform distribution within specified intervals to simulate varying relationships between variables. These intervals, categorized as low–medium and medium–high correlation ranges, are detailed in the table below.

Table 3.4: Random correlation ranges for relevant inputs.

Inputs	Correlation range: low–medium	Correlation range: medium–high
$x_1 - x_2$	[0, 0.3]	[0.5, 0.7]
$x_3 - x_4$	[0.3, 0.5]	[0.7, 1]
$x_5 - x_6$	[0, 0.3]	[0.5, 0.7]
$x_7 - x_8$	[0.3, 0.5]	[0.7, 1]

The correlation between relevant and irrelevant inputs follows a defined pattern: each irrelevant feature is directly correlated with exactly one relevant feature, while a relevant feature may be correlated with multiple irrelevant features. Tables 3.5 and 3.6 present the correlation structures for scenarios where the number of relevant inputs exceeds the number of irrelevant inputs.

Table 3.5: Random correlation ranges for a scenario with 4 relevant and 2 irrelevant inputs.

Inputs	Correlation range: low–medium	Correlation range: medium–high
$x_1 - \tilde{x}_1$	[0, 0.3]	[0.5, 0.7]
$x_2 - \tilde{x}_2$	[0.3, 0.5]	[0.7, 1]

Table 3.6: Random correlation ranges for a scenario with 8 relevant and 4 irrelevant inputs.

Inputs	Correlation range: low–medium	Correlation range: medium–high
$x_1 - \tilde{x}_1$	[0, 0.3]	[0.5, 0.7]
$x_2 - \tilde{x}_2$	[0.3, 0.5]	[0.7, 1]
$x_3 - \tilde{x}_3$	[0, 0.3]	[0.5, 0.7]
$x_4 - \tilde{x}_4$	[0.3, 0.5]	[0.7, 1]

Conversely, Tables 3.7 and 3.8 depict the correlation structures for scenarios where the number of irrelevant inputs surpasses the number of relevant inputs.

Table 3.7: Random correlation ranges for a scenario with 4 relevant and 6 irrelevant inputs.

Inputs	Correlation range: low–medium	Correlation range: medium–high
$x_1 - \tilde{x}_1$	[0, 0.3]	[0.5, 0.7]
$x_2 - \tilde{x}_2$	[0.3, 0.5]	[0.7, 1]
$x_3 - \tilde{x}_3$	[0, 0.3]	[0.5, 0.7]
$x_4 - \tilde{x}_4$	[0.3, 0.5]	[0.7, 1]
$x_1 - \tilde{x}_5$	[0.3, 0.5]	[0.7, 1]
$x_2 - \tilde{x}_6$	[0, 0.3]	[0.5, 0.7]

Table 3.8: Random correlation ranges for a scenario with 8 relevant and 12 irrelevant inputs.

Inputs	Correlation range: low–medium	Correlation range: medium–high
$x_1 - \tilde{x}_1$	[0, 0.3]	[0.5, 0.7]
$x_2 - \tilde{x}_2$	[0.3, 0.5]	[0.7, 1]
$x_3 - \tilde{x}_3$	[0, 0.3]	[0.5, 0.7]
$x_4 - \tilde{x}_4$	[0.3, 0.5]	[0.7, 1]
$x_5 - \tilde{x}_5$	[0, 0.3]	[0.5, 0.7]
$x_6 - \tilde{x}_6$	[0.3, 0.5]	[0.7, 1]
$x_7 - \tilde{x}_7$	[0, 0.3]	[0.5, 0.7]
$x_8 - \tilde{x}_8$	[0.3, 0.5]	[0.7, 1]
$x_1 - \tilde{x}_9$	[0.3, 0.5]	[0.7, 1]
$x_2 - \tilde{x}_{10}$	[0, 0.3]	[0.5, 0.7]
$x_3 - \tilde{x}_{11}$	[0.3, 0.5]	[0.7, 1]
$x_4 - \tilde{x}_{12}$	[0, 0.3]	[0.5, 0.7]

Finally, to incorporate inefficiency into the model, we define the observed output y as:

$$y = y^* \cdot e^{-u}, \tag{3.64}$$

where $u \sim |\mathcal{N}(0, 0.4)|$ represents the inefficiency component.

3.4.2 Evaluation metrics

All experimental designs involve estimating the true radial output score (ϕ). To this end, we use the available data $\{\boldsymbol{x}_i, \boldsymbol{y}_i\}_{i=1}^n$ to obtain estimates of ϕ through a variety of techniques, denoted generically as $\hat{\phi}^{\mathcal{A}}$, where $\mathcal{A}$ represents the estimation method under consideration.

For Experimental Designs 1 and 4, the true radial efficiency score is computed as:

$$\phi = \frac{y_1^*}{y_1}, \tag{3.65}$$

where y_1^* denotes the frontier output and y_1 the observed output.

For Experimental Design 2, which involves two outputs, the true score is analogously defined under the assumption of proportional reductions in both dimensions:

$$\phi = \frac{y_1^*}{y_1} = \frac{y_2^*}{y_2}. \tag{3.66}$$

In contrast, for experimental design 3, the true radial score is obtained as the ratio between the production function (CD) and the transformation function (CET):

$$\phi = \frac{f(\boldsymbol{x})}{h(\boldsymbol{y})}, \tag{3.67}$$

where $f(\boldsymbol{x})$ denotes the frontier output given inputs $\boldsymbol{x}$, and $h(\boldsymbol{y})$ is the implicit value of the CET function corresponding to the observed output vector $\boldsymbol{y}$.

To assess the accuracy of the estimated efficiency scores $\hat{\phi}^{\mathcal{A}}$, we employ three standard performance metrics: mean squared error (MSE), bias, and Pearson's linear correlation coefficient (ρ). The MSE captures the average squared deviation between the estimated and the true scores and is defined as

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{\phi}_i^{\mathcal{A}} - \phi_i\right)^2. \tag{3.68}$$

This metric penalizes larger deviations more heavily and thus reflects both the variance and the bias of the estimator. Lower values of MSE indicate better predictive accuracy, as the estimates are closer to the true efficiency scores on average.

To complement this, we compute the bias[4] as the average signed deviation between the estimated and the true scores:

$$\text{Bias} = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{\phi}_i^{\mathcal{A}} - \phi_i\right). \tag{3.69}$$

This measure reveals the direction and magnitude of any systematic error in the estimation. A positive bias indicates that the method tends to overestimate efficiency, while a negative bias implies underestimation.

Lastly, we evaluate the rank-preserving capacity of each method through Pearson's linear correlation coefficient:

[4]This definition departs from the usual statistical notion of bias based on absolute deviations, but the signed version facilitates comparison with methods such as CCNLS from Kuosmanen and Johnson (2010).

$$\rho = \frac{\text{Cov}(\hat{\phi}^{\mathcal{A}}, \phi)}{\sigma_{\hat{\phi}^{\mathcal{A}}} \cdot \sigma_{\phi}}. \tag{3.70}$$

This coefficient takes values in the range $[-1, 1]$ and measures the strength and direction of the linear association between the estimated and true scores. A value near 1 indicates that the method preserves the relative ordering of the DMUs well, even if the estimates are not numerically identical to the true values.

3.4.3 Hardware and software specifications

All simulations were performed on the Dantzig Cluster at the Center of Operations Research (CIO) of the Universidad Miguel Hernández[5]. Specifically, the computational resources included a Supermicro SYS-120GQ-TNRT node equipped with two Intel® Xeon® Silver 4316 CPUs operating at 2.30GHz, offering a total of 80 cores and 768GB of RAM. The simulations were run under the Rocky Linux 8.7 operating system.

The proposed methodologies—ACES, RF-ACES, and Q-ACES—as well as DEA, BDEA, and CEAT benchmarks, were implemented in `R` (version 4.2.3). LP models related to the proposed methods were solved using the `Rglpk` package (Theussl, 2009). All source code is publicly available on GitHub[6] and will be released as a package on CRAN upon publication. The CEAT models were implemented using the `eat` package (Esteve et al., 2022). StoNED estimations were carried out in `Python` using the `pystoned` library (Dai et al., 2024), with models solved locally using the default `MOSEK` solver (Mosek and Copenhagen, 2021). Finally, BDEA estimations were conducted using the `Benchmarking` R package (Bogetoft et al., 2015).

[5] https://cio.umh.es/en
[6] https://github.com/Victor-Espana/aces

Chapter 4

An adaptation of Multivariate Adaptive Regression Splines to estimate production functions

This chapter presents a novel methodology for estimating (single-output) production functions that comply with fundamental microeconomic postulates, such as monotonicity and concavity. The approach builds upon an additive variant of the Multivariate Adaptive Regression Splines (MARS) algorithm by Friedman (1991), adapted to construct a piecewise linear estimator that envelops the data. Unlike traditional nonparametric estimators like Data Envelopment Analysis (DEA) by Banker et al. (1984), which often suffer from overfitting due to their strict adherence to the minimal extrapolation principle, the proposed method mitigates this issue by incorporating a pruning procedure based on generalized cross-validation (GCV). The technique is framed as a constrained regression problem and allows for an estimation of the production frontier while preserving the key theoretical axioms. The performance of the estimator is evaluated through a series of simulation experiments and compared against two benchmark methods: DEA and Corrected Concave Nonparametric Least Squares (CC-NLS) (Kuosmanen and Johnson, 2010), both widely used in the literature. This chapter formalizes the methodology, outlines the necessary algorithmic modifications to MARS, and assesses its performance in the context of Experimental Design 1 (see Section 3.4.1). The content of this chapter is mainly based on the work published in España et al. (2024).

Estimating production functions can be framed as a regression problem subject to shape constraints, where the objective is to uncover the underlying relationship between a single output y and a set of input variables x_j, $j = 1, \ldots, m$, at its upper bound. The data generating process (DGP) is defined as:

$$y = f(x_1, \ldots, x_m) + \upsilon - u, \tag{4.1}$$

where $f(x_1, \ldots, x_m)$ represents the joint predictive relationship between the single output y and the set of inputs $(x_1, \ldots, x_m)$, while υ and u capture random noise and technical inefficiency, respectively. Throughout this approach, we adopt a deterministic framework by assuming $\upsilon = 0$ and $u \geq 0$, where $u = 0$ indicates that the evaluated Decision Making Unit (DMU) is technically efficient.

Under this setting, the production technology described in Equation (3.8) for the single-output case ($s = 1$) is expressed as:

$$\varphi = \{(\boldsymbol{x}, y) \in \mathbb{R}_+^{m+1} : y \leq f(\boldsymbol{x})\}, \tag{4.2}$$

where the production function f is required to satisfy two classical axioms from production theory: (A1) monotonicity, such that if $\boldsymbol{x} \leq \boldsymbol{x}'$, then $f(\boldsymbol{x}) \leq f(\boldsymbol{x}')$; and (A2) concavity. Finally, f must envelop the data from above, satisfying the previous axioms.

In this context, the estimator $\hat{f}$ must serve as a valid approximation of the true production function f, while satisfying the shape constraints imposed by axioms (A1) and (A2). We construct $\hat{f}$ as a linear combination of hinge-based basis functions (BFs) generated via MARS framework (see Equation (3.34)). This choice yields a piecewise linear and continuous function that can flexibly adapt to the data. However, the original algorithm proposed by Friedman (1991) does not enforce any shape restrictions and is designed to estimate mean behaviors. Therefore, to ensure that $\hat{f}$ behaves as a proper production function—monotonic, concave, and enveloping—we introduce specific adaptations to both the forward selection and backward pruning stages of the MARS procedure.

The next two subsections introduce a preliminary framework for adapting the forward and backward procedures of the additive MARS algorithm to the context of efficiency analysis.

4.1 The new adapted forward algorithm

To begin, we identify and describe the two key components that require modification in order for the standard additive version of the MARS model to be effectively applied within the framework of production function estimation.

1. Limit the maximum degree of the BFs in Equation (3.34) to generate a purely additive MARS model. This means that the proposed technique only allows the creation of univariate BF of degree $q = 1$.

2. Reformulate the Quadratic Programming (QP) model defined in Equation (3.36) to estimate a function that envelops the data from above and satisfies the monotonicity and concavity conditions.

We begin with Point 1. The satisfaction of the monotonicity (A1) and concavity (A2) axioms can be ensured by restricting the maximum degree in the construction of the BF set to one. This eliminates variable interactions, making the parent term in Equation (3.35) always $B_0^{(0)}(\boldsymbol{x})$. While this may limit predictive capacity in some scenarios, it offers significant advantages in terms of computational efficiency. It is worth noting that the computationally intensive part of the code is the coefficient estimation, which is formulated as the minimization problem in Equation (3.36). Specifically, the total computation time is mainly related to the sample size (n), the number of available predictors (m), and the interaction degree among variables ($q_{\max}$). Therefore, by constraining the BFs degree to $q_{\max} = 1$, computational costs are significantly reduced. Additionally, another advantage arises from this restriction: the additive model estimator is piecewise linear, which allows a direct comparison with the DEA estimator. In fact, the new model—after the backward phase—could be interpreted as a pruned version of DEA that overcomes its overfitting issues.

Regarding Point 2, the specific adaptations made to the additive MARS model to satisfy axioms A1 and A2 are described progressively throughout the text. Figure 3.3 shows how MARS was not originally designed by Friedman (1991) to estimate production frontiers in microeconomic settings. Clearly, some commonly accepted techniques in regression analysis—such as Ordinary Least Squares—are not applicable here, as they are oriented towards estimating conditional averages. As an alternative, we propose reformulating the linear optimization problem by introducing additional shape constraints to capture the estimation of upper envelopes rather than mean trends. In doing so, the imposed monotonicity (A1) and concavity (A2) constraints enable a natural adaptation of the additive MARS version to the production analysis field.

The approximation function of the additive MARS model in the context of production analysis, after the forward selection phase, can be expressed as:

$$\hat{f}(\boldsymbol{x}) \;=\; \tau_0 \;+\; \sum_{j=1}^{m} \sum_{p \in P_j} \Big[\gamma_{j_p}^{+} \left(x_j - \tilde{\kappa}_{j_p} \right)_+ + \gamma_{j_p}^{-} \left(\tilde{\kappa}_{j_p} - x_j \right)_+ \Big], \tag{4.3}$$

where x_j is the j-th input, for $j = 1, \ldots, m$, τ_0 is the intercept term, and $\gamma_{j_p}^{+}$ and $\gamma_{j_p}^{-}$ are the slope coefficients of the right- and left-sided hinge functions of the p-th reflected pair for input j. The knot location $\tilde{\kappa}_{j_p} \in \{x_{1j}, x_{2j}, \ldots, x_{nj}\}$ defines the split point for the p-th reflected pair corresponding to input j. The index family $P = \{P_j\}_{j=1}^{m}$ contains, for each input j, the set of reflected pairs selected during the forward stage.

As explained in Section 3.2.1, the forward selection algorithm iteratively incorporates, at each step, the BF from the candidate set defined in Equation (3.33) that yields the greatest improvement in model fit. Therefore, by collecting all splines selected during this phase, the set of BF that form the additive model is given by:

$$\mathcal{B} = \left\{1,\ (x_j - \tilde{\kappa}_{j_p})_+,\ (\tilde{\kappa}_{j_p} - x_j)_+\right\}, \quad j = 1, \ldots, m,\ \forall p \in P_j. \tag{4.4}$$

It is important to emphasize that we use the symbol k to denote any candidate knot considered for generating a reflected pair in the set (3.33), while the notation $\tilde{\kappa}$ is reserved for those specific knot locations that are actually selected and incorporated into the model.

We now detail the requirements that must be satisfied to ensure compliance with the conditions outlined in the previous point. The first property to be enforced relates to the enveloping nature of the production function $\hat{f}$ estimated by the proposed approach. This condition implies that, given any observed input–output pair $(\boldsymbol{x}_i, y_i)$, the estimate $\hat{f}(\boldsymbol{x}_i)$ must lie above the observed output y_i. Mathematically, this is expressed as $y_i \leq \hat{f}(\boldsymbol{x}_i)$ for all $i = 1, \ldots, n$. It is therefore natural to require the estimator to satisfy this inequality across all training samples.

To implement this constraint, we reformulate the estimation as a linear optimization problem as follows:

$$\min_{\boldsymbol{\varepsilon},\, \tau_0,\, \boldsymbol{\gamma}} \sum_{i=1}^{n} \varepsilon_i \tag{4.5}$$

subject to:

$$\tau_0 + \sum_{j=1}^{m} \sum_{p \in P_j} \left[\gamma_{j_p}^{+} \left(x_{ij} - \tilde{\kappa}_{j_p}\right)_+ + \gamma_{j_p}^{-} \left(\tilde{\kappa}_{j_p} - x_{ij}\right)_+\right] - \varepsilon_i = y_i, \quad i = 1, \ldots, n, \tag{4.6}$$

$$\varepsilon_i \geq 0, \quad i = 1, \ldots, n, \tag{4.7}$$

where the new non-negative variable $\varepsilon_i = \tau_0 + \sum_{j=1}^{m} \sum_{p \in P_j} \left[\gamma_{j_p}^{+} \left(x_{ij} - \tilde{\kappa}_{j_p}\right)_+ + \gamma_{j_p}^{-} \left(\tilde{\kappa}_{j_p} - x_{ij}\right)_+\right] - y_i$ captures the deviation between the estimated frontier and the observed output. Due to the non-negativity constraint, there is no compensation between overestimations and underestimations, which is consistent with the enveloping property of the production frontier. As a result, a Linear Programming (LP) formulation is adopted in place of a QP model, leading to reduced computational complexity in the estimation process.

This new estimator is, in some sense, equivalent to the additive forward MARS model, but it now estimates a frontier that envelops the data cloud instead of modeling its average behavior. This shift is achieved through the incorporation of Constraints (4.6) and (4.7), which ensure that all observations lie below the estimated surface.

However, the model may still produce estimators that violate monotonicity, thereby failing to satisfy Axiom A1. Similarly, the concavity of the estimated function is not guaranteed, which contradicts the theoretical requirement stated in Axiom A2. These issues are illustrated in Figure 4.1, where an extremely overfitted surface—resulting from an unbounded forward expansion of BFs—is shown to satisfy the enveloping property, yet lacks the monotonicity and concavity needed to ensure a convex technology set.

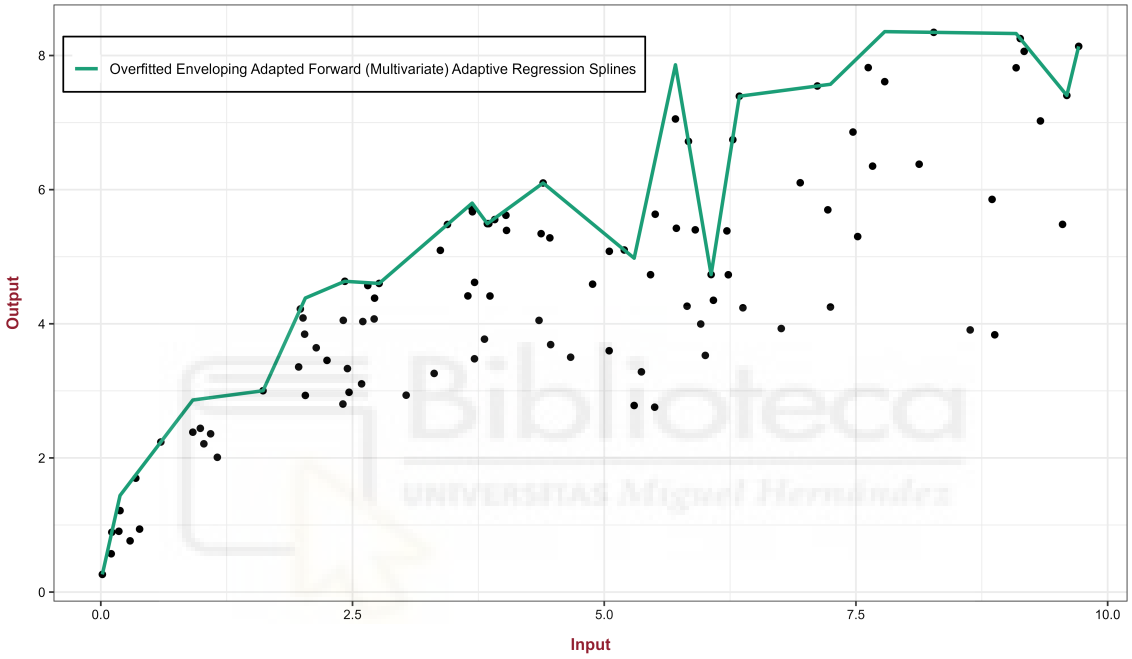

Figure 4.1: Estimator obtained by Model (4.5)

.

The rationale behind satisfying monotonicity (A1) and concavity (A2) in the estimation of production functions using the proposed approach is conceptually straightforward. The sum of non-decreasing monotonic functions yields a non-decreasing monotonic function, and likewise, the sum of concave functions yields a concave function. Therefore, our strategy consists of treating each reflected pair in Equation (4.3) independently, ensuring that both properties hold. This ensures that the resulting estimator $\hat{f}$ inherits monotonicity and concavity from the additive combination of monotonic and concave components.

We next establish sufficient conditions to enforce monotonicity and concavity in the estimator defined in Equation (4.3) under the proposed framework:

Proposition 1.1. [Monotonicity] Let

$$\hat{f}(\boldsymbol{x}) = \tau_0 + \sum_{j=1}^{m} \sum_{p \in P_j} \left[\gamma_{j_p}^{+} (x_j - \tilde{\kappa}_{j_p})_+ + \gamma_{j_p}^{-} (\tilde{\kappa}_{j_p} - x_j)_+ \right]$$

be the estimator obtained at any stage of the forward selection phase. If $\gamma_{j_p}^{+} \geq 0$ and $\gamma_{j_p}^{-} \leq 0$ for all $j = 1, \ldots, m$ and all $p \in P_j$, then $\hat{f}(\boldsymbol{x})$ is monotonically non-decreasing with respect to each input x_j.

Proof. Fix an arbitrary input index j and a reflected pair $p \in P_j$. The contribution of this pair to the estimator is

$$\gamma_{j_p}^{+} (x_j - \tilde{\kappa}_{j_p})_+ + \gamma_{j_p}^{-} (\tilde{\kappa}_{j_p} - x_j)_+,$$

a continuous, piecewise linear function with a kink at $x_j = \tilde{\kappa}_{j_p}$. We consider two cases:

Case 1: $x_j \leq \tilde{\kappa}_{j_p}$. In this region, we have:

$$(x_j - \tilde{\kappa}_{j_p})_+ = 0, \qquad (\tilde{\kappa}_{j_p} - x_j)_+ = \tilde{\kappa}_{j_p} - x_j.$$

Thus, the contribution simplifies to:

$$\gamma_{j_p}^{-} (\tilde{\kappa}_{j_p} - x_j),$$

whose derivative with respect to x_j is $-\gamma_{j_p}^{-}$. Since $\gamma_{j_p}^{-} \leq 0$ by assumption, it follows that $-\gamma_{j_p}^{-} \geq 0$, confirming non-decreasing behavior in this region.

Case 2: $x_j \geq \tilde{\kappa}_{j_p}$. In this case, we obtain:

$$(\tilde{\kappa}_{j_p} - x_j)_+ = 0, \qquad (x_j - \tilde{\kappa}_{j_p})_+ = x_j - \tilde{\kappa}_{j_p},$$

so the contribution becomes:

$$\gamma_{j_p}^{+} (x_j - \tilde{\kappa}_{j_p}),$$

with derivative $\gamma_{j_p}^{+} \geq 0$, again ensuring non-decreasing behavior.

Since each reflected pair is non-decreasing in x_j, and $\hat{f}(\boldsymbol{x})$ is the sum of such components across all inputs, the overall estimator is non-decreasing in every input coordinate. □

Proposition 1.2. [Concavity] Let

$$\hat{f}(\boldsymbol{x}) = \tau_0 + \sum_{j=1}^{m} \sum_{p \in P_j} \left[\gamma_{j_p}^{+} (x_j - \tilde{\kappa}_{j_p})_+ + \gamma_{j_p}^{-} (\tilde{\kappa}_{j_p} - x_j)_+ \right]$$

be the estimator obtained at any stage of the forward selection phase. If $\gamma_{j_p}^{+} + \gamma_{j_p}^{-} \leq 0$ for all $j = 1, \ldots, m$ and $p \in P_j$, then $\hat{f}(\boldsymbol{x})$ is concave.

Proof. Fix an arbitrary input index j and a reflected pair $p \in P_j$. The contribution of this pair to the estimator can be rewritten as:

$$\gamma_{j_p}^{+}(x_j - \tilde{\kappa}_{j_p})_+ + \gamma_{j_p}^{-}(\tilde{\kappa}_{j_p} - x_j)_+ = (\gamma_{j_p}^{+} + \gamma_{j_p}^{-})(x_j - \tilde{\kappa}_{j_p})_+ - \gamma_{j_p}^{-} x_j + \gamma_{j_p}^{-}\tilde{\kappa}_{j_p},$$

where we use the identity $(\tilde{\kappa}_{j_p} - x_j)_+ = (x_j - \tilde{\kappa}_{j_p})_+ - (x_j - \tilde{\kappa}_{j_p})$.

The function $(x_j - \tilde{\kappa}_{j_p})_+$ is convex, x_j is linear, and constants are trivially both convex and concave. Therefore, when $\gamma_{j_p}^{+} + \gamma_{j_p}^{-} \leq 0$, the coefficient of the convex component is non-positive, which ensures the overall concavity of the expression. Since $\hat{f}(\boldsymbol{x})$ is a sum of concave functions, it is itself concave. □

By incorporating the three constraints derived from Propositions 1 and 2 into the model, the LP model to solve during the forward selection phase becomes:

$$\min_{\boldsymbol{\varepsilon},\boldsymbol{\gamma},\tau_0} \sum_{i=1}^{n} \varepsilon_i \tag{4.8}$$

subject to:

$$\tau_0 + \sum_{j=1}^{m} \sum_{p \in P_j} \left[\gamma_{j_p}^{+}(x_{ij} - \tilde{\kappa}_{j_p})_+ + \gamma_{j_p}^{-}(\tilde{\kappa}_{j_p} - x_{ij})_+\right] - \varepsilon_i = y_i, \quad i = 1, \ldots, n, \tag{4.9}$$

$$\varepsilon_i \geq 0, \quad i = 1, \ldots, n, \tag{4.10}$$

$$\gamma_{j_p}^{+} + \gamma_{j_p}^{-} \leq 0, \quad j = 1, \ldots, m, \ \forall p \in P_j, \tag{4.11}$$

$$\gamma_{j_p}^{+} \geq 0, \quad j = 1, \ldots, m, \ \forall p \in P_j, \tag{4.12}$$

$$\gamma_{j_p}^{-} \leq 0, \quad j = 1, \ldots, m, \ \forall p \in P_j. \tag{4.13}$$

Constraints (4.9) and (4.10) correspond to the envelopment formulation introduced in Model (4.5). The additional Constraints (4.11), (4.12), and (4.13) are incorporated to ensure that the resulting estimator satisfies monotonicity and concavity. As a result, the estimated function complies with the fundamental microeconomic conditions required to represent a valid production function.

It is important to distinguish between the loss function used to estimate the coefficients in Model (4.8) and the lack-of-fit (LOF) criterion applied to evaluate model performance. Coefficients are always estimated by minimizing the sum of non-negative residuals to preserve computational efficiency inherent to the LP framework. However, alternative LOF metrics, such as the sum of squared residuals, can be employed to guide model selection. While the sum of non-negative residuals is the only metric guaranteed to be minimized during estimation, introducing alternative metrics in the selection phase can improve model fit in specific contexts.

To conclude the forward phase, the algorithm stops when either (i) a user-defined maximum number of BFs is reached (η), or (ii) the marginal improvement falls below a threshold (ξ). Typically, no limit is imposed on the number of terms, and ξ is set between 0% and 2%, where approaching 0% may increase computational time without significant gains in accuracy. Importantly, the forward phase is intentionally designed to overfit the model to the observed data by constructing a rich set of pre-selected BFs. This ensures that potentially relevant terms are not prematurely excluded. A more parsimonious structure is then recovered during the backward phase, where redundant or suboptimal terms are systematically pruned. Prematurely limiting model complexity during the forward pass may result in an underfitted model, compromising overall performance. In line with this philosophy, Friedman (1991) recommends pruning approximately half of the terms selected during the forward step, striking a balance between flexibility and generalization.

Up to this point, the proposed model produces results that closely resemble those of DEA. Consequently, the new technique remains vulnerable to overfitting. Despite its strong in-sample performance, the estimator exhibits high variance and limited generalization capacity, primarily due to its tight adherence to the observed sample. This behavior reflects a classic low-bias, high-variance tradeoff and is consistent with the limitations commonly associated with traditional DEA methods. In fact, Figure 4.2 illustrates that the constrained forward model and DEA produce almost identical estimators:

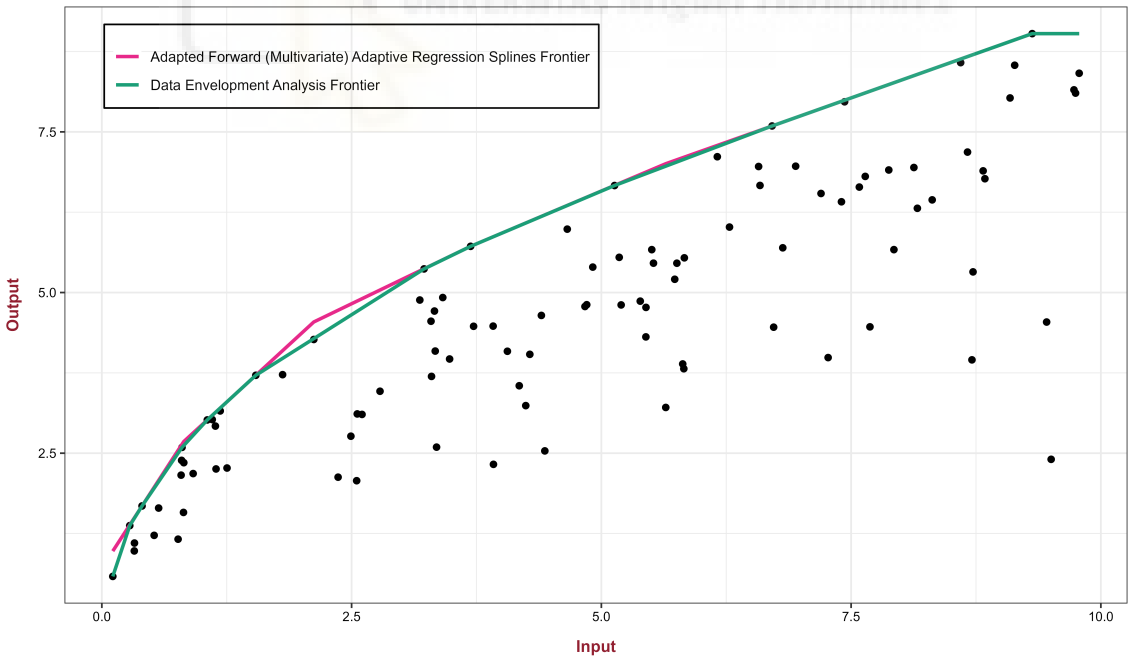

Figure 4.2: Comparison between DEA and the forward estimator of the adapted MARS approach.

Example 4.1. *Consider a dataset with three input variables denoted as x_1, x_2 and x_3, along with a single output variable. Suppose the following additive model is obtained through the forward phase of the adapted MARS algorithm:*

$$\hat{f}(\boldsymbol{x}) = 5 + 2(x_1 - 5)_+ - 3(5 - x_1)_+ + 4(x_1 - 7)_+ - 8(7 - x_1)_+ + 3(x_2 - 3)_+ - 5(3 - x_2)_+.$$

This representation involves three reflected pairs defined on coordinates x_1 and x_2, and no contribution from x_3.

First, the index family is given by

$$P = \{P_1, P_2, P_3\}, \quad P_1 = \{1, 2\},\ P_2 = \{1\},\ P_3 = \emptyset,$$

The coefficients for the right-sided BFs are given by

$$\gamma^+ = \{\gamma_1^+, \gamma_2^+, \gamma_3^+\}, \quad \gamma_1^+ = \{2, 4\},\ \gamma_2^+ = \{3\},\ \gamma_3^+ = \emptyset,$$

while left-sided counterparts are

$$\gamma^- = \{\gamma_1^-, \gamma_2^-, \gamma_3^-\}, \quad \gamma_1^- = \{-3, -8\},\ \gamma_2^- = \{-5\},\ \gamma_3^- = \emptyset.$$

The knot set is defined as

$$\tilde{\kappa} = \{\tilde{\kappa}_1, \tilde{\kappa}_2, \tilde{\kappa}_3\}, \quad \tilde{\kappa}_1 = \{5, 7\},\ \tilde{\kappa}_2 = \{3\},\ \tilde{\kappa}_3 = \emptyset.$$

Lastly, the set of BFs retained after the forward phase is:

$$\mathcal{B} = \{1, (x_1 - 5)_+,\ (5 - x_1)_+,\ (x_1 - 7)_+,\ (7 - x_1)_+,\ (x_2 - 3)_+,\ (3 - x_2)_+\ \},$$

with $|\mathcal{B}| = 7$.

As a summary of the methodology described above, the steps required to implement the forward procedure for estimating the production frontier using the proposed additive MARS-based approach are systematically presented in Algorithm 1. This procedure iteratively builds the model by selecting and incorporating pairs of BFs that contribute to reducing the LOF, while adhering to the shape constraints imposed by production theory. The resulting representation serves as an initial, overparameterized approximation that will later be refined through a backward pruning stage.

Algorithm 1: Forward procedure of the adapted MARS-based approach

Input:

η: maximum number of BF to include into the model ;

ξ: minimum relative reduction in training error required to proceed to the next iteration.

Output:

$\hat{f}$: estimator defined in Equation (4.3) ;

$\mathcal{B}$: set of selected BF ;

$\tilde{\boldsymbol{\kappa}}$: set of selected knot locations.

Initialization

  $\text{LOF}_{\min} \leftarrow \infty$;

  $P_j \leftarrow \emptyset$, for $j = 1, \ldots, m$; $\tilde{\kappa}_j \leftarrow \emptyset$, for $j = 1, \ldots, m$; $\mathcal{B} \leftarrow \{1\}$;

while $|\mathcal{B}| < \eta$ **do**

  Store previous best

  $\text{LOF}^{\circ}_{\min} \leftarrow \text{LOF}_{\min}$;

  $P^{\circ} \leftarrow P$; $\tilde{\kappa}^{\circ} \leftarrow \tilde{\kappa}$; $\mathcal{B}^{\circ} \leftarrow \mathcal{B}$;

  for $j \leftarrow 1$ **to** m **do**

    for $i \leftarrow 1$ **to** n **do**

      Select a candidate knot location: $\tilde{\kappa}_{j_{|\tilde{\kappa}_j|+1}} = x_{ij}$;

      Form a candidate reflected pair: $\left\{(x_j - \tilde{\kappa}_{j_{|\tilde{\kappa}_j|+1}})_+, (\tilde{\kappa}_{j_{|\tilde{\kappa}_j|+1}} - x_j)_+\right\}$;

      Update the index set: $P'_j \leftarrow P_j \cup \{|P_j| + 1\}$

      Update the knot set: $\tilde{\kappa}'_j \leftarrow \tilde{\kappa}_j \cup \{\tilde{\kappa}_{j_{|\tilde{\kappa}_j|+1}}\}$;

      Add the new pair to the basis: $\mathcal{B}' \leftarrow \mathcal{B} \cup \left\{(x_j - \tilde{\kappa}_{j_{|\tilde{\kappa}_j|+1}})_+, (\tilde{\kappa}_{j_{|\tilde{\kappa}_j|+1}} - x_j)_+\right\}$;

      Estimate coefficients using the model defined in Equation (4.8) ;

      Compute the user-specified LOF criterion ;

      if $LOF < LOF^{\circ}_{\min}$ **then**

        $\text{LOF}^{\circ}_{\min} \leftarrow \text{LOF}$;

        $P^{\circ} \leftarrow P'$, $\tilde{\kappa}^{\circ} \leftarrow \tilde{\kappa}'$, $\mathcal{B}^{\circ} \leftarrow \mathcal{B}'$

  Update if improvement is sufficient

  if $LOF_{\min} < LOF^{\circ}_{\min} \cdot (1 - \xi)$ **then**

    $\text{LOF}_{\min} \leftarrow \text{LOF}^{\circ}_{\min}$;

    $P \leftarrow P^{\circ}$, $\tilde{\kappa} \leftarrow \tilde{\kappa}^{\circ}$, $\mathcal{B} \leftarrow \mathcal{B}^{\circ}$;

  else

    break

4.2 The new adapted backward algorithm

Overfitting is a critical issue that undermines the reliability of any statistical model. As in the original MARS algorithm, the proposed methodology heavily relies on the response variable to define the set of BFs. While this often reduces model bias, it simultaneously increases variance. This behavior can lead the model to "memorize" the training data, resulting in poor generalization to new, unseen samples. This is a well-known limitation in many machine learning (ML) algorithms. Fortunately, most of these algorithms incorporate specific mechanisms to enhance generalization performance. In our case, we propose integrating the forward procedure with a pruning strategy that enables a more accurate evaluation of production functions. To this end, we adopt the standard backward phase proposed by Friedman (1991) in the MARS framework—originally based on GCV—and introduce targeted modifications to ensure compliance with the theoretical requirements of frontier analysis.

The forward phase concludes with the construction of an overparameterized model consisting of a collection of reflected pairs, in addition to the constant function $B_0^{(0)}(\boldsymbol{x})$ (intercept). Due to the tendency to overfit (see Figure 4.2), a backward elimination process is initiated, where BFs that do not significantly contribute to predictive performance are removed. The backward procedure seeks to balance model complexity and accuracy. This process intentionally breaks the pairing structure of the BFs created in the forward stage. Consequently, some functions may be retained in the model, while others are fully or partially discarded. As a result, each BF can end up in one of three possible configurations: as part of a preserved pair, as an unpaired left-sided BF, or as an unpaired right-sided BF. The full model in Equation (4.3) is then redefined accordingly:

$$\begin{aligned}\hat{f}(\boldsymbol{x}) = \tau_0 + \sum_{j=1}^{m}\sum_{p\in P_j}\left[\gamma_{j_p}^{+}\left(x_j - \tilde{\kappa}_{j_p}\right)_+ + \gamma_{j_p}^{-}\left(\tilde{\kappa}_{j_p} - x_j\right)_+\right] + \sum_{j=1}^{m}\sum_{h\in H_j}\left[\delta_{j_h}^{+}\left(x_j - \tilde{\kappa}_{j_h}^{(\mathrm{R})}\right)_+\right] \\ + \sum_{j=1}^{m}\sum_{w\in W_j}\left[\delta_{j_w}^{-}\left(\tilde{\kappa}_{j_w}^{(\mathrm{L})} - x_j\right)_+\right].\end{aligned} \tag{4.14}$$

Equation (4.14) extends the additive MARS frontier estimator by incorporating unpaired BFs in addition to the reflected pairs already defined in Equation (4.3). The sets $P = \{P_j\}_{j=1}^m$, $\gamma^+ = \{\gamma_j^+\}_{j=1}^m$, $\gamma^- = \{\gamma_j^-\}_{j=1}^m$, and $\tilde{\kappa} = \{\tilde{\kappa}_j\}_{j=1}^m$ retain their definitions and refer to the index sets, coefficients, and knot locations associated with the reflected pairs built during the forward stage. In addition, let H_j denote the subset of indices corresponding to unpaired right-sided BF for input j, so that $H = \{H_j\}_{j=1}^m$. Each index $h \in H_j$ identifies a function of the form $(x_j - \tilde{\kappa}_{j_h}^{(\mathrm{R})})_+$, where $\tilde{\kappa}_{j_h}^{(\mathrm{R})} \in \{x_{1j}, x_{2j}, \ldots, x_{nj}\}$ is the associated knot location, and $\delta_{j_h}^+$ is its corresponding coefficient. Similarly, the sets W_j, grouped into the family $W = \{W_j\}_{j=1}^m$, index the unpaired left-sided BF $(\tilde{\kappa}_{j_w}^{(\mathrm{L})} - x_j)_+$, each defined by its knot $\tilde{\kappa}_{j_w}^{(\mathrm{L})} \in \{x_{1j}, x_{2j}, \ldots, x_{nj}\}$ and coefficient $\delta_{j_w}^-$.

We now establish the sufficient conditions under which the extended estimator defined in Equation (4.14) satisfies the monotonicity and concavity constraints. The proofs follow the same reasoning previously applied, relying on the classical result that the sum of non-decreasing and concave functions is also non-decreasing and concave. This ensures that the final estimator remains consistent with the axiomatic framework of production theory.

Proposition 1.3. [Monotonicity and concavity] Let

$$\hat{f}(\boldsymbol{x}) = \tau_0 + \sum_{j=1}^{m} \sum_{p \in P_j} \left[\gamma_{j_p}^{+} \left(x_j - \tilde{\kappa}_{j_p} \right)_+ + \gamma_{j_p}^{-} \left(\tilde{\kappa}_{j_p} - x_j \right)_+ \right] + \sum_{j=1}^{m} \sum_{h \in H_j} \left[\delta_{j_h}^{+} \left(x_j - \tilde{\kappa}_{j_h}^{(\mathrm{R})} \right)_+ \right] + \sum_{j=1}^{m} \sum_{w \in W_j} \left[\delta_{j_w}^{-} \left(\tilde{\kappa}_{j_w}^{(\mathrm{L})} - x_j \right)_+ \right].$$

Assume for every j, p

$$\gamma_{j_p}^{+} + \gamma_{j_p}^{-} \leq 0, \quad \gamma_{j_p}^{+} \geq 0, \quad \gamma_{j_p}^{-} \leq 0,$$

and for every j, w

$$\delta_{j_w}^{-} \leq 0.$$

Moreover, to keep both non-decreasingness and concavity, we require

$$H = \emptyset \quad \text{(no unpaired right-sided BF)}.$$

Then $\hat{f}(\boldsymbol{x})$ is coordinate-wise non-decreasing and concave.

Proof. By Proposition 1.1., the contribution of each reflected pair $\gamma_{j_p}^{+}(x_j - \tilde{\kappa}_{j_p})_+ + \gamma_{j_p}^{-}(\tilde{\kappa}_{j_p} - x_j)_+$ is coordinate-wise non-decreasing provided that $\gamma_{j_p}^{+} \geq 0$ and $\gamma_{j_p}^{-} \leq 0$, and is concave whenever $\gamma_{j_p}^{+} + \gamma_{j_p}^{-} \leq 0$. These conditions are assumed here.

We now turn to the unpaired BF. Each left-sided unpaired BF has the form $\delta_{j_w}^{-} \left(\tilde{\kappa}_{j_w}^{(\mathrm{L})} - x_j \right)_+$, where the hinge function $\left(\tilde{\kappa}_{j_w}^{(\mathrm{L})} - x_j \right)_+$ is convex and non-increasing in j. Thus, multiplying it by a non-positive coefficient $\delta_{j_w}^{-} \leq 0$ yields a non-decreasing and concave function, preserving the desired shape constraints. In contrast, right-sided unpaired BFs of the form $\delta_{j_h}^{+}(x_j - \tilde{\kappa}_{j_h}^{(\mathrm{R})})_+$ are convex and increasing. Even when $\delta_{j_h}^{+} \geq 0$, the resulting function is also increasing but fails to be concave. To ensure the overall concavity of the estimator, we therefore assume the absence of unpaired right-sided BFs, i.e., $H = \emptyset$.

Under these conditions, all components of the estimator are both non-decreasing and concave. Since the sum of such functions preserves these properties, the full estimator $\hat{f}(\boldsymbol{x})$ is coordinate-wise non-decreasing and concave. □

Following Proposition 1.3., the need arises to modify the backward procedure described in Friedman (1991). The adapted backward algorithm remains identical to the original procedure detailed in Section 3.2.2, except for the criteria used to determine which BF should be removed at each iteration.

In the standard MARS algorithm, any BF is treated as a potential candidate for removal. However, under the proposed approach, two specific conditions must be satisfied before a BF can be considered eligible for elimination, in order to ensure that no unpaired right-sided BFs are retained ($H = \emptyset$):

1. Right-sided BFs may only be removed if they belong to a reflected pair.

2. Left-sided BFs can be removed only if they appear unpaired.

The LP model solved at each iteration of the backward procedure is then formulated as follows:

$$\min_{\boldsymbol{\varepsilon},\boldsymbol{\gamma},\boldsymbol{\delta}} \sum_{i=1}^{n} \varepsilon_i \tag{4.15}$$

subject to:

$$\begin{aligned} \hat{f}(\boldsymbol{x}_i) - \varepsilon_i &= y_i, \quad i = 1,\dots,n, \\ \varepsilon_i &\geq 0, \quad i = 1,\dots,n, \\ \gamma_{j_p}^{+} + \gamma_{j_p}^{-} &\leq 0, \quad \forall p \in P_j,\ j = 1,\dots,m, \\ \gamma_{j_p}^{+} &\geq 0, \quad \forall p \in P_j,\ j = 1,\dots,m, \\ \gamma_{j_p}^{-} &\leq 0, \quad \forall p \in P_j,\ j = 1,\dots,m, \\ \delta_{j_w}^{-} &\leq 0, \quad \forall w \in W_j,\ j = 1,\dots,m, \end{aligned} \tag{4.16}$$

$$\tag{4.17}$$

where:

$$\hat{f}(\boldsymbol{x}_i) = \tau_0 + \sum_{j=1}^{m} \sum_{p \in P_j} \left[\gamma_{j_p}^{+} (x_{ij} - \tilde{\kappa}_{j_p})_+ + \gamma_{j_p}^{-} (\tilde{\kappa}_{j_p} - x_{ij})_+ \right] + \sum_{j=1}^{m} \sum_{w \in W_j} \left[\delta_{j_w}^{-} \left(\tilde{\kappa}_{j_w}^{(\mathrm{L})} - x_{ij} \right)_+ \right]. \tag{4.18}$$

Compared to the LP model utilized for the forward algorithm defined in Equation (4.8), the only additional constraint in (4.17) restricts the coefficients associated with unpaired left-sided BFs. This ensures that the resulting estimator also satisfies the monotonicity and concavity conditions required for a valid production function.

As a continuation of the modeling process, the steps to implement the backward procedure for refining the production frontier estimator are outlined in Algorithm 2. Starting from the overparameterized model obtained in the forward phase, this stage iteratively removes BFs that contribute least to performance while remains consistent with production theory.

Algorithm 2: Backward procedure of the adapted MARS-based approach

Input:

$\mathcal{B}$: set of BFs selected after the forward phase;

P: index family of reflected pairs associated with $\mathcal{B}$;

$\tilde{\kappa}$: set of selected knots after the forward phase;

d: complexity penalty parameter used in the GCV criterion.

Output:

$\hat{f}$: estimator defined in Equation (4.18);

Initialization

- $\mathcal{B}^{(\mathrm{P})} \subset \mathcal{B}$: set of paired BFs in $\mathcal{B}$ (excluding the intercept);
- $\mathcal{B}^{(\mathrm{PR})} \subset \mathcal{B}^{(\mathrm{P})}$: set of paired right-sided BFs;
- $\mathcal{B}^{(\mathrm{UL})} \leftarrow \emptyset$: set of unpaired left-sided BFs;
- $\mathcal{B}^{(\mathrm{CR})} \leftarrow \mathcal{B}^{(\mathrm{PR})} \cup \mathcal{B}^{(\mathrm{UL})}$: set of candidate BFs for removal;
- $W_j \leftarrow \emptyset$, for $j = 1, ..., m$: index set of unpaired left-sided BFs;
- $\mathrm{GCV}_{|\mathcal{B}|} \leftarrow \mathrm{GCV}(\mathcal{B})$ (based on Equation (3.40))

while $|\mathcal{B}| > 1$ **do**

- Store previous best
- $\mathrm{GCV}_{\min} \leftarrow \infty$ $P^{\circ} \leftarrow P$; $W^{\circ} \leftarrow W$; $\tilde{\kappa}^{\circ} \leftarrow \tilde{\kappa}$; $\mathcal{B}^{\circ} \leftarrow \mathcal{B}$;
- **foreach** $l \in \mathcal{B}^{(CR)}$ **do**
  - **if** $l \in \mathcal{B}^{(PR)}$ **then**
    - Identify the left-sided BF l' paired with l;
    - $\mathcal{B}^{(\mathrm{P})'} \leftarrow \mathcal{B}^{(\mathrm{P})} \setminus \{l, l'\}$; $\mathcal{B}^{(\mathrm{PR})'} \leftarrow \mathcal{B}^{(\mathrm{PR})} \setminus \{l\}$; $\mathcal{B}^{(\mathrm{UL})'} \leftarrow \mathcal{B}^{(\mathrm{UL})} \cup \{l'\}$;
  - **else**
    - $\mathcal{B}^{(\mathrm{UL})'} \leftarrow \mathcal{B}^{(\mathrm{UL})} \setminus \{l\}$;
  - Update full basis set:
  - $\mathcal{B}' \leftarrow \{1\} \cup \mathcal{B}^{(\mathrm{P})'} \cup \mathcal{B}^{(\mathrm{UL})'}$;
  - Update index and knot sets P', W' and $\tilde{\kappa}'$ based on new $\mathcal{B}'$;
  - Estimate model coefficients based on Equation (4.15) using basis $\mathcal{B}'$;
  - Compute $\mathrm{GCV}(\mathcal{B}')$ based on Equation (3.40) ;
  - **if** $GCV(\mathcal{B}') < GCV_{min}$ **then**
    - $\mathrm{GCV}_{\min} \leftarrow \mathrm{GCV}(\mathcal{B}')$;
    - $\mathcal{B}^{\circ} \leftarrow \mathcal{B}'$ $P^{\circ} \leftarrow P'$; $W^{\circ} \leftarrow W'$; $\tilde{\kappa}^{\circ} \leftarrow \tilde{\kappa}'$
- Update current configuration
- $\mathcal{B}^{(\mathrm{P})} \leftarrow \mathcal{B}^{(\mathrm{P})} \cap \mathcal{B}^{\circ}$; $\mathcal{B}^{(\mathrm{PR})} \leftarrow \mathcal{B}^{(\mathrm{PR})} \cap \mathcal{B}^{\circ}$; $\mathcal{B}^{(\mathrm{UL})} \leftarrow \mathcal{B}^{(\mathrm{UL})} \cap \mathcal{B}^{\circ}$; $\mathcal{B}^{(\mathrm{CR})} \leftarrow \mathcal{B}^{(\mathrm{PR})} \cup \mathcal{B}^{(\mathrm{UL})}$;
- $\mathcal{B} \leftarrow \mathcal{B}^{\circ}$; $P \leftarrow P^{\circ}$; $W \leftarrow W^{\circ}$; $\tilde{\kappa} \leftarrow \tilde{\kappa}^{\circ}$;

Select final set of BFs: $\tilde{\mathcal{B}} = \mathcal{B}$ for which $\mathrm{GCV}(\mathcal{B})$ is the minimum GCV observed ;

Example 4.2. *As previously illustrated, we now present the formulation of an additive MARS-based model for a dataset with three inputs after the backward phase. Consider the following production function estimator:*

$$\hat{f}(\boldsymbol{x}) = 6 + 3(x_1 - 5)_+ - 5(5 - x_1)_+ - 2(3 - x_2)_+.$$

In this case, the sets P, γ^+, γ^-, and $\tilde{\kappa}$ are defined as in Equation (4.3), corresponding to the reflected pair $\{(x_1-5)_+, (5-x_1)_+\}$. The unpaired left-sided BF $(3-x_2)_+$, retained after the backward phase, is indexed as follows. The index family of unpaired left-sided BFs is $W = \{W_1, W_2, W_3\}$, where $W_1 = \emptyset$, $W_2 = \{1\}$, and $W_3 = \emptyset$. The corresponding set of coefficients is $\delta^- = \{\delta_1^-, \delta_2^-, \delta_3^-\}$, with $\delta_1^- = \emptyset$, $\delta_2^- = \{-2\}$, and $\delta_3^- = \emptyset$. The associated knot locations are defined as $\tilde{\kappa}^{(L)} = \{\tilde{\kappa}_1^{(L)}, \tilde{\kappa}_2^{(L)}, \tilde{\kappa}_3^{(L)}\}$, where $\tilde{\kappa}_1^{(L)} = \emptyset$, $\tilde{\kappa}_2^{(L)} = \{3\}$, and $\tilde{\kappa}_3^{(L)} = \emptyset$.

Note that the model defined above is a pruned submodel of the one introduced in Example 4.1, where one of the two reflected pairs has been removed, and a single unpaired left-sided BF has been retained.

Finally, Figure 4.3 illustrates the impact of the pruning procedure on the estimator obtained after the forward phase. To this end, different values of the hyperparameter d are tested. As defined in Equation (3.40), this parameter represents the cost of retaining a knot in the model. Higher values of d result in fewer selected knots and, consequently, a simpler model that is less likely to overfit. The optimal value of d can be selected via standard cross-validation (CV) procedures.

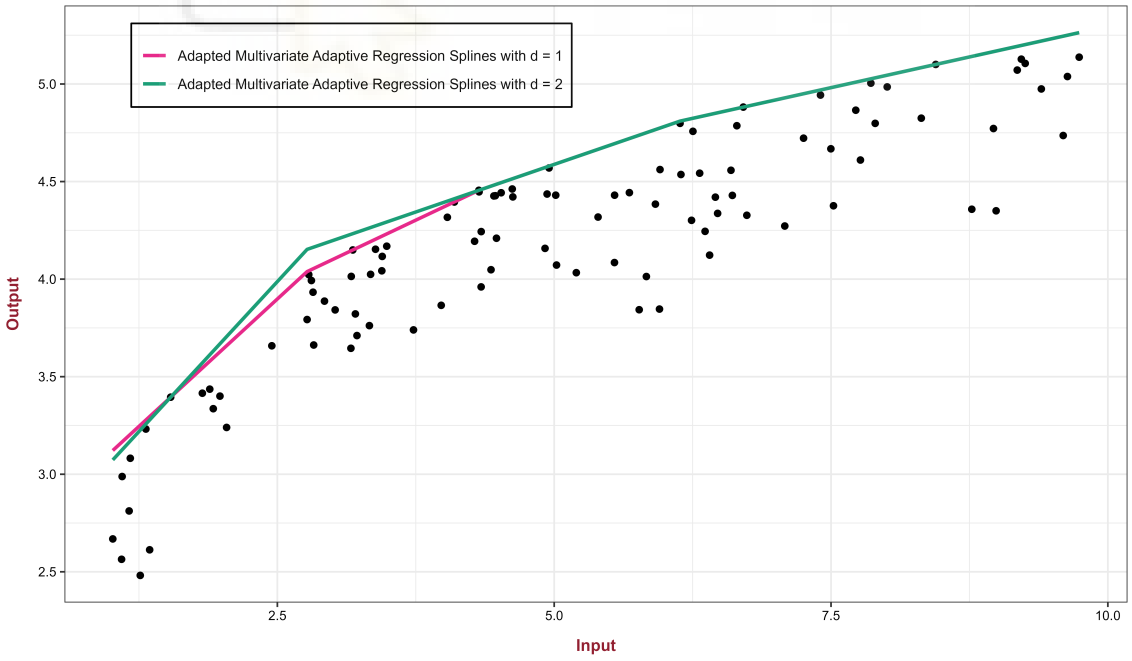

Figure 4.3: Comparison of the pruning effect under different values of hyperparameter d.

4.3 The new adapted smoothing procedure

This section introduces two alternative smoothed versions of the previously described approach: one using cubic BFs and another using quintic BFs, as described in the works of Friedman (1991) and Chen et al. (1999), respectively. These procedures enable the construction of models with continuous first-order or second-order derivatives. It is important to emphasize that this technique should only be applied when the underlying function is assumed to be smooth, which is often the case in production theory—for example, in typical specifications such as the Cobb–Douglas production function. In our particular setting, we identify a main difference compared to Friedman's original smoothing procedure: we adopt an alternative method to construct the new set of knots for the smoothed model. This aspect is discussed in detail below.

To construct the new set of knots, we begin with the collection of surviving knots obtained after the backward algorithm, denoted by $\mathcal{K} = \{K_1, \ldots, K_m\}$, as shown in Equation (3.42). Each set $K_j = \{k_{j_1}, \ldots, k_{j_{|K_j|}}\}$ contains the surviving knots (in increasing order) for input j. In the smoothing phase, each cubic or quintic BF is constructed based on a central knot $k_{j_c} \in K_j$, together with two side knots: a right-sided knot $k_{j_c}^+$ and a left-sided knot $k_{j_c}^-$. These side knots are computed as the midpoints between the central knot and its neighboring central knots:

$$k_{j_c}^+ = \frac{k_{j_c} + k_{j_{c+1}}}{2}, \qquad k_{j_c}^- = \frac{k_{j_{c-1}} + k_{j_c}}{2}, \qquad j = 1, \ldots, m, \quad c = 1, \ldots, |K_j|. \tag{4.19}$$

For boundary cases—i.e., when k_{j_c} is the first or last knot in K_j—the missing neighbor is replaced by the minimum or maximum value observed in the training data for the corresponding input dimension. As a result, each central knot is associated with a triplet $\{k_{j_c}^-,\ k_{j_c},\ k_{j_c}^+\}$, which defines the support of the smoothed BF centered at k_{j_c}.

Importantly, every central knot $k_{j_c} \in K_j$ corresponds to a knot already present in the piecewise linear model—either as part of a reflected pair or as the knot of an unpaired left-sided BF. More precisely, for each $k_{j_c} \in K_j$, either

$$\exists\, c = 1, \ldots, |K_j|,\ \exists\, p \in P_j \text{ such that } k_{j_c} = \tilde{\kappa}_{j_p} \quad \text{or} \quad \exists\, c = 1, \ldots, |K_j|,\ \exists\, w \in W_j \text{ such that } k_{j_c} = \tilde{\kappa}_{j_w}^{(\mathrm{L})}.$$

For notational convenience, let $\mathcal{T}(k) = \{k^-, k, k^+\}$ denote the triplet of knots associated with a given central knot k, where k^- and k^+ are the side knots defined as midpoints between k and its adjacent central knots. Each smoothed BF $B_\iota^{\pm}(\boldsymbol{x}, \mathcal{T}(k))$, with $\iota \in \{\mathrm{C}, \mathrm{Q}\}$, is defined on this triplet and corresponds to either a cubic ($\iota = \mathrm{C}$) or quintic ($\iota = \mathrm{Q}$) function. Using this representation, the smoothed model takes the following form:

$$\hat{f}^{\mathrm{S}}(\boldsymbol{x}) = \tau_0 + \sum_{j=1}^{m} \sum_{p \in P_j} \left[\gamma_{j_p}^{+} B_{\iota}^{+}\left(x_j, \mathcal{T}(\tilde{\kappa}_{j_p})\right) + \gamma_{j_p}^{-} B_{\iota}^{-}\left(x_j, \mathcal{T}(\tilde{\kappa}_{j_p})\right) \right] + \sum_{j=1}^{m} \sum_{w \in W_j} \delta_{j_w}^{-} B_{\iota}^{-}\left(x_j, \mathcal{T}(\tilde{\kappa}_{j_w}^{(\mathrm{L})})\right). \tag{4.20}$$

At this stage we can state four sets of sufficient conditions that guarantee the smoothed estimator in (4.20) is (i) coordinate-wise non-decreasing (r1 and r2) and (ii) concave (r3 and r4):

$$\text{r1:} \quad \gamma_{j_p}^{+} \frac{\partial B_{\iota}^{+}\left(x_j, \mathcal{T}(k_{j_c})\right)}{\partial x_j} + \gamma_{j_p}^{-} \frac{\partial B_{\iota}^{-}\left(x_j, \mathcal{T}(k_{j_c})\right)}{\partial x_j} \geq 0, \tag{4.21}$$

$$\text{r2:} \quad \delta_{j_w}^{-} \frac{\partial B_{\iota}^{-}\left(x_j, \mathcal{T}(k_{j_c})\right)}{\partial x_j} \geq 0, \tag{4.22}$$

$$\text{r3:} \quad \gamma_{j_p}^{+} \frac{\partial^2 B_{\iota}^{+}\left(x_j, \mathcal{T}(k_{j_c})\right)}{\partial x_j^2} + \gamma_{j_p}^{-} \frac{\partial^2 B_{\iota}^{-}\left(x_j, \mathcal{T}(k_{j_c})\right)}{\partial x_j^2} \leq 0, \tag{4.23}$$

$$\text{r4:} \quad \delta_{j_w}^{-} \frac{\partial^2 B_{\iota}^{-}\left(x_j, \mathcal{T}(k_{j_c})\right)}{\partial x_j^2} \leq 0, \tag{4.24}$$

for all $j = 1, \ldots, m$, $c = 1, \ldots, |K_j|$, $\iota \in \{C, Q\}$, $p \in P_j$, and $w \in W_j$.

Linear BFs—defined in Equation (3.32)—correspond exactly to the smoothed BFs—defined in Equations (3.43) and (3.44) or (3.45) and (3.46)—whenever $x_j \leq k^-$ and $x_j \geq k^+$, respectively. Therefore, it is reasonable to use the LP model defined in (4.15) to estimate the coefficient vector, modifying Constraint (4.16) to incorporate the smoothed predictor defined in Equation (4.20). However, additional constraints must be introduced to ensure the desired shape of the predictor in the interval $[k^-, k^+]$, where the smooth estimator and the original piecewise linear model do not coincide.

To this end, we must consider the placement of the side knot locations. We assume that, in the univariate case, the side knots, k^- and k^+, are lateral translations of the central knot, that is:

$$k^+ = k + \zeta^+, \quad \zeta^+ \geq 0, \qquad k^- = k - \zeta^-, \quad \zeta^- \geq 0. \tag{4.25}$$

As a result, we define the distance between side knots as:

$$k^+ - k^- = \zeta^+ + \zeta^-, \qquad (\text{or equivalently, } k^- - k^+ = -\zeta^+ - \zeta^-).$$

The main idea is that the closer the side knots are to the central knot (i.e., as $\zeta^+ \to 0$ and $\zeta^- \to 0$), the closer the smoothed predictor defined in Equation (4.20) approaches the submodel from the backward phase defined in Equation (4.18), that is,

$$\hat{f}^{\mathrm{S}}(\boldsymbol{x}) \to \hat{f}(\boldsymbol{x}).$$

Thus, we aim to determine the maximum tolerable distance between each side knot and its corresponding central knot such that the resulting smoothed predictor does not deviate from the desired shape inherited from the backward submodel.

In the following, we explore two types of smoothing procedures based on the degree of the BFs employed. Subsection 4.3.1 is devoted to the case of cubic BFs, originally proposed by Friedman (1991), and Subsection 4.3.2 analyzes their quintic counterparts, introduced by Chen et al. (1999). Each subsection presents the formal definition of the corresponding smoothed BFs, examines their first- and second-order derivatives, and derives sufficient conditions under which the resulting predictor remains non-decreasing and concave,

4.3.1 Cubic basis function smoothing

The cubic BFs introduced by Friedman (1991), reparameterized in terms of the new variables ζ^- and ζ^+, are given by:

$$B_C^+(x;k,k^+,k^-) = \begin{cases} 0, & x \le k^-, \\ \dfrac{2\zeta^+ - \zeta^-}{(\zeta^+ + \zeta^-)^2}(x-k^-)^2 + \dfrac{\zeta^- - \zeta^+}{(\zeta^+ + \zeta^-)^3}(x-k^-)^3, & k^- < x < k^+, \\ x-k, & x \ge k^+, \end{cases} \tag{4.26}$$

$$B_C^-(x;k,k^+,k^-) = \begin{cases} k-x, & x \le k^-, \\ \dfrac{2\zeta^- - \zeta^+}{(\zeta^+ + \zeta^-)^2}(x-k^+)^2 - \dfrac{\zeta^+ - \zeta^-}{(\zeta^+ + \zeta^-)^3}(x-k^+)^3, & k^- < x < k^+, \\ 0, & x \ge k^+. \end{cases} \tag{4.27}$$

First-order derivatives of previous expressions are given by:

$$\frac{\partial B_C^+(x;k,k^+,k^-)}{\partial x} = \begin{cases} 0, & x \le k^-, \\ 2 \cdot \dfrac{2\zeta^+ - \zeta^-}{(\zeta^+ + \zeta^-)^2}(x-k^-) + 3 \cdot \dfrac{\zeta^- - \zeta^+}{(\zeta^+ + \zeta^-)^3}(x-k^-)^2, & k^- < x < k^+, \\ 1, & x \ge k^+, \end{cases} \tag{4.28}$$

$$\frac{\partial B_C^-(x;k,k^+,k^-)}{\partial x} = \begin{cases} -1, & x \le k^-, \\ 2 \cdot \dfrac{2\zeta^- - \zeta^+}{(\zeta^+ + \zeta^-)^2}(x-k^+) - 3 \cdot \dfrac{\zeta^+ - \zeta^-}{(\zeta^+ + \zeta^-)^3}(x-k^+)^2, & k^- < x < k^+, \\ 0, & x \ge k^+. \end{cases} \tag{4.29}$$

In addition, the second derivatives of the previously defined expressions are given by:

$$\frac{\partial^2 B_C^+(x;k,k^+,k^-)}{\partial x^2} = \begin{cases} 0, & x \le k^-, \\ 2\cdot\dfrac{2\zeta^+-\zeta^-}{(\zeta^++\zeta^-)^2} + 6\cdot\dfrac{\zeta^--\zeta^+}{(\zeta^++\zeta^-)^3}(x-k^-), & k^- < x < k^+, \\ 0, & x \ge k^+, \end{cases} \tag{4.30}$$

$$\frac{\partial^2 B_C^-(x;k,k^+,k^-)}{\partial x^2} = \begin{cases} 0, & x \le k^-, \\ 2\cdot\dfrac{2\zeta^--\zeta^+}{(\zeta^++\zeta^-)^2} - 6\cdot\dfrac{\zeta^+-\zeta^-}{(\zeta^++\zeta^-)^3}(x-k^+), & k^- < x < k^+, \\ 0, & x \ge k^+. \end{cases} \tag{4.31}$$

We now provide sufficient conditions to ensure that the smoothed estimator complies with the fundamental shape constraints required by production theory.

Proposition 1.4. [Monotonicity for smoothed reflected pairs] Let

$$\hat{f}^{\mathrm{S}}(\boldsymbol{x}) = \tau_0 + \sum_{j=1}^{m}\sum_{p\in P_j}\left[\gamma_{j_p}^+ B_{\mathrm{C}}^+(x_j,\mathcal{T}(\tilde{\kappa}_{j_p})) + \gamma_{j_p}^- B_{\mathrm{C}}^-(x_j,\mathcal{T}(\tilde{\kappa}_{j_p}))\right] + \sum_{j=1}^{m}\sum_{w\in W_j}\delta_{j_w}^- B_{\mathrm{C}}^-(x,\mathcal{T}(\tilde{\kappa}_{j_w}^{(\mathrm{L})}))$$

be the smoothed estimator defined in Equation (4.20) using cubic BFs. Assume that $\gamma_{j_p}^+ \ge 0$, $\gamma_{j_p}^- \le 0$, and $\gamma_{j_p}^+ + \gamma_{j_p}^- \le 0$ for all $j = 1,\dots,m$ and all $p \in P_j$. If the ratio between side knots $\pi_1 = \zeta^-/\zeta^+$ satisfies $\pi_1 \in [1,2]$, then the estimator $\hat{f}^{\mathrm{S}}(\boldsymbol{x})$ is coordinate-wise non-decreasing on the interval $[k^-,k^+]$.

Proof. To prove the result, we show that the derivative of the smoothed estimator is non-negative within the interval $[k^-,k^+]$. We begin with the right-sided cubic BF. According to Equation (4.28), the first derivative of $B_C^+(x;k,k^+,k^-)$ on the interval $x \in (k^-,k^+)$ is given by:

$$\frac{\partial B_C^+(x;k,k^+,k^-)}{\partial x} = 3\cdot\frac{\zeta^--\zeta^+}{(\zeta^++\zeta^-)^3}(x-k^-)^2 + 2\cdot\frac{2\zeta^+-\zeta^-}{(\zeta^++\zeta^-)^2}(x-k^-).$$

To ensure monotonicity, we must verify that the expression remains non-negative over the entire interval, that is, $\frac{\partial B_C^+(x;k,k^+,k^-)}{\partial x} \ge 0$ for all $x \in [k^-,k^+]$. For ease of analysis, we rewrite this condition in an equivalent form by multiplying both sides of the inequality by the positive quantity $\frac{(\zeta^++\zeta^-)^2}{x-k^-}$, leading to:

$$t(x) := 3\cdot\frac{\zeta^--\zeta^+}{\zeta^++\zeta^-}(x-k^-) + 2(2\zeta^+-\zeta^-) \ge 0.$$

This is a straight line with positive slope. Its minimum is attained at $x = k^-$, where:

$$t(k^-) = 2(2\zeta^+ - \zeta^-) = 2\zeta^+(2 - \pi_1) \geq 0.$$

Thus, the condition $t(k^-) \geq 0$ holds for all $\pi_1 \in [1, 2]$.

We now consider the left-sided cubic BF. According to Equation (4.29), its first derivative over the interval $x \in (k^-, k^+)$ is given by:

$$\frac{\partial B_C^-(x; k, k^+, k^-)}{\partial x} = -3 \cdot \frac{\zeta^+ - \zeta^-}{(\zeta^+ + \zeta^-)^3}(x - k^+)^2 + 2 \cdot \frac{2\zeta^- - \zeta^+}{(\zeta^+ + \zeta^-)^2}(x - k^+).$$

To ensure non-decreasing monotonicity, we require that $\frac{\partial B_C^-(x;k,k^+,k^-)}{\partial x} \leq 0$ for all $x \in [k^-, k^+]$. To facilitate the analysis, we consider the equivalent condition obtained by multiplying both sides of the inequality by the negative quantity $\frac{(\zeta^+ + \zeta^-)^2}{x - k^+}$, which leads to:

$$t(x) := -3 \cdot \frac{\zeta^+ - \zeta^-}{\zeta^+ + \zeta^-}(x - k^+) + 2(2\zeta^- - \zeta^+) \geq 0.$$

This is a straight line with positive slope. Its minimum is attained at $x = k^-$, where we compute:

$$t(k^-) = 3(\zeta^+ - \pi_1\zeta^+) + 2(2\pi_1\zeta^+ - \zeta^+) \geq 0.$$

Hence, under $\pi_1 \in [1, 2]$, the following condition $t(k^-) \geq 0$ is satisfied. □

Corollary 1.1. The unpaired left-sided cubic BFs also satisfy the non-decreasing monotonicity condition established in Proposition 1.4., provided that $\delta_{j_w}^- \leq 0$, $j = 1, \ldots, m$, $w \in W_j$.

Proposition 1.5. [Concavity for smoothed reflected pairs] Let

$$\hat{f}^{\mathrm{S}}(\boldsymbol{x}) = \tau_0 + \sum_{j=1}^{m} \sum_{p \in P_j} \left[\gamma_{j_p}^+ B_{\mathrm{C}}^+(x_j, \mathcal{T}(\tilde{\kappa}_{j_p})) + \gamma_{j_p}^- B_{\mathrm{C}}^-(x_j, \mathcal{T}(\tilde{\kappa}_{j_p}))\right] + \sum_{j=1}^{m} \sum_{w \in W_j} \delta_{j_w}^- B_{\mathrm{C}}^-(x, \mathcal{T}(\tilde{\kappa}_{j_w}^{(\mathrm{L})}))$$

be the smoothed estimator defined in Equation (4.20) using cubic BFs. Assume that $\gamma_{j_p}^+ \geq 0$, $\gamma_{j_p}^- \leq 0$, and $\gamma_{j_p}^+ + \gamma_{j_p}^- \leq 0$ for all $j = 1, \ldots, m$ and all $p \in P_j$. If the ratio between side knots $\pi_1 = \zeta^-/\zeta^+$ satisfies $\pi_1 \in [1, 2]$, then the estimator $\hat{f}^{\mathrm{S}}(\boldsymbol{x})$ is coordinate-wise concave on the interval $[k^-, k^+]$.

Proof. We analyze the concavity of each reflected pair. For the left-sided BF, according to Equation (4.31), the second derivative over the interval $x \in (k^-, k^+)$ is:

$$\frac{\partial^2 B_C^-(x; k, k^+, k^-)}{\partial x^2} = -6 \cdot \frac{\zeta^+ - \zeta^-}{(\zeta^+ + \zeta^-)^3}(x - k^+) + 2 \cdot \frac{2\zeta^- - \zeta^+}{(\zeta^+ + \zeta^-)^2}.$$

To ensure concavity, we must verify that this expression is non-negative throughout the interval, that is, $\frac{\partial^2 B_C^-(x;k,k^+,k^-)}{\partial x^2} \geq 0$ for all $x \in [k^-, k^+]$. To simplify the analysis, we multiply both sides of the inequality by the positive term $(\zeta^+ + \zeta^-)^2$, which leads to:

$$t(x) := -6 \cdot \frac{\zeta^+ - \zeta^-}{\zeta^+ + \zeta^-}(x - k^+) + 2(2\zeta^- - \zeta^+) \geq 0.$$

This is a straight line with positive slope. Its minimum is attained at $x = k^-$, where we compute:

$$t(k^-) = 6(\zeta^+ - \pi_1\zeta^+) + 2(2\pi_1\zeta^+ - \zeta^+) \geq 0.$$

Hence, under $\pi_1 \in [1, 2]$, the following condition $t(k^-) \geq 0$ is satisfied.

To preserve concavity across the entire reflected pair, a natural strategy is to impose an analogous condition on the second derivative of the right-sided BF. Specifically, one may require:

$$\gamma_{j_p}^+ \cdot \frac{\partial^2 B_C^+(x_j; \zeta_{j_p}^+, \zeta_{j_p}^-)}{\partial x_j^2} + \gamma_{j_p}^- \cdot \frac{\partial^2 B_C^-(x_j; \zeta_{j_p}^+, \zeta_{j_p}^-)}{\partial x_j^2} \leq 0.$$

However, if we evaluate this expression at a point $x = k^+ - \Delta$, with $0 \leq \Delta \leq \zeta^+ + \zeta^-$, it can be shown that:

$$\left|\frac{\partial^2 B_C^+(x; k, k^+, k^-)}{\partial x^2}\right| = \left|\frac{\partial^2 B_C^-(x; k, k^+, k^-)}{\partial x^2}\right| \qquad \text{for all } x \in [k^-, k^+].$$

Consequently, if the condition $|\gamma_{j_p}^-| > |\gamma_{j_p}^+|$ holds, the concavity of the smoothed reflected pair is guaranteed. □

Corollary 1.2. The unpaired left-sided cubic BFs also satisfy the concavity property established in Proposition 1.5., provided that $\delta_{j_w}^- \leq 0$, $j = 1, \ldots, m$, $w \in W_j$.

If the conditions established in Propositions 1.4 and 1.5 are satisfied for every triplet of knots associated with each reflected pair $p \in P_j$ and for every unpaired left-sided BF $w \in W_j$, then the resulting estimator $\hat{f}^{\mathrm{S}}(\boldsymbol{x})$ is globally coordinate-wise non-decreasing monotonic and concave.

4.3.2 Quintic basis function smoothing

On the other hand, the quintic BFs introduced by Chen et al. (1999), reparameterized in terms of the new variables ζ^- and ζ^+, can be written using the following coefficients:

$$\Lambda_1^+ = \frac{6\zeta^+ - 4\zeta^-}{(\zeta^+ + \zeta^-)^3}, \qquad \Lambda_2^+ = \frac{7\zeta^- - 8\zeta^+}{(\zeta^+ + \zeta^-)^4}, \qquad \Lambda_3^+ = \frac{3\zeta^+ - 3\zeta^-}{(\zeta^+ + \zeta^-)^5},$$

$$\Lambda_1^- = \frac{4\zeta^+ - 6\zeta^-}{(\zeta^+ + \zeta^-)^3}, \qquad \Lambda_2^- = \frac{7\zeta^+ - 8\zeta^-}{(\zeta^+ + \zeta^-)^4}, \qquad \Lambda_3^- = \frac{3\zeta^+ - 3\zeta^-}{(\zeta^+ + \zeta^-)^5}.$$

Using these definitions, the right- and left-sided quintic BFs take the form:

$$B_{\mathrm{Q}}^{+}(x;k,k^{+},k^{-})=\begin{cases}0, & x\leq k^{-},\\ \Lambda_{1}^{+}(x-k^{-})^{3}+\Lambda_{2}^{+}(x-k^{-})^{4}+\Lambda_{3}^{+}(x-k^{-})^{5}, & k^{-}<x<k^{+},\\ x-k, & x\geq k^{+},\end{cases} \tag{4.32}$$

$$B_{\mathrm{Q}}^{-}(x;k,k^{+},k^{-})=\begin{cases}k-x, & x\leq k^{-},\\ \Lambda_{1}^{-}(x-k^{+})^{3}+\Lambda_{2}^{-}(x-k^{+})^{4}+\Lambda_{3}^{-}(x-k^{+})^{5}, & k^{-}<x<k^{+},\\ 0, & x\geq k^{+}.\end{cases} \tag{4.33}$$

The first-order derivatives of the aforementioned expressions are as follows:

$$\frac{\partial B_{\mathrm{Q}}^{+}(x;k,k^{+},k^{-})}{\partial x}=\begin{cases}0, & x\leq k^{-},\\ 3\Lambda_{1}^{+}(x-k^{-})^{2}+4\Lambda_{2}^{+}(x-k^{-})^{3}+5\Lambda_{3}^{+}(x-k^{-})^{4}, & k^{-}<x<k^{+},\\ 1, & x\geq k^{+},\end{cases} \tag{4.34}$$

$$\frac{\partial B_{\mathrm{Q}}^{-}(x;k,k^{+},k^{-})}{\partial x}=\begin{cases}-1, & x\leq k^{-},\\ 3\Lambda_{1}^{-}(x-k^{+})^{2}+4\Lambda_{2}^{-}(x-k^{+})^{3}+5\Lambda_{3}^{-}(x-k^{+})^{4}, & k^{-}<x<k^{+},\\ 0, & x\geq k^{+}.\end{cases} \tag{4.35}$$

Finally, the second-order derivatives of the previously introduced expressions are presented below:

$$\frac{\partial^{2} B_{\mathrm{Q}}^{+}(x;k,k^{+},k^{-})}{\partial x^{2}}=\begin{cases}0, & x\leq k^{-},\\ 6\Lambda_{1}^{+}(x-k^{-})+12\Lambda_{2}^{+}(x-k^{-})^{2}+20\Lambda_{3}^{+}(x-k^{-})^{3}, & k^{-}<x<k^{+},\\ 0, & x\geq k^{+},\end{cases} \tag{4.36}$$

$$\frac{\partial^{2} B_{\mathrm{Q}}^{-}(x;k,k^{+},k^{-})}{\partial x^{2}}=\begin{cases}0, & x\leq k^{-},\\ 6\Lambda_{1}^{-}(x-k^{+})+12\Lambda_{2}^{-}(x-k^{+})^{2}+20\Lambda_{3}^{-}(x-k^{+})^{3}, & k^{-}<x<k^{+},\\ 0, & x\geq k^{+}.\end{cases} \tag{4.37}$$

We now establish conditions under which the smoothed estimator satisfies the standard shape constraints of production theory, namely non-decreasing monotonicity and concavity.

Proposition 1.6. [Monotonicity] Let

$$\hat{f}^{\mathrm{S}}(\boldsymbol{x}) = \tau_0 + \sum_{j=1}^{m} \sum_{p \in P_j} \left[\gamma_{j_p}^{+} \, B_{\mathrm{Q}}^{+}(x_j, \mathcal{T}(\tilde{\kappa}_{j_p})) + \gamma_{j_p}^{-} \, B_{\mathrm{Q}}^{-}(x_j, \mathcal{T}(\tilde{\kappa}_{j_p})) \right] + \sum_{j=1}^{m} \sum_{w \in W_j} \delta_{j_w}^{-} \, B_{\mathrm{Q}}^{-}(x_j, \mathcal{T}(\tilde{\kappa}_{j_w}^{(\mathrm{L})}))$$

be the smoothed estimator defined in Equation (4.20) using quintic BFs. Assume that $\gamma_{j_p}^{+} + \gamma_{j_p}^{-} \leq 0$, $\gamma_{j_p}^{+} \geq 0$, and $\gamma_{j_p}^{-} \leq 0$ for all $p \in P_j$, $j = 1, \ldots, m$. If the ratio between side knots $\pi_2 = \zeta^{+}/\zeta^{-}$ satisfies $\pi_2 \in \left[\frac{8}{7}, \ \frac{3}{2}\right]$, then the estimator $\hat{f}^{\mathrm{S}}(\boldsymbol{x})$ is coordinate-wise non-decreasing on the interval $[k^{-}, k^{+}]$.

Proof. To prove this result, we show that the derivative of the smoothed estimator is non-negative over the interval $[k^{-}, k^{+}]$. First, we analyze the right-sided BF. Its first derivative over $x \in (k^{-}, k^{+})$ is:

$$\frac{\partial B_{\mathrm{Q}}^{+}(x; k, k^{+}, k^{-})}{\partial x} = 5 \cdot \frac{3\zeta^{+} - 3\zeta^{-}}{(\zeta^{+} + \zeta^{-})^{5}} (x - k^{-})^{4} + 4 \cdot \frac{7\zeta^{-} - 8\zeta^{+}}{(\zeta^{+} + \zeta^{-})^{4}} (x - k^{-})^{3} + 3 \cdot \frac{6\zeta^{+} - 4\zeta^{-}}{(\zeta^{+} + \zeta^{-})^{3}} (x - k^{-})^{2}.$$

To ensure monotonicity, we must verify that the expression remains non-negative over the entire interval, that is, $\frac{\partial B_{\mathrm{Q}}^{+}(x; k, k^{+}, k^{-})}{\partial x} \geq 0$ for all $x \in [k^{-}, k^{+}]$. For ease of analysis, we rewrite this condition in an equivalent form by multiplying both sides of the inequality by the positive quantity $\frac{(\zeta^{+} + \zeta^{-})^{3}}{(x - k^{-})^{2}}$, leading to the following convex quadratic polynomial:

$$t(x) := 5 \cdot \frac{3\zeta^{+} - 3\zeta^{-}}{(\zeta^{+} + \zeta^{-})^{2}} (x - k^{-})^{2} + 4 \cdot \frac{7\zeta^{-} - 8\zeta^{+}}{\zeta^{+} + \zeta^{-}} (x - k^{-}) + 3 \cdot (6\zeta^{+} - 4\zeta^{-}) \geq 0.$$

This parabola is defined on the interval $[k^{-}, k^{+}]$, and its minimum occurs at vertex

$$x^{*} = -\frac{2(\zeta^{+} + \zeta^{-})(7\zeta^{-} - 8\zeta^{+})}{15(\zeta^{+} - \zeta^{-})} + k^{-}.$$

Evaluating the minimum value at the vertex, we obtain:

$$t(x^{*}) = \frac{4(7\zeta^{-} - 8\pi_2\zeta^{-})^{2}}{15(\pi_2\zeta^{-} - \zeta^{-})} - \frac{8(7\zeta^{-} - 8\pi_2\zeta^{-})^{2}}{15(\pi_2\zeta^{-} - \zeta^{-})} + 3(6\pi_2\zeta^{-} - 4\zeta^{-}) \geq 0.$$

Finally, it can be verified that this value is non-negative when $\pi_2 \geq \frac{8}{7}$.

Next, we analyze the effect of left-sided BF on reflected pairs. Its first derivative over $x \in (k^{-}, k^{+})$ is:

$$\frac{\partial B_{\mathrm{Q}}^{-}(x; k, k^{+}, k^{-})}{\partial x} = 5 \cdot \frac{3\zeta^{+} - 3\zeta^{-}}{(\zeta^{+} + \zeta^{-})^{5}} (x - k^{+})^{4} + 4 \cdot \frac{7\zeta^{+} - 8\zeta^{-}}{(\zeta^{+} + \zeta^{-})^{4}} (x - k^{+})^{3} + 3 \cdot \frac{4\zeta^{+} - 6\zeta^{-}}{(\zeta^{+} + \zeta^{-})^{3}} (x - k^{+})^{2}.$$

To ensure non-decreasing monotonicity, we require that $\frac{\partial B_{\mathrm{Q}}^{-}(x;k,k^{+},k^{-})}{\partial x} \leq 0$ for all $x \in [k^{-}, k^{+}]$. To facilitate the analysis, we consider the equivalent condition obtained by multiplying both sides of the inequality by the positive quantity $\frac{(\zeta^{+}+\zeta^{-})^{3}}{(x-k^{+})^{2}}$, leading to the following convex quadratic polynomial:

$$t(x) := 5 \cdot \frac{3\zeta^{+} - 3\zeta^{-}}{(\zeta^{+}+\zeta^{-})^{2}}(x-k^{+})^{2} + 4 \cdot \frac{7\zeta^{+} - 8\zeta^{-}}{\zeta^{+}+\zeta^{-}}(x-k^{+}) + 3(4\zeta^{+} - 6\zeta^{-}) \leq 0.$$

The maximum of this parabola occurs at one of the endpoints of $[k^{-}, k^{+}]$. It can be verified that:

$$t(k^{-}) = -5(3\pi_2\zeta^{-} - 3\zeta^{-}) - 4(7\pi_2\zeta^{-} - 8\zeta^{-}) + 3(4\pi_2\zeta^{-} - 6\zeta^{-}) \leq 0 \quad \text{when } \pi_2 \geq \frac{29}{31},$$

and that:

$$t(k^{+}) = 3(4\pi_2\zeta^{-} - 6\zeta^{-}) \leq 0 \quad \text{when } \pi_2 \leq \frac{3}{2}.$$

Therefore, if $\pi_2 \in \left[\frac{8}{7},\ \frac{3}{2}\right]$, both derivatives satisfy the required sign conditions over the entire interval, and the estimator is non-decreasing monotonic. □

Corollary 1.3. The unpaired left-sided quintic BFs also satisfy the non-decreasing monotonicity condition established in Proposition 1.6., provided that $\delta_{j_w}^{-} \leq 0$, $j = 1, \ldots, m$, $w \in W_j$.

Proposition 1.7 [Concavity] Let

$$\hat{f}^{\mathrm{S}}(\boldsymbol{x}) = \tau_0 + \sum_{j=1}^{m} \sum_{p \in P_j} \left[\gamma_{j_p}^{+} B_{\mathrm{Q}}^{+}(x_j, \mathcal{T}(\tilde{\kappa}_{j_p})) + \gamma_{j_p}^{-} B_{\mathrm{Q}}^{-}(x_j, \mathcal{T}(\tilde{\kappa}_{j_p}))\right] + \sum_{j=1}^{m} \sum_{w \in W_j} \delta_{j_w}^{-} B_{\mathrm{Q}}^{-}(x, \mathcal{T}(\tilde{\kappa}_{j_w}^{(\mathrm{L})}))$$

be the smoothed estimator in (4.20) based on quintic BFs. Assume that $\gamma_{j_p}^{+} + \gamma_{j_p}^{-} \leq 0$, $\gamma_{j_p}^{+} \geq 0$, and $\gamma_{j_p}^{-} \leq 0$ for every $p \in P_j$, $j = 1, \ldots, m$. If the ratio between side knots $\pi_2 = \zeta^{+}/\zeta^{-}$ satisfies $\pi_2 \in \left[\frac{8}{7},\ \frac{3}{2}\right]$, then the estimator $\hat{f}^{\mathrm{S}}(\boldsymbol{x})$ is coordinate-wise concave on the interval $[k^{-}, k^{+}]$.

Proof. We analyze the concavity of each reflected pair. First, we prove that the second derivative of the left-sided BFs is non-negative over (k^{-}, k^{+}). That is,

$$\frac{\partial^2 B_{\mathrm{Q}}^{-}(x;k,k^{+},k^{-})}{\partial x^2} = 20\frac{3\zeta^{+} - 3\zeta^{-}}{(\zeta^{+}+\zeta^{-})^{5}}(x-k^{+})^{3} + 12\frac{7\zeta^{+} - 8\zeta^{-}}{(\zeta^{+}+\zeta^{-})^{4}}(x-k^{+})^{2} + 6\frac{4\zeta^{+} - 6\zeta^{-}}{(\zeta^{+}+\zeta^{-})^{3}}(x-k^{+}) \geq 0.$$

Multiplying both sides by the negative term $\frac{(\zeta^{+}+\zeta^{-})^{3}}{x-k^{+}}$, we obtain the equivalent convex quadratic expression:

$$t(x) := 20 \cdot \frac{3\zeta^{+} - 3\zeta^{-}}{(\zeta^{+}+\zeta^{-})^{2}}(x-k^{+})^{2} + 12 \cdot \frac{7\zeta^{+} - 8\zeta^{-}}{\zeta^{+}+\zeta^{-}}(x-k^{+}) + 6(4\zeta^{+} - 6\zeta^{-}) \leq 0.$$

Since this parabola is only defined over $[k^-, k^+]$, its maximum occurs at the endpoints. We evaluate:

$$t(k^-) = -20(3\pi_2\zeta^- - 3\zeta^-) - 12(7\pi_2\zeta^- - 8\zeta^-) + 6(4\pi_2\zeta^- - 6\zeta^-) \leq 0 \quad \text{for } \pi_2 \geq 1,$$

$$t(k^+) = 6(4\pi_2\zeta^- - 6\zeta^-) \leq 0 \quad \text{for } \pi_2 \leq \tfrac{3}{2}.$$

Hence, under $\pi_2 \in \left[\frac{8}{7},\ \frac{3}{2}\right]$, the following conditions $t(k^-) \leq 0$ and $t(k^+) \leq 0$ are satisfied.

Preserving concavity over the reflected pair can be achieved by enforcing a corresponding condition on the second derivative of the right-sided BF. Specifically, one may require:

$$\gamma_{j_p}^+ \cdot \frac{\partial^2 B_Q^+(x_j; \zeta_{j_p}^+, \zeta_{j_p}^-)}{\partial x_j^2} + \gamma_{j_p}^- \cdot \frac{\partial^2 B_Q^-(x_j; \zeta_{j_p}^+, \zeta_{j_p}^-)}{\partial x_j^2} \leq 0.$$

However, if we define $x = k^- + \Delta^-$, with $\Delta^- \geq 0$, and $x = k^+ - \Delta^+$, with $\Delta^+ \geq 0$, such that $\Delta^+ + \Delta^- = k^+ - k^- = \Delta$, then can be easily shown that

$$\left|\frac{\partial^2 B_Q^+(x; k, k^+, k^-)}{\partial x^2}\right| = \left|\frac{\partial^2 B_Q^-(x; k, k^+, k^-)}{\partial x^2}\right| \quad \text{for all } x \in [k^-, k^+].$$

Therefore, since $|\gamma_{j_p}^-| > |\gamma_{j_p}^+|$, concavity is ensured. □

Corollary 1.4. The unpaired left-sided quintic BFs also satisfy the concavity property established in Proposition 1.7., provided that $\delta_{j_w}^- \leq 0$, $j = 1, \ldots, m$, $w \in W_j$.

Finally, it is important to note that the distance between each central knot k and its associated lateral knots k^- and k^+ is not treated as a variable to be optimized. Instead, this spacing—governed by the pair (ζ^-, ζ^+)—is parameterized through the ratio $\pi_1 = \zeta^-/\zeta^+$ or $\pi_2 = \zeta^+/\zeta^-$, which is introduced as a hyperparameter. Although a valid alternative would be to jointly estimate the knot locations and model coefficients by solving a Nonlinear Programming problem, this would significantly increase the complexity of the estimation procedure. In particular, the model would lose its LP structure, as the BFs would become nonlinear in the knot locations. To avoid this, we adopt a simpler and computationally efficient strategy that decouples the knot placement from the coefficient estimation: π_1 and π_2 is treated as tuning hyperparameters, selected externally based on validation performance.

Finally, Figure 4.4 provides an example of the smoothed cubic and quintic estimators constructed using the procedures detailed in this chapter.

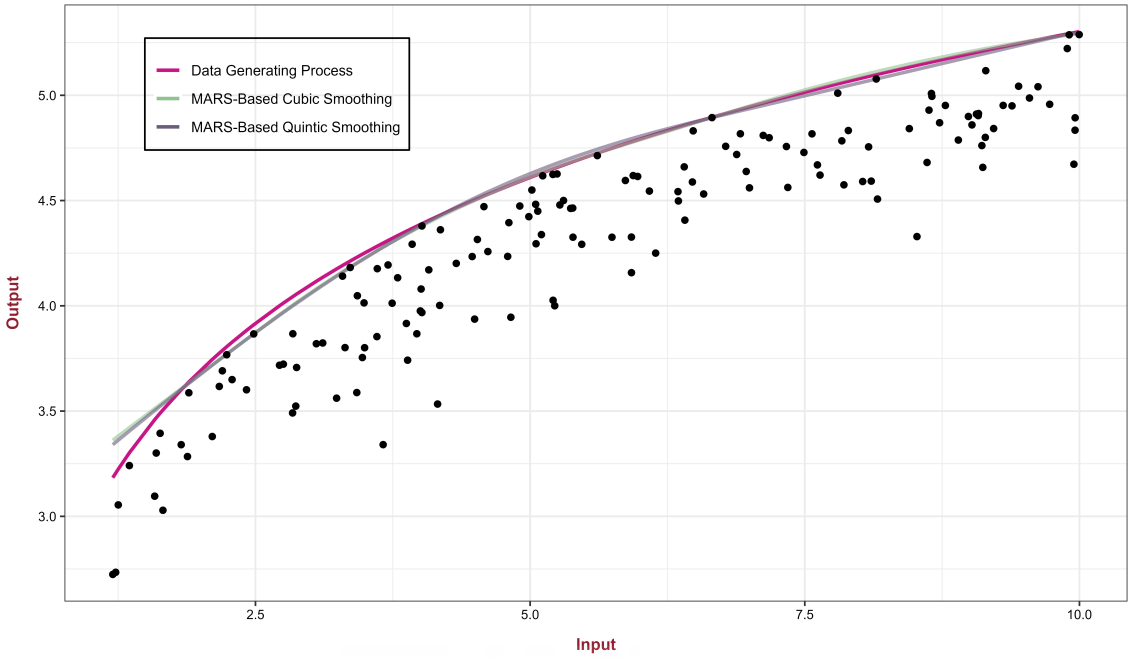

Figure 4.4: A graphical representation of the predictor generated by the smoothed estimators.

4.4 Computational experiments

This section presents a set of computational analyses designed to support and illustrate the contributions introduced throughout the chapter. First, we characterize the overfitting phenomenon frequently encountered in DEA by means of a concrete illustrative example, emphasizing its impact on model generalization. Next, we compile the full list of hyperparameters introduced in the proposed framework, clarifying their role and interpretation within the estimation process. In addition to listing them, we also propose a practical procedure for determining suitable values in empirical applications. Finally, we apply the Experimental Design 1 described in Section 3.4.1 to compare the performance of our estimator against two widely used benchmark models: the DEA estimator proposed by Banker et al. (1984) and the CCNLS introduced by Kuosmanen and Johnson (2010). All models are evaluated in terms of their mean squared error (MSE) and bias, both of which provide complementary insights into the accuracy and systematic deviation of each method. In addition to evaluating empirical performance, this analysis also helps identify specific areas where the proposed technique may require further refinement in order to become more competitive with existing methods, laying the groundwork for the improvements proposed in the next chapter.

4.4.1 The overfitting problem in DEA models

The statement that DEA and related non-parametric envelopment techniques such as Free Disposal Hull (FDH) suffer from an overfitting problem is relatively recent in the literature. Statistically, overfitting occurs when a model fits the observed data perfectly, but fails to generalize well to unseen samples, leading to poor predictive performance and a high generalization error (Hastie et al., 2009). Standard ML techniques aim to recover the actual function behind the DGP by balancing two competing objectives: minimizing the error on the observed sample, and maintaining robustness to new data (Vapnik, 1999). When the right equilibrium is achieved between these goals, the model can yield an accurate estimate of the underlying function. This trade-off is closely linked to the concept of generalization error. In non-parametric frameworks, the generalization error cannot be computed in closed form, but it can be approximated using hold-out test sets or CV. In the context of efficiency analysis, envelopment techniques such as DEA are designed to wrap the frontier tightly around the observed sample under the minimal extrapolation principle. While this property allows DEA to accurately assess the efficiency of observed DMUs within the sample, it also makes it highly prone to overfitting. When the true frontier lies above the cloud of observed points, the DEA surface becomes excessively tailored to the data, thereby limiting its inferential capability—especially in small-sample settings. This limitation is particularly relevant when the objective is not only to describe observed units, but to approximate the function that governs the production process. A direct consequence of this overfitting effect is that a large portion of the units under evaluation are classified as technically efficient despite being located far from the true frontier. As a result, DEA scores tend to be overly optimistic, systematically underestimating inefficiency. Only recently has the literature begun to explicitly recognize and address this issue. Several contributions have pointed out the structural overfitting tendencies of DEA and FDH and their implications for inference and predictive validity (see Esteve et al. 2020; Tsionas 2022; Valero-Carreras et al. 2022; Molinos-Senante et al. 2023).

Now, we present a graphical example to illustrate the overfitting phenomenon inherent in DEA. In Figure 4.5, we display a sample of DMUs assessed for efficiency using a Variable returns to scale (VRS) frontier. Simultaneously, the figure reveals the underlying DGP, which represents the maximum attainable output given an input profile. Notably, DEA tends to produce overoptimistic efficiency estimates, which can distort performance evaluations. To highlight this issue, we examine the specific cases of DMUs A and B. While DMU A is deemed efficient, DMU B would need to slightly increase its output—keeping input constant—to be considered efficient under the output-oriented radial measure of Banker et al. (1984). However, both units remain considerably distant from the true frontier defined by the DGP, thereby underscoring the fact that DEA estimates are entirely driven by the observed sample, and may not reflect the actual efficiency landscape.

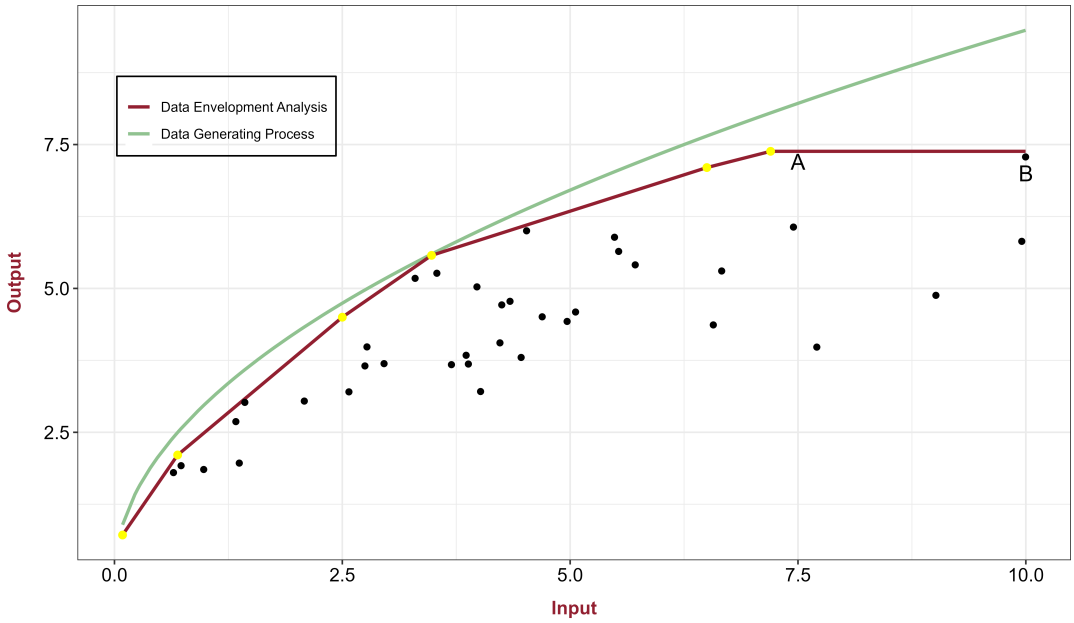

Figure 4.5: An example of overoptimistic DEA scores.

4.4.2 Hyperparameter configuration

The proposed methodology incorporates several hyperparameters that govern different aspects of the model construction process, including BF selection, knot placement, and regularization. These hyperparameters are crucial for balancing model flexibility and generalization, as well as for managing computational efficiency. Table 4.1 provides a complete list of all hyperparameters considered throughout the study, along with a brief description of their purpose and implementation.

Table 4.1: List of hyperparameters used in the proposed MARS-based adaptation.

Hyperparameter	**Description**
η	Maximum number of BFs allowed after the forward phase.
ξ	Minimum relative improvement in the LOF criterion required to add a new pair of BFs.
`LOF`	Criterion to evaluate BF performance (e.g., MSE, MAE).
`grid`	Candidate set of knot locations.
`minspan`	Minimum number of observations between two consecutive knots.
`endspan`	Minimum number of observations from extreme knots to data boundaries.
d	Penalty factor on the number of knots used in the backward pruning phase.
$\pi_1 \in [1, 2]$	Ratio between distances to adjacent knots in cubic BFs to ensure shape-constraints.
$\pi_2 \in \left[\frac{8}{7}, \frac{3}{2}\right]$	Ratio between distances to adjacent knots in quintic BFs to ensure shape-constraints.

In predictive modeling, model evaluation typically relies on two standard validation strategies: the training–test split and k-fold CV. In the training–test approach, the dataset is divided into two mutually exclusive subsets. The model is fitted on the training set and subsequently evaluated on the test set, which serves as a proxy for new, unseen data. This method provides a straightforward estimate of generalization performance but may suffer from high variance depending on how the split is performed. Alternatively, k-fold CV offers a more comprehensive assessment by partitioning the dataset into k equally size folds. The model is trained and validated k times, each time using a different fold for validation and the remaining k-1 folds for training. This procedure reduces variance in the performance estimate and is particularly well-suited for hyperparameter tuning, as it ensures that all data points contribute to both model training and validation.

For Experimental Design 1, we conducted an exploration of the hyperparameter space by testing a selected set of combinations of the predefined values introduced above. Specifically, we fixed the maximum number of BFs to be included in the model (η) as a large value—typically equal to the sample size—and controlled the growth of the forward algorithm through the hyperparameter ξ, which determines the minimum relative improvement in the LOF function required to add a new BF. Based on prior experience, we considered the finite set $\xi \in \{0,\ 0.001,\ 0.01,\ 0.1\}$. Coefficients were estimated using Model (4.15); however, the MSE was used as the LOF criterion to determine which BFs were added during the forward selection. The grid of available knots was defined using the observed data, following the approach of Friedman (1991). Regarding `minspan` and `endspan`, we adopted the spacing rules proposed by Friedman (1991), as described in Equations (3.37) and (3.38), as well as the alternative formulation of Zhang (1994), presented in Equation (3.39). For the backward algorithm, the hyperparameter d, which penalizes the number of knots in Equation (3.40), was also tuned. Finally, for the smoothing procedure, π_1 and π_2 were fixed at the midpoints of their respective admissible intervals. These values generated a total of 48 different hyperparameter combinations for the proposed simulations. Each configuration was evaluated using a training–test split, with 70% of the data used for training and the remaining 30% for testing.

Table 4.2 reports the mean, standard deviation (in brackets), and median of the best-performing hyperparameters across our simulations. It is important to note that this table does not reflect the configurations that are closest to the true optimal values (which are known in the simulation setting). Instead, it summarizes the hyperparameter values most frequently selected as optimal based on the proposed 70/30 training–test split. That is, the reported statistics correspond to the configurations that most consistently led to the best test (out-of-sample) performance across repeated trials. For instance, values of ξ around 0.01 (both slightly above and below) were most frequently chosen, striking a balance between model flexibility and computational efficiency. In the case of `minspan`, the Friedman-based spacing rule was generally preferred. For `endspan`, the Friedman and Zhang approaches yielded

similar results and were selected with comparable frequency. Regarding the regularization parameter d, value 1 emerged as the most reliable, while value 3 was rarely selected, likely due to underfitting. Table 4.2 also reports the average computation time of the proposed technique. A notable drawback of the approach is its computational cost, especially when compared to simpler techniques like DEA, which only rely on LP models. The results show an exponential increase in execution time with sample size, a trend that becomes even more pronounced as the number of inputs increases.

4.4.3 Results from Experimental Design 1

Tables 4.3 and 4.4 summarize the results of Experimental Design 1 in terms of MSE, comparing the performance of the proposed linear, cubic, and quintic MARS-based estimators with two benchmark methods: DEA and CCNLS. In both tables, the first two columns describe the experimental scenario, including the functional form, number of inputs, and sample size.

Overall, the proposed methods demonstrate substantial improvements in estimation accuracy relative to DEA and CCNLS across most scenarios. As expected, all methods exhibit some deterioration in performance as the dimensionality increases, i.e., when moving from one-input to two- and three-input settings. However, the increase in MSE is significantly more pronounced in DEA. For example, from Scenario 1 (1 input) to Scenario 4 (3 inputs), the MSE in DEA rises from 0.010 to 0.065 for the smallest sample size, which represents a 550% increase. In contrast, our linear MARS-based estimator increases only from 0.007 to 0.018, showing a considerably more robust behavior under higher dimensionality.

When comparing relative performance (Table 4.4), the proposed estimators consistently outperform DEA, with improvements reaching up to 86.23% in Scenario 6 with 150 observations. Similar improvements are observed over CCNLS, with relative gains of up to 69.50%. The proposed methods also perform well in scenarios with two or three inputs; however, their performance is influenced by the presence and strength of input interactions. For instance, Scenarios 5 and 6 exhibit the strongest absolute performance, achieving the best MSE improvements across all sample sizes. Scenario 4 also delivers competitive results, but its performance relative to CCNLS becomes less favorable as the sample size increases, in some cases even being slightly worse. In contrast, Scenario 3 stands out as the most challenging. In this setting, the linear model performs poorly and even underperforms DEA. This is likely due to the strong interaction term in the true production function, which conflicts with the additive structure of the MARS-based estimators. Interestingly, when the exponent of the interaction term is reduced, as in Scenario 5, performance improves significantly. To further investigate this, an additional computational experiment was conducted where the interaction exponent was varied from 0.1 to 1. Results suggest a performance threshold around 0.5, above which the estimation error of additive models increases rapidly, regardless of sample size.

Table 4.5 reports the estimation bias of each method under Experimental Design 1. The table mirrors the structure of Table 4.3. We begin by examining the performance of DEA. As shown, DEA consistently underestimates the true production function across all scenarios, as evidenced by the uniformly negative bias values (Korostelëv et al., 1995). This confirms DEA's known tendency to produce conservative frontier estimates. While such behavior allows DEA to provide a descriptive envelope of the data, it limits its inferential accuracy unless large-sample properties—such as consistency—are explicitly invoked. Additionally, DEA's bias decreases in magnitude as the sample size increases, reflecting improved approximation with more data. On the other side, CCNLS consistently exhibits a positive bias across all scenarios, suggesting a tendency to overestimate the production frontier

Turning to the proposed MARS-based estimators, results reveal that the bias is generally closer to zero than in DEA or CCNLS. These models tend to slightly underestimate the production function in most scenarios, particularly for smaller sample sizes. However, the magnitude of this underestimation diminishes as the sample size grows, indicating model convergence. An exception to this pattern is Scenario 3, where all three variants of the proposed method (linear, cubic, and quintic) exhibit positive bias values. This suggests an overestimation of the production function, likely caused by the strong interaction term embedded in the true model. As previously discussed in the MSE analysis, additive models may struggle to capture complex interaction effects, especially when such interactions are not mild.

Table 4.2: Optimal hyperparameters for the new MARS-based approach.

Scenario (#inputs)	Sample size	ξ		`minspan`		`endspan`		d		Execution Time (in seconds)
		Mean (std)	Median	Friedman	Zhang	Friedman	Zhang	Mean (std)	Median	
1 (1)	50	0.059 (0.048)	0.100	67	33	53	47	1.45 (0.64)	1	0.32
	100	0.025 (0.040)	0.001	64	36	65	35	1.42 (0.70)	1	2.73
	150	0.011 (0.027)	0.001	60	40	75	25	1.24 (0.55)	1	6.93
2 (1)	50	0.043 (0.048)	0.010	71	29	60	40	1.48 (0.75)	1	0.46
	100	0.014 (0.031)	0.001	55	45	63	37	1.24 (0.53)	1	3.44
	150	0.004 (0.011)	0.001	58	42	62	38	1.22 (0.46)	1	9.16
3 (2)	50	0.045 (0.047)	0.010	63	37	39	61	1.84 (0.85)	2	1.05
	100	0.038 (0.046)	0.010	64	36	52	48	1.64 (0.84)	1	4.69
	150	0.028 (0.041)	0.010	79	21	68	32	1.65 (0.82)	1	11.02
4 (3)	50	0.034 (0.045)	0.010	58	42	48	52	1.92 (0.87)	2	2.16
	100	0.018 (0.034)	0.001	66	34	49	51	1.42 (0.68)	1	14.11
	150	0.016 (0.030)	0.010	68	32	45	55	1.47 (0.73)	1	36.72
5 (2)	50	0.046 (0.047)	0.010	69	31	49	51	1.97 (0.87)	2	1.00
	100	0.042 (0.047)	0.010	76	24	66	34	1.56 (0.77)	1	4.73
	150	0.034 (0.044)	0.010	76	24	67	33	1.50 (0.70)	1	12.12
6 (3)	50	0.036 (0.046)	0.010	73	27	40	60	2.07 (0.77)	2	2.32
	100	0.024 (0.039)	0.010	60	40	41	39	1.56 (0.74)	1	12.50
	150	0.020 (0.037)	0.001	60	40	64	36	1.43 (0.69)	1	39.46

Table 4.3: MSE and standard deviation (in brackets) for each method in Experimental Design 1.

Scenario (#inputs)	Sample size	Linear MARS-based	Cubic MARS-based	Quintic MARS-based	DEA	CCNLS
1 (1)	50	0.007 (0.007)	0.005 (0.005)	0.007 (0.008)	0.010 (0.007)	0.006 (0.005)
	100	0.003 (0.002)	0.002 (0.002)	0.002 (0.002)	0.005 (0.002)	0.005 (0.005)
	150	0.002 (0.001)	0.001 (0.001)	0.002 (0.002)	0.003 (0.002)	0.004 (0.005)
2 (1)	50	0.007 (0.006)	0.006 (0.006)	0.009 (0.009)	0.011 (0.008)	0.009 (0.006)
	100	0.004 (0.003)	0.002 (0.002)	0.003 (0.003)	0.006 (0.003)	0.006 (0.007)
	150	0.002 (0.001)	0.001 (0.001)	0.002 (0.002)	0.003 (0.002)	0.006 (0.006)
3 (2)	50	0.018 (0.008)	0.024 (0.024)	0.022 (0.023)	0.030 (0.014)	0.013 (0.008)
	100	0.018 (0.006)	0.028 (0.028)	0.024 (0.024)	0.018 (0.005)	0.009 (0.007)
	150	0.018 (0.004)	0.026 (0.026)	0.024 (0.024)	0.012 (0.004)	0.010 (0.006)
4 (3)	50	0.018 (0.007)	0.021 (0.021)	0.020 (0.020)	0.065 (0.020)	0.024 (0.012)
	100	0.016 (0.005)	0.020 (0.020)	0.018 (0.018)	0.046 (0.011)	0.017 (0.007)
	150	0.015 (0.005)	0.022 (0.022)	0.018 (0.018)	0.035 (0.008)	0.015 (0.009)
5 (2)	50	0.008 (0.005)	0.006 (0.006)	0.007 (0.007)	0.029 (0.013)	0.014 (0.008)
	100	0.004 (0.002)	0.003 (0.003)	0.004 (0.004)	0.015 (0.005)	0.008 (0.007)
	150	0.003 (0.002)	0.003 (0.003)	0.004 (0.004)	0.011 (0.004)	0.007 (0.005)
6 (3)	50	0.012 (0.006)	0.010 (0.010)	0.011 (0.011)	0.060 (0.019)	0.022 (0.012)
	100	0.006 (0.003)	0.006 (0.006)	0.007 (0.007)	0.040 (0.011)	0.016 (0.007)
	150	0.004 (0.002)	0.005 (0.005)	0.006 (0.006)	0.030 (0.008)	0.014 (0.008)

Table 4.4: Comparison of relative MSE performance: proposed methods vs. DEA and CCNLS in Experimental Design 1.

Scenario (#inputs)	Sample size	Linear MARS-based		Cubic MARS-based		Quintic MARS-based	
		DEA	CCNLS	DEA	CCNLS	DEA	CCNLS
1 (1)	50	-30.45%	+14.38%	-51.40%	-19.00%	-34.80%	+8.67%
	100	-44.59%	-46.78%	-67.20%	-67.20%	-51.80%	-51.80%
	150	-35.71%	-55.33%	-60.33%	-70.25%	-43.00%	-57.25%
2 (1)	50	-32.99%	-15.64%	-42.73%	-30.00%	-19.18%	-1.22%
	100	-41.08%	-36.82%	-62.83%	-62.83%	-45.17%	-45.17%
	150	-36.81%	-63.08%	-54.33%	-77.17%	-27.33%	-63.67%
3 (2)	50	-39.29%	+38.52%	-20.00%	+84.62%	-25.17%	+72.69%
	100	-3.39%	+86.79%	+56.39%	+369.17%	+30.56%	+291.67%
	150	+51.86%	+76.32%	+116.33%	+159.60%	+100.33%	+140.40%
4 (3)	50	-71.86%	-23.77%	-67.29%	-11.42%	-69.22%	-16.63%
	100	-65.11%	-6.80%	-56.28%	+18.29%	-60.72%	+6.29%
	150	-56.08%	+1.12%	-37.17%	+46.60%	-49.89%	+16.93%
5 (2)	50	-72.39%	-41.79%	-78.59%	-55.64%	-74.41%	-47.00%
	100	-73.57%	-53.96%	-77.00%	-56.88%	-71.53%	-46.63%
	150	-75.04%	-60.51%	-74.82%	-60.43%	-66.45%	-47.29%
6 (3)	50	-80.30%	-46.59%	-83.95%	-56.23%	-82.03%	-51.00%
	100	-84.34%	-60.52%	-84.75%	-61.88%	-82.78%	-56.94%
	150	-86.23%	-69.50%	-82.77%	-63.07%	-80.03%	-57.21%

Table 4.5: Bias of each method in Experimental Design 1.

Scenario (#inputs)	Sample size	Linear MARS-based	Cubic MARS-based	Quintic MARS-based	DEA	CCNLS
1 (1)	50	-0.036	-0.031	-0.033	-0.070	+0.019
	100	-0.021	-0.018	-0.020	-0.044	+0.038
	150	-0.017	-0.015	-0.017	-0.035	+0.044
2 (1)	50	-0.031	-0.035	-0.031	-0.074	+0.023
	100	-0.024	-0.020	-0.019	-0.051	+0.044
	150	-0.018	-0.016	-0.015	-0.037	+0.050
3 (2)	50	+0.002	+0.026	+0.015	-0.124	+0.011
	100	+0.034	+0.074	+0.061	-0.092	+0.040
	150	+0.049	+0.078	+0.070	-0.072	+0.049
4 (3)	50	-0.019	+0.002	-0.009	-0.199	-0.023
	100	+0.010	+0.043	+0.029	-0.162	+0.023
	150	+0.032	+0.064	+0.044	-0.140	+0.045
5 (2)	50	-0.042	-0.030	-0.027	-0.120	+0.015
	100	-0.021	-0.006	-0.004	-0.081	+0.035
	150	-0.012	+0.001	+0.002	-0.067	+0.044
6 (3)	50	-0.055	-0.037	-0.041	-0.190	-0.017
	100	-0.029	-0.013	-0.015	-0.149	+0.025
	150	-0.012	+0.002	-0.003	-0.124	+0.045

4.5 Discussion and final remarks

This chapter has established a novel connection between production theory and ML. In particular, we have introduced a new technique for estimating production functions based on splines and recursive partitioning. The proposed method builds on an additive version of the MARS approach developed by Friedman (1991), incorporating shape constraints to ensure that the estimated surface envelops the data from above while satisfying monotonicity and concavity conditions.

The main limitation of the proposed approach lies in its methodological foundation: the model is inherently additive, which may hinder its performance when the true (and unknown) production function exhibits strong interactions among inputs. Consequently, it is important to acknowledge that, in practical applications, the actual performance will depend on the nature and intensity of such interactions, as well as the degree of curvature in the underlying technology—features that are typically unobservable to researchers. These limitations underscore the importance of designing models that are capable of capturing complex dependencies when they arise. To illustrate the impact of this limitation, we present an example from Scenario 3 with a sample size of 100, where the performance of the proposed method is notably reduced due to the presence of a strong interaction terms in the true production function.

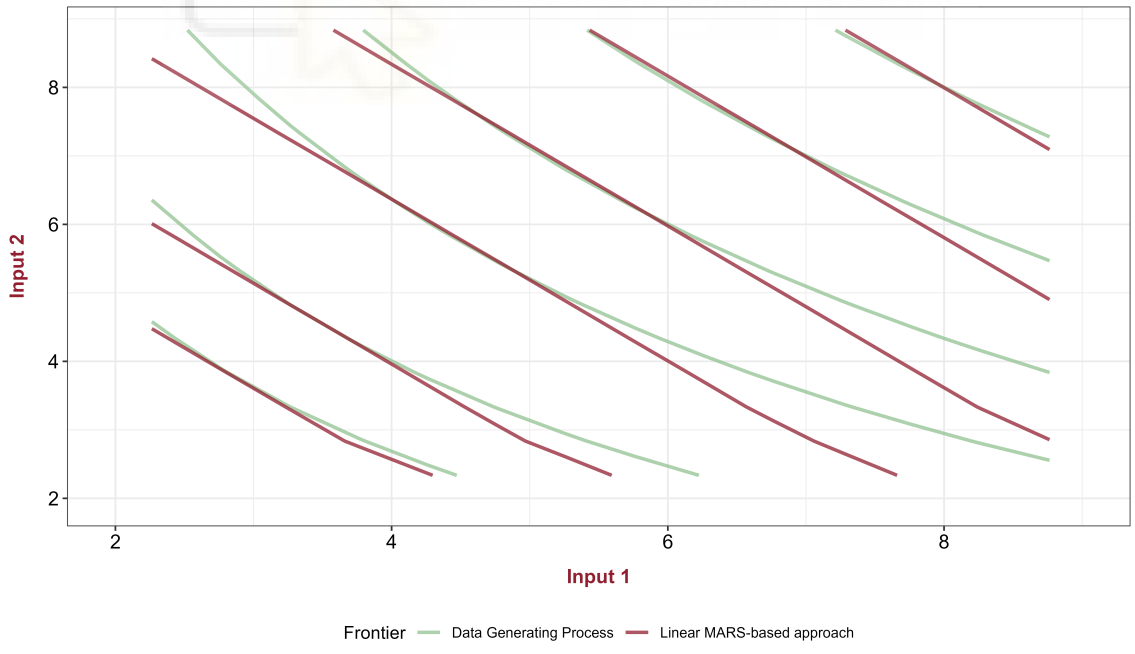

Figure 4.6: Contour plot illustrating the performance of the proposed method in a two-input, one-output example.

However, despite this structural constraint, we have demonstrated—through a set of simulation scenarios adapted from Kuosmanen and Johnson (2010)—that the proposed method can outperform both DEA and CCNLS, even in settings that deviate from strict additivity. These results suggest that the method is capable of capturing certain types of curvature and nonlinearities, particularly when interactions are moderate or when the dimensionality is low. Nevertheless, the limitations identified in more complex scenarios provide a clear motivation for the methodological developments presented in the following chapter, where we introduce the Adaptive Constrained Enveloping Splines (ACES) technique—a more flexible and robust alternative designed to better accommodate interaction effects and complex frontier structures.

Beyond its core application, the proposed approach is versatile and can be applied to estimate production functions in a wide range of contexts, including banking, education, management, and public policy, among others. Beyond the estimation of production functions, the method can also be adapted to estimate revenue functions by using monetary output as the response variable and treating inputs as covariates. Moreover, production frontier estimation is closely related to the edge estimation problem, as discussed in the literature (see, for instance, Daouia et al. 2016). In this regard, Korostelev and Tsybakov (1993) provide foundational work on edge estimation within the field of image reconstruction. Other promising applications of the proposed methodology arise in disciplines where monotonicity and concavity are natural assumptions. In medicine, for example, the probability of contracting a disease may increase monotonically with respect to specific risk factors. In environmental studies, monotonic relationships are observed between ozone concentration and inversion base temperature, as shown in Croux et al. (2012). Finally, a novel and mathematically rich area of application lies in the inverse problem for Hamilton–Jacobi equations and semiconcave envelopes—see, for example, Esteve and Zuazua (2020)—where our approach could provide useful insights.

Chapter 5

Adaptive Constrained Enveloping Splines

Building on the methodology introduced earlier, this chapter presents a general framework that broadens the scope of shape-constrained regression techniques for efficiency analysis. While the additive Multivariate Adaptive Regression Splines (MARS)-based approach proved effective for estimating single-output production functions under monotonicity and concavity constraints, key limitations arise when applying it to more complex, real-world technologies. In practice, production settings often involve estimating multiple outputs, input interactions, and the need to generate efficiency scores compatible with Data Envelopment Analysis (DEA). These features call for more advanced modeling tools that capture complex relationships while respecting core economic assumptions. To address these challenges, we introduce the Adaptive Constrained Enveloping Splines (ACES) methodology, which significantly extends the original model. The three main methodological contributions of ACES are outlined below. This chapter draws extensively on the work published in España et al. (2025b), where the ACES methodology was first introduced and analyzed in detail.

The first improvement involves relaxing the conditions for satisfying monotonicity and concavity assumptions. In the original approach, these constraints were enforced additively, meaning that each univariate function within the estimator was independently required to comply with the shape restrictions. In contrast, our new approach enforces these constraints dimensionally, applying them across the (dimensional) aggregated function rather than its individual components. This refinement eliminates the need for all component functions to be individually monotonic and concave.

Second, we address a key limitation of the original additive framework: its inability to model interactions between variables. By introducing the capability to account for these interactions, our approach captures intricate, non-linear relationships present in the data, overcoming the restrictive additive structure of the original method. This improvement directly resolves one of the critical shortcomings identified in the earlier work, where cross-variable effects were insufficiently represented. Together with the relaxation of monotonicity and concavity constraints described earlier, this advancement considerably enhances the model's ability to accurately estimate the underlying data structure and account for the complexities of real-worlds production processes.

Third, we propose a novel procedure that enables the definition of a DEA-type technology within the MARS framework, thereby extending the methodology beyond the estimation of single-output production functions. This extension allows for the estimation of multi-input, multi-output technologies while preserving consistency with the shape constraints required by production theory. The procedure unfolds in three main steps. First, a machine learning (ML) technique adapted to the context of efficiency analysis is used to approximate the observed outputs to the underlying Data-Generating Process (DGP), ensuring that the sample is enveloped from above. Second, a refinement phase is applied in which selected predictions—particularly those deemed inaccurate—are conservatively replaced by the corresponding observed values. Finally, a DEA-type technology under variable returns to scale (VRS) is constructed by substituting the original output data with the refined predictions. This construction allows for the computation of standard efficiency measures while mitigating the overfitting issues typical of traditional DEA models.

5.1 A more relaxed approach

The additive MARS model for estimating production functions, as presented in Chapter 4, builds on the principle that the sum of non-decreasing monotonic and concave functions preserves these properties. Equation (4.18) imposes shape constraints on each reflected pair and unpaired left-sided basis function (BF), ensuring that all components individually satisfy monotonicity and concavity. While this ensures theoretical validity, it may overly constrain model flexibility, limiting its ability to capture the true data structure. To address this, we propose enforcing shape constraints at the dimensional level—that is, on the overall function with respect to each input—rather than on each BF. This enhances adaptability without violating theoretical assumptions. Accordingly, we reformulate Equation (4.18) as a sum of both right-side and left-side BFs, including paired and unpaired components (España et al. 2025b):

$$\hat{f}^{\text{ACES}}(\boldsymbol{x}) = \tau_0 + \sum_{j=1}^{m} \sum_{h \in H_j} \alpha_{j_h} \left(x_j - \kappa_{j_h}^{(R)} \right)_+ + \sum_{j=1}^{m} \sum_{w \in W_j} \beta_{j_w} \left(\kappa_{j_w}^{(L)} - x_j \right)_+ . \tag{5.1}$$

Here, x_j denotes the j-th input, $j = 1, \ldots, m$. The coefficients τ_0, α_{j_h}, and β_{j_w} correspond, respectively, to the intercept term, the coefficient of the h-th right-side BF and the coefficient of the w-th left-side BF for the j-th input, respectively. The knot locations for these BFs, denoted as $\kappa_{j_h}^{(R)}$ for right-side and $\kappa_{j_w}^{(L)}$ for left-side BFs, are selected from the set of input variables $\{x_{1j}, x_{2j}, \ldots, x_{nj}\}$. The sets H_j and W_j index the right-side and left-side BFs for the j-th input, respectively. In the enumeration of BFs, paired BFs are listed first, followed by unpaired BFs (right-side first, then left-side), with all knots ordered by increasing values. Finally, $\left(x_j - \kappa_{j_h}^{(R)}\right)_+$ and $\left(\kappa_{j_w}^{(L)} - x_j\right)_+$ form a reflected pair if $\exists\, h \in H_j$, $\exists\, w \in W_j$ such that the corresponding knot locations satisfy $\kappa_{j_h}^{(R)} = \kappa_{j_w}^{(L)}$.

Next, the set of knots selected by the left-side BFs for the j-th input is specified as $\kappa_j^{(L)*} = \kappa_j^{(L)}$, such that $\kappa_{j_w}^{(L)*} \in \kappa_j^{(L)}$, $\forall w \in W_j$. Besides, the set of selected knots associated with the right-side BFs for the j-th input is determined by excluding the knots already used by the left-side BFs, as $\kappa_j^{(R)*} = \kappa_j^{(R)} \setminus \left(\kappa_j^{(R)} \cap \kappa_j^{(L)}\right)$, such that $\kappa_{j_h}^{(R)*} \in \kappa_j^{(R)}$, $\forall h \in H_j$. This omission is necessary to avoid repeated values when the BFs form a reflected pair.

Thus, the set of knots, defined in (3.42), selected at any step of the algorithm is reformulated as follows:

$$\mathcal{K} = \left\{K_j : K_j = \kappa_j^{(R)*} \cup \kappa_j^{(L)*}\right\}_{j=1}^{m} = \left\{\left\{k_{1_1}, \ldots, k_{1_{|K_1|}}\right\}, \ldots, \left\{k_{m_1}, \ldots, k_{m_{|K_m|}}\right\}\right\}, \tag{5.2}$$

with $k_{j_1} < \cdots < k_{j_{|K_j|}}$.

At the end of the algorithm, each set K_j is expanded to include both the minimum and maximum observed values of the variable x_j, defined as $k_{j_0} = \min_{1 \le i \le n}(x_{ij})$ and $k_{j_{|K_j|+1}} = \max_{1 \le i \le n}(x_{ij})$, respectively. This ensures the creation of a collection of intervals $\left[k_{j_{v-1}}, k_{j_v}\right)$, for $v = 1, \ldots, |K_j| + 1$. Then, the new set of knots is defined as:

$$\mathcal{K}^* = \left\{k_{j_0} \cup K_j \cup k_{j_{|K_j|+1}}\right\}_{j=1}^{m}. \tag{5.3}$$

Under this new approach, our objective is to guarantee non-decreasing monotonicity and concavity within each interval $\left[k_{j_{v-1}}, k_{j_v}\right)$. To achieve this, we use the closed-form expression of the first-order partial derivatives determined from Equation (5.1) to establish estimation conditions on the coefficients that satisfy the shape requirements of the estimator. The idea is to gather all BFs that involve identical inputs, as expressed below:

$$\hat{f}^{\text{ACES}}(\boldsymbol{x}) = \tau_0 + \sum_{j=1}^{m} \hat{f}_j^{\text{ACES}}(\boldsymbol{x}) = \tau_0 + \sum_{j=1}^{m} \left(\sum_{h \in H_j} \alpha_{j_h} \left(x_j - \kappa_{j_h}^{(R)}\right)_+ + \sum_{w \in W_j} \beta_{j_w} \left(\kappa_{j_w}^{(L)} - x_j\right)_+ \right). \tag{5.4}$$

Hence, the j-th first-order partial derivative of $\hat{f}^{\text{ACES}}(\boldsymbol{x})$ is defined as a piecewise function:

$$\frac{\partial \hat{f}^{\text{ACES}}(\boldsymbol{x})}{\partial x_j} = \sum_{h \in H_j} \alpha_{j_h} \cdot \mathbb{I}(x_j > \kappa_{j_h}^{(R)}) - \sum_{w \in W_j} \beta_{j_w} \cdot \mathbb{I}(x_j < \kappa_{j_w}^{(L)}), \tag{5.5}$$

where $\mathbb{I}(\cdot)$ is an indicator function that takes the value 1 when its condition is met and 0 otherwise. These indicator functions determine the regions in the input space where a BF associated with some coefficient α_{j_h} or β_{j_w} is activated. Specifically, $(x_j - \kappa_{j_h}^{(R)})_+$ is activated in all intervals to the right-side of the knot $\kappa_{j_h}^{(R)}$, while $(\kappa_{j_w}^{(L)} - x_j)_+$ is activated in all intervals to the left-side of the knot $\kappa_{j_w}^{(L)}$. In view of this, we can define the j-th partial derivative of $\hat{f}^{\text{ACES}}(\boldsymbol{x})$ for the v-th interval $\left[k_{j_{v-1}}, k_{j_v}\right)$ as:

$$\left.\frac{\partial \hat{f}^{\text{ACES}}(\boldsymbol{x})}{\partial x_j}\right|_v = \sum_{h \in H_j} \alpha_{j_h} \cdot \mathbb{I}\left(k_{j_0} \leq \kappa_{j_h}^{(R)} < k_{j_v}\right) - \sum_{w \in W_j} \beta_{j_w} \cdot \mathbb{I}\left(k_{j_{v-1}} < \kappa_{j_w}^{(L)} \leq k_{j_{|K_j|+1}}\right). \tag{5.6}$$

From this point onward, we proceed to define the conditions required for ensuring both non-decreasing monotonicity and concavity properties of our estimator. Non-decreasing monotonicity is achieved by imposing that the estimated function increases in each interval $\left[k_{j_{v-1}}, k_{j_v}\right)$:

$$\left.\frac{\partial \hat{f}^{\text{ACES}}(\boldsymbol{x})}{\partial x_j}\right|_v \geq 0, \tag{5.7}$$

while concavity is guaranteed by imposing that the rate of growth decreases between two consecutive intervals $\left(\left[k_{j_{v-1}}, k_{j_v}\right), \left[k_{j_v}, k_{j_{v+1}}\right)\right)$:

$$\left.\frac{\partial \hat{f}^{\text{ACES}}(\boldsymbol{x})}{\partial x_j}\right|_v \geq \left.\frac{\partial \hat{f}^{\text{ACES}}(\boldsymbol{x})}{\partial x_j}\right|_{v+1}, \tag{5.8}$$

which is equivalent to:

$$-\sum_{h \in H_j} \alpha_{j_h} \cdot \mathbb{I}\left(k_{j_v} \leq \kappa_{j_h}^{(R)} < k_{j_{v+1}}\right) + \sum_{w \in W_j} \beta_{j_w} \cdot \mathbb{I}\left(k_{j_{v-1}} < \kappa_{j_w}^{(L)} \leq k_{j_v}\right) \geq 0. \tag{5.9}$$

We now present the Linear Programming (LP) model used to estimate the optimal values of the coefficients. Constraints (5.7) and (5.8) are applied for each interval v and each input dimension j. Additionally, an important improvement is introduced with respect to the model used in the previous chapter. In particular, we assign a different weight to each residual in the objective function to prioritize the accuracy of the fit for those units located closer to the production frontier. This is achieved by weighting the residuals by the inverse of the technical efficiency score ϕ_i, where $\phi_i \in [1, \infty)$ is the score of the i-th DMU obtained from the output-oriented radial DEA model defined in Equation (3.13):

$$\min_{\boldsymbol{\varepsilon},\,\tau_0,\,\boldsymbol{\alpha},\,\boldsymbol{\beta}} \sum_{i=1}^{n} \frac{1}{\phi_i} \cdot \varepsilon_i \tag{5.10}$$

subject to:

$$\hat{f}^{\mathrm{ACES}}(\boldsymbol{x}_i) - \varepsilon_i = y_i, \quad i = 1, \ldots, n, \tag{5.11}$$

$$\varepsilon_i \geq 0, \quad i = 1, \ldots, n, \tag{5.12}$$

$$\left.\frac{\partial \hat{f}^{\mathrm{ACES}}(\boldsymbol{x})}{\partial x_j}\right|_{v} \geq 0, \quad j = 1, \ldots, m, \quad v = 1, \ldots, |K_j| + 1, \tag{5.13}$$

$$\left.\frac{\partial \hat{f}^{\mathrm{ACES}}(\boldsymbol{x})}{\partial x_j}\right|_{v} - \left.\frac{\partial \hat{f}^{\mathrm{ACES}}(\boldsymbol{x})}{\partial x_j}\right|_{v+1} \geq 0, \quad j = 1, \ldots, m, \quad v = 1, \ldots, |K_j|, \tag{5.14}$$

where $\hat{f}^{\mathrm{ACES}}(\boldsymbol{x}_i) = \tau_0 + \sum_{j=1}^{m} \sum_{h \in H_j} \alpha_{j_h} \left(x_{ij} - \kappa_{j_h}^{(R)}\right)_+ + \sum_{j=1}^{m} \sum_{w \in W_j} \beta_{j_w} \left(\kappa_{j_w}^{(L)} - x_{ij}\right)_+$ is the optimal output estimation for DMU i.

Furthermore, the relaxation of the problem has a direct and significant impact on the design of the backward elimination algorithm. Unlike the methodology proposed in the previous chapter, where the removal of BFs was subject to strict structural constraints—specifically, right-sided BFs could only be removed if they were part of a reflected pair, and left-sided BFs were eligible for removal only when they appeared unpaired—the current approach simplifies this process considerably. Since the shape constraints are now enforced at the dimensional level and the feasibility region of the model is substantially broader, any BF can be evaluated for elimination regardless of its pairing status, as long as the resulting LP model remains feasible.

Finally, as an illustrative example, Figure 5.1 shows the partitioning of the input space and the corresponding first-order partial derivatives within each subspace for the following ACES estimator:

$$\begin{aligned}\hat{f}^{\mathrm{ACES}}(\boldsymbol{x}) = \tau_0 &+ \alpha_{1_1} (x_1 - 5)_+ + \beta_{1_1} (5 - x_1)_+ + \alpha_{1_2} (x_1 - 7)_+ + \beta_{1_2} (7 - x_1)_+ \\ &+ \alpha_{2_1} (x_2 - 3)_+ + \beta_{2_1} (3 - x_2)_+ .\end{aligned}$$

The estimator consists of three reflected pairs: the first two pairs correspond to the variable x_1, with knots located at 5 and 7, while the third pair is associated with the variable x_2, with a single knot at 3. This setup divides the input space into six distinct two-dimensional hyperrectangles. Each hyperrectangle is associated with specific partial derivative values, which are displayed within their respective regions in the following figure.

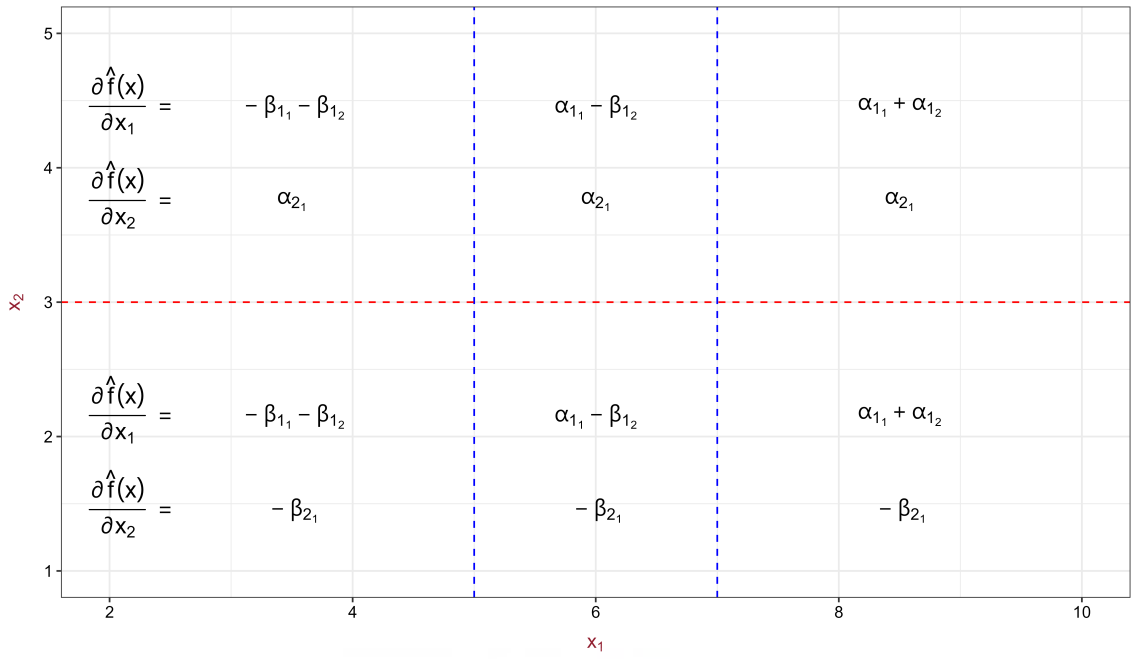

Figure 5.1: Visualization of first-order partial derivatives of the ACES estimator.

5.2 A model incorporating variable interaction

The methodology presented so far operates as a purely additive model, which can present challenges when estimating production functions that involve significant interactions between input variables. While the additive structure may initially appear limiting, simulations from Experimental Design 1 and the results discussed in the first chapter demonstrate that the method can outperform techniques such as DEA and Corrected Concave Non-parametric Least Squares (CCNLS), even when non-additive relationships are present. However, the performance of this approach ultimately depends on the degree of variable interactions and the extent of non-additivity in the true production function. For example, computational experiments using the production function $f(\boldsymbol{x}) = 0.1x_1 + 0.1x_2 + 0.3(x_1x_2)^{0.5}$ show that the method presented in Chapter 4 performs worse than DEA and CCNLS when the sample size reaches or exceeds 150 and 50, respectively.

If we review the literature on ML, we find that the standard (non-additive) MARS framework builds multivariate BFs of degree q by combining a new 1-degree univariate spline with previously introduced $(q-1)$-degree BFs. To minimize sensitivity to extreme values, each factor in a multivariate BF must involve distinct predictor variables. Furthermore, variable interactions follow a hierarchical structure, requiring a $(q-1)$-degree BF to be included in the model before a q-degree BF is introduced.

Finally, the interaction effect is local. For example, the 2-degree BF $(x_2 - 6)_+ \cdot (10 - x_1)_+$ captures interactions between x_1 and x_2 within the region where $x_2 > 6$ and $x_1 < 10$. However, the product of non-decreasing monotonic and concave functions does not necessarily yield a function with these properties. In this regard, Martinez et al. (2015) proposed alternative versions of MARS to build a convex (if f is a convex function, then $-f$ is a concave function) but non-monotonic estimator.

To overcome the aforementioned limitation, we adopt a simple yet effective strategy to incorporate variable interactions by expanding the original input space with multiplicative terms (España et al. 2025b). Let $x_1, \ldots, x_{m_1}$ denote the original set of inputs in $\mathbb{R}_+^{m_1}$. We define a transformation $\mathfrak{g} : \mathbb{R}_+^{m_1} \to \mathbb{R}_+^{m_1+m_2}$ that augments the input vector with multiplicative interactions among the inputs:

$$\mathfrak{g}(x_1, \ldots, x_{m_1}) = (x_1, \ldots, x_{m_1}, z_1, \ldots, z_{m_2}),$$

where each interaction term z_l, for $l = 1, \ldots, m_2$, is computed as

$$z_l = \prod_{j \in I_l} x_j,$$

with $I_l \subseteq \{1, \ldots, m_1\}$ and $2 \leq |I_l| \leq q_{\max}$, where $q_{\max}$ is a hyperparameter that controls the maximum order of interactions allowed in the transformation. Therefore, the total number of interaction terms is given by $m_2 = \sum_{a=2}^{q_{\max}} \binom{m_1}{a}$.

The inclusion of these artificial variables allows the model to account for interactions while preserving the shape constraints originally imposed on the inputs, since interaction terms are treated analogously to the original variables. Self-interactions are excluded, and although we do not enforce a strict hierarchical structure as in Friedman's original MARS procedure, our method prioritizes lower-degree BFs over higher-degree ones, following the strategy of Tsai and Chen (2005). To regulate this selection, we use a hyperparameter $\xi^{(q)}$, $q = 2, \ldots, q_{\max}$, which sets the minimum improvement threshold a q-degree BF must achieve relative to the best one-degree BF to be incorporated. This ensures that only interactions with substantial contribution are retained.

Following this procedure, the input space expands from m_1 to $m_1 + m_2$ through the transformation $\mathfrak{g}$. Although shape constraints are applied to the estimated production function $\hat{f}^{\text{ACES}}$ in the extended space, these constraints may not hold when considering the original inputs alone. To address this, we project the predicted output $\hat{f}^{\text{ACES}}(\mathfrak{g}(x))$ back onto the original space by applying a standard DEA-VRS model. This ensures that the final frontier estimation satisfies free disposability and convexity in the original m_1-dimensional input space. In this way, the estimated frontier integrates variable interaction effects into the prediction of optimal output while preserving the axiomatic properties of the original domain (see Section 5.3).

Finally, Figure 5.2 presents a comparative example showing the improvements achieved by the proposed approach over the linear model introduced in the previous chapter, thanks to the inclusion of variable interactions and the relaxed model formulation. As can be observed, this leads to better approximations to the DGP.

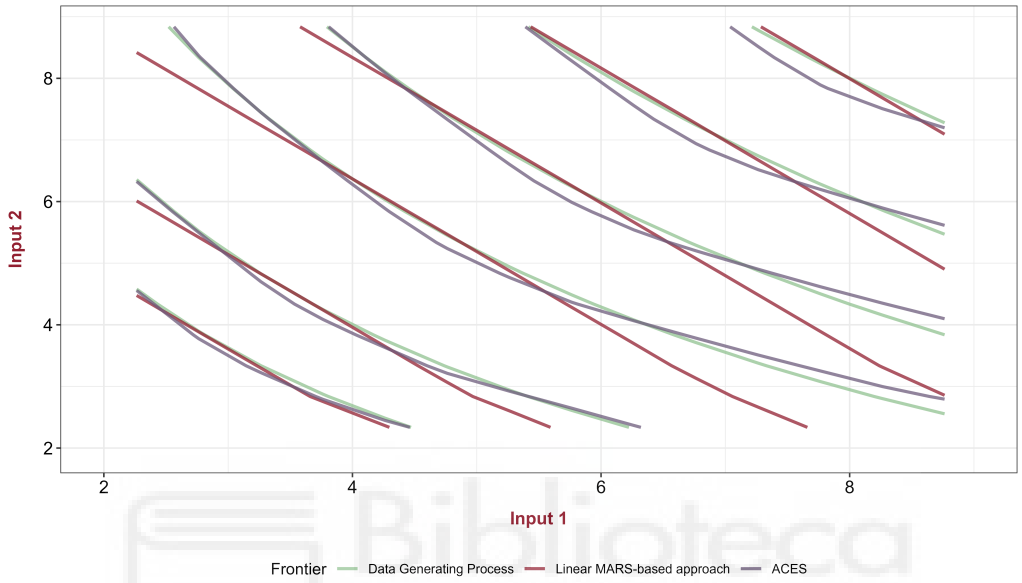

Figure 5.2: Contour plot comparing ACES performance with its previous version.

5.3 Estimating a DEA-type technology with ACES

In this section, we extend the previously described methodology, which estimates production functions under shape constraints such as concavity and monotonicity, to accommodate a multi-output production framework (España et al. 2025b). This extension defines an estimator for the production technology using a piecewise linear frontier that satisfies these constraints, offering an alternative to traditional DEA in scenarios with multiple inputs and outputs, which are very usual in practice[1].

Consider a set of n DMUs, where each unit i is characterized by an observed input-output pair $(\boldsymbol{x}_i, \boldsymbol{y}_i)$ randomly drawn from a true population. Our approach aims to identify optimal input-output combinations $(\boldsymbol{x}_i^*, \boldsymbol{y}_i^*)$, $i = 1, \ldots, n$, which define the true efficient frontier within the production space. This process is carried out through a structured 3-stage procedure designed to improve alignment with the underlying production capabilities.

[1] All developments in this section assume an input space $\mathbb{R}_+^m$ based on the original variables, but the methodology remains valid when applied to an extended space $\mathbb{R}_+^{m_1+m_2}$ that includes interaction terms, as long as the final technology is projected onto the original space.

In the first stage, the objective is to estimate the optimal outputs ($\boldsymbol{y}_i^*$) by determining approximations ($\hat{\boldsymbol{y}}_i$) that better reflect the true production capabilities of each DMU. This stage operates under the assumption that the observed inputs are optimal, i.e., $\boldsymbol{x}_i = \boldsymbol{x}_i^*$, indicating that there is no inefficiency or random error in input measurements. This step generates a new dataset $\{(\boldsymbol{x}_i^*, \hat{\boldsymbol{y}}_i)\}_{i=1}^n$, representing a projection of the observed data that lies closer to the true efficient frontier and is therefore better aligned with the underlying production process.

In the second stage of the method, we address potential overestimations in the predicted output values. Specifically, if any component of the estimated output vector is suspected to be overestimated, i.e., $\hat{y}_{ij} > y_{ij}^*$ for some j, it can be corrected by replacing $\hat{y}_{ij}$ with the corresponding observed value y_{ij}. This conservative approach is proposed because it is sufficient to project a few units to the true frontier to achieve a good characterization of the technology. Additionally, any estimate that falls between the observed value y_{ij} and the true (but unknown) optimal value y_{ij}^* improves DEA's performance under no random error. However, an overestimation of $\hat{y}_{ij}$ beyond y_{ij}^* could degrade the accuracy of the method. Therefore, correcting overestimations maintains the reliability of the resulting technology.

As a final step in the methodology, we construct a standard DEA-VRS technology using the refined version of $\{(\boldsymbol{x}_i^*, \hat{\boldsymbol{y}}_i)\}_{i=1}^n$. This stage offers two significant advantages. First, it provides the flexibility to relax monotonicity and concavity constraints during the initial stage if needed. It is important to note that the first step is primarily focused on adjusting the observed data to better align with the true underlying production technology, rather than directly estimating it. By initially relaxing these constraints, the model can capture data complexities more effectively in some scenarios, while in others, maintaining the constraints may yield more accurate results. However, even if all the form constraints were imposed in the first stage, the correction phase could break with the proper form of the technology. For these reasons, constraints are rigorously re-imposed in the final step when constructing the DEA-VRS technology, ensuring that the final estimated technology satisfies all necessary shape restrictions. Thus, the first two stages effectively "push" the observed data toward the true DGP ensuring $\boldsymbol{y}_i \leq \hat{\boldsymbol{y}}_i$, $\forall i = 1, \ldots, n$, with the full enforcement of form constraints occurring in this last step. Finally, the second advantage lies in that, resorting to the DEA-VRS technology, it is easy to measure technical inefficiency through any measure established within the standard DEA framework.

We begin by adapting the iterative procedure from the original single-output to a multi-output context. While the formal application of MARS for estimating multiple response variables is limited, Milborrow (2014) describes a method for handling multiple outputs within a standard MARS model, which is implemented in the R `earth` package (Milborrow et al. 2017). In this approach, while the same set of BFs is applied across all models, the coefficients differ for each response variable. Next, during the backward algorithm, the usual procedure is followed, but with the minimization of the overall generalized cross-validation (GCV) score in Equation (3.40) across all response variables.

This formulation is a direct extension of the LP model defined in Equation (5.10) to the multi-output case. For each output $r = 1, \ldots, s$, we simultaneously estimate a shape-constrained production frontier while minimizing deviations from the observed values. The resulting piecewise linear estimators preserve key properties of production technologies and mitigate the risk of overfitting.

$$\min_{\boldsymbol{\varepsilon}, \boldsymbol{\tau}_0, \boldsymbol{\alpha}, \boldsymbol{\beta}} \sum_{r=1}^{s} \sum_{i=1}^{n} \frac{1}{\phi_i} \cdot \varepsilon_{ir} \tag{5.15}$$

subject to:

$$\hat{f}_r^{\text{ACES}}(\boldsymbol{x}_i) - \varepsilon_{ir} = y_{ir}, \quad r = 1, \ldots, s, \quad i = 1, \ldots, n, \tag{5.16}$$

$$\varepsilon_{ir} \geq 0, \quad r = 1, \ldots, s, \quad i = 1, \ldots, n, \tag{5.17}$$

$$\left.\frac{\partial \hat{f}_r^{\text{ACES}}(\boldsymbol{x})}{\partial x_j}\right|_v \geq 0, \quad r = 1, \ldots, s, \quad j = 1, \ldots, m, \quad v = 1, \ldots, |K_j| + 1, \tag{5.18}$$

$$\left.\frac{\partial \hat{f}_r^{\text{ACES}}(\boldsymbol{x})}{\partial x_j}\right|_v - \left.\frac{\partial \hat{f}_r^{\text{ACES}}(\boldsymbol{x})}{\partial x_j}\right|_{v+1} \geq 0, \quad r = 1, \ldots, s, \quad j = 1, \ldots, m, \quad v = 1, \ldots, |K_j|, \tag{5.19}$$

where

$$\hat{f}_r^{\text{ACES}}(\boldsymbol{x}_i) = \tau_0^{(r)} + \sum_{j=1}^{m} \sum_{h \in H_j} \alpha_{j_h}^{(r)} \left(x_{ij} - \kappa_{j_h}^{(R)}\right)_+ + \sum_{j=1}^{m} \sum_{w \in W_j} \beta_{j_w}^{(r)} \left(\kappa_{j_w}^{(L)} - x_{ij}\right)_+ \tag{5.20}$$

is the estimated value of output r for DMU i. Each output r is associated with its own intercept $\tau_0^{(r)}$, slope coefficients $\alpha_{j_h}^{(r)}$ and $\beta_{j_w}^{(r)}$, while the knot sets $\kappa_{j_h}^{(R)}$, $\kappa_{j_w}^{(L)}$, and index sets H_j, W_j are shared across outputs. The piecewise partial derivative of output r with respect to input j in interval v (Constraint 5.18) is given by:

$$\sum_{h \in H_j} \alpha_{j_h}^{(r)} \cdot \mathbb{I}\left(k_{j_0} \leq \kappa_{j_h}^{(R)} < k_{j_v}\right) - \sum_{w \in W_j} \beta_{j_w}^{(r)} \cdot \mathbb{I}\left(k_{j_{v-1}} < \kappa_{j_w}^{(L)} \leq k_{j_{|K_j|+1}}\right) \tag{5.21}$$

and its difference across adjacent intervals (Constraint 5.19) is:

$$-\sum_{h \in H_j} \alpha_{j_h}^{(r)} \cdot \mathbb{I}\left(k_{j_v} \leq \kappa_{j_h}^{(R)} < k_{j_{v+1}}\right) + \sum_{w \in W_j} \beta_{j_w}^{(r)} \cdot \mathbb{I}\left(k_{j_{v-1}} < \kappa_{j_w}^{(L)} \leq k_{j_v}\right) \tag{5.22}$$

The separability of Model (5.15) implies that each output $r = 1, \ldots, s$ is estimated independently, conditional on a common set of BFs. While this decomposition into s parallel subproblems enhances computational tractability and enables output-specific flexibility, it introduces a structural limitation: the model does not account for the joint distribution of outputs. Specifically, there is no mechanism in place to ensure that the vector of predicted outputs $(\hat{y}_{i1}, \ldots, \hat{y}_{is})$ remains consistent with the joint production structure implied by the production technology.

This issue becomes particularly evident when a DMU's output vector is jointly efficient, but individual components of the vector are estimated from marginal models that fail to recognize this joint optimality. Let $(\boldsymbol{x}_i, \boldsymbol{y}_i) \in \mathbb{R}^m_+ \times \mathbb{R}^s_+$ denote the observed input–output combination of a given DMU. Suppose that $\boldsymbol{y}_i = \boldsymbol{y}_i^*$ is technically efficient under the true production technology $\varphi \subseteq \mathbb{R}^m_+ \times \mathbb{R}^s_+$, meaning that $(\boldsymbol{x}_i, \boldsymbol{y}_i) \in \partial(\varphi)$. In this case, the output vector $\boldsymbol{y}_i$ is maximal for the input bundle $\boldsymbol{x}_i$, and no additional output can be produced without increasing some input or violating feasibility. Now consider the estimator $\hat{\boldsymbol{y}}_i = (\hat{y}_{i1}, \ldots, \hat{y}_{is})$ obtained by solving s independent subproblems as in Model (5.15). Each component $\hat{y}_{ir}$ is constructed from a marginal envelopment model (Constraints 5.16 and 5.17) that seeks to approximate the true frontier for output r, conditional on $\boldsymbol{x}_i$, and subject to the monotonicity and concavity constraints. Since these subproblems operate independently, they do not preserve the joint feasibility of the original vector $\boldsymbol{y}_i^*$. Suppose now that the marginal model for output r leads to a significant overestimation, so that $\hat{f}_r^{\text{ACES}}(\boldsymbol{x}_i) \gg y_{ir}^*$. Then, even though the vector $(\boldsymbol{x}_i, \boldsymbol{y}_i)$ is efficient with respect to the true joint technology, the marginal model for output r will indicate that the unit is far below its estimated frontier for that dimension. As a result, the optimization procedure will suggest that the DMU should increase output r to align with the marginal frontier. This correction is misguided: it artificially inflates $\hat{y}_{ir}$, yielding a prediction $\hat{y}_{ir} \gg y_{ir}^*$, and ultimately produces a distorted estimation of the joint production capacity of the unit.

To illustrate this situation, consider a dataset of 50 DMUs, each characterized by two inputs (x_1, x_2) and two outputs (y_1, y_2), generated according to the methodology outlined by Perelman and Santín (2009) (Experimental Design 2) that meets usual microeconomic postulates. Additionally, the maximum output producible for a given set of inputs is denoted as (y_1^*, y_2^*). Data are generated without random error. Subsequently, we estimate these optimal outputs, yielding $(\hat{y}_1, \hat{y}_2)$. To further evaluate the method, three additional columns are included in the table below, representing the output-oriented radial score calculated by standard DEA, as defined by Banker et al. (1984). Specifically, $\phi^{(r)}$ is the output-oriented radial score utilizing all inputs and only the r-th output, while ϕ is the output-oriented radial score utilizing all the available variables. The following table presents data for four representative units, highlighting the performance of our approach:

Table 5.1: Illustrative example of the performance of our methodology for four selected units.

Index	x_1	x_2	y_1	y_2	y_1^*	y_2^*	$\hat{y}_1$	$\hat{y}_2$	$\phi^{(1)}$	$\phi^{(2)}$	ϕ
4	39.96	15.47	161.77	459.22	189.98	539.29	600.71	572.42	3.59	1.03	1.03
22	32.34	40.57	667.48	811.30	675.93	821.57	765.65	880.22	1.15	1.00	1.00
23	38.30	14.85	446.22	190.38	563.14	240.27	558.46	529.54	1.22	2.36	1.22
49	17.41	12.26	143.98	173.14	233.85	281.22	255.62	222.19	1.97	1.72	1.42

Table 5.1 reveals that for the resource levels of $x_1 = 39.96$ and $x_2 = 15.47$, unit 4 needs to produce $y_1^* = 189.98$ and $y_2^* = 539.29$ to achieve technical efficiency. Similarly, unit 23, with resources $x_1 = 38.30$ and $x_2 = 14.85$, should aim to produce $y_1^* = 563.14$ and $y_2^* = 240.27$. In these cases, production should be mainly concentrated in y_2 or y_1, respectively. Conversely, for resource levels presented in units 22 and 49, the optimal production levels are similar for both outputs. Additionally, we can evaluate each unit's relative position with respect to single-output and multi-output analyses by using the scores $\phi^{(1)}$, $\phi^{(2)}$, and ϕ. For example, units 22 and 49 have similar positions relative to the frontier in both single-output and multi-output contexts. However, unit 4 is significantly farther from the frontier when only output y_1 is considered compared to when both outputs are accounted for. A similar situation occurs for output y_2 and unit 23. The ACES model estimates $(\hat{y}_1, \hat{y}_2)$ were generated using Model (5.15). The results show that prediction for both outputs are similar across all four units, exhibiting poor performance in estimating y_1^* for unit 4 and y_2^* for unit 23. Precisely, these units exhibit greater discrepancies when comparing single-output and multi-output analyses. This situation underscores the importance of the refinement step, where replacing the predicted values with observed values could significantly enhance the technique's performance.

When applying an ACES model in a multi-output context, the model's accuracy is highly dependent on its ability to predict each output variable effectively. Poor predictions, even for a single output, can significantly distort the overall efficiency assessment of DMUs. In contrast, projecting only a few units onto the frontier can achieve a good characterization of the technology. Adopting an approach where only a portion of the observed output is substituted, while preserving the remainder, has proven beneficial in improving the model performance, as we will show. As illustrated in Table 5.1, the ACES model performs well when the DMU's relative position remains consistent across both single-output and multi-output analyses, thereby accurately capturing the DMU's maximum production capacity. Considering these factors, we implement the following strategy to refine the ACES estimates:

$$\hat{y}_{ir} = \begin{cases} \hat{f}_r^{\text{ACES}}(\boldsymbol{x}_i), & \left|\phi^{(r)} - \phi\right| < \psi \\ y_{ir}, & \left|\phi^{(r)} - \phi\right| \geq \psi \end{cases}, \quad i = 1, \ldots, n, \quad r = 1, \ldots, s, \tag{5.23}$$

where $\phi^{(r)}$ is the DEA-based output-oriented radial score utilizing all inputs and only the r-th output, ϕ is the DEA-based output-oriented radial score utilizing all the available variables, $\hat{f}_r^{\text{ACES}}(\boldsymbol{x}_i)$ is the r-th output prediction for the i-th DMU by an ACES model, and ψ is a threshold that prioritizes either predicted output $\hat{f}_r^{\text{ACES}}(\boldsymbol{x}_i)$ or observed output y_{ir} to constitute $\{(\boldsymbol{x}_i^*, \hat{y}_i)\}_{i=1}^n$. Following our experience, a threshold value of $\psi = 0.05$ generally performs well across different scenarios, while a value of $\psi = 0$ corresponds to the standard DEA technology. Notably, in the case of considering a single output all observed units are replaced by the predicted units.

Once the refined outputs $\hat{\boldsymbol{y}}_i$ are obtained, the ACES-based technology is defined as the convex hull of the sample $\{(\boldsymbol{x}_i, \hat{\boldsymbol{y}}_i)\}_{i=1}^n$, imposing the standard VRS constraint. Formally, we define:

$$\hat{\varphi}_{\text{ACES}} = \left\{ (\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}_+^{m+s} : \begin{array}{ll} x_j \geq \sum_{i=1}^n \lambda_i x_{ij}, & j = 1, \ldots, m, \\ y_r \leq \sum_{i=1}^n \lambda_i \hat{y}_{ir}, & r = 1, \ldots, s, \\ \sum_{i=1}^n \lambda_i = 1, & \\ \lambda_i \geq 0, & i = 1, \ldots, n \end{array} \right\}. \tag{5.24}$$

Several aspects merit consideration when comparing this new DEA-type technology with the well-established DEA-VRS technology ($\hat{\varphi}_{\text{DEA}}$) defined in Equation (3.10). In the standard DEA, the technology construction hinges on historical data: observed input–output combinations form the core dataset, and new virtual productions are added based on assumptions such as convexity, free disposability, and minimal extrapolation. Traditional DEA addresses the question of 'What other input–output combinations can be guaranteed as producible based on the observed units?' assuming that some efficient DMUs are always observed. In contrast, the ACES framework relaxes this assumption. Rather than relying solely on observed data, it aims to build a (more realistic) technology by including input–output combinations that improve upon the observed ones. The primary dataset is thus modified to include a virtual sample generated by the ACES model, which preserves the the m original input vectors and replaces the outputs with a set of r (new) "pushed-up" output vectors.

Convexity and free disposability of inputs and outputs in Equation (5.24) are easily verified. It can also be proved that $\hat{\varphi}_{\text{DEA}} \subseteq \hat{\varphi}_{\text{ACES}}$. The minimal extrapolation assumption is not imposed; instead, our approach positions the frontier as closely as possible to the new virtual data sample. In this way, the issue of overfitting can be addressed. Below, we summarize the main advantages of constructing the final ACES technology as formulated in Equation (5.24):

- It ensures that the estimated production frontier satisfies the shape constraints in the original $\mathbb{R}^m$ input space, even after incorporating variable interactions through artificial covariates (via the transformation $\mathfrak{g}$ described in Section 5.2). Although estimation is performed in the extended space, the projection using the DEA-VRS model guarantees theoretical consistency. When interactions are included, shape constraints are also enforced on the artificial covariates through Equations (5.18) and (5.19), while the refined outputs are computed using the extended-input prediction $\hat{f}_r^{\text{ACES}}(\mathfrak{g}(\boldsymbol{x}_i))$ in Equation (5.23). On the other hand, the inputs used in the final technology defined in Equation (5.24) remain the original variables.

- It allows the replacement of potentially overestimated output components with their observed values during the refinement phase. This conservative correction reduces the risk that the projected frontier exceeds the true production capacity of any DMU, preserving the reliability of the resulting technology. Since only a subset of units needs to be projected onto the frontier to ensure an accurate characterization of the production set, this refinement strategy improves robustness without requiring perfect prediction for all outputs.

- It allows the relaxation of monotonicity and concavity constraints during the initial estimation phase when needed. This flexibility also applies to the artificial covariates introduced through variable interactions. Since the final DEA-VRS construction reimposes these shape restrictions on the original input space, the model preserves economic coherence even if such constraints are relaxed to improve predictive performance during estimation. However, the envelopment condition must still be enforced throughout the process, as the estimated optimal output should not fall below the observed value in the absence of random noise. This ensures consistency with the foundational assumption that observed outputs represent feasible production levels. This flexibility is further discussed in Subsection 5.6.2.

5.4 Measuring technical efficiency through ACES

Expression (5.24) defines a technology under VRS that is separated from the observed data set by eliminating the minimum extrapolation axiom. Next, we show how to measure the efficiency of a DMU with input–output bundle $(\boldsymbol{x}_0, \boldsymbol{y}_0)$ using $\hat{\varphi}_{\text{ACES}}$, depending on the type of measure considered (see, for example, Pastor et al., 2012).

The output-oriented radial measure (Banker et al. 1984):

$$\max_{\phi,\boldsymbol{\lambda}} \phi \tag{5.25}$$

subject to:

$$\begin{aligned}
\sum_{i=1}^{n} \lambda_i x_{ij} &\leq x_{0j}, \quad j = 1, \ldots, m, \\
\sum_{i=1}^{n} \lambda_i \hat{y}_{ir} &\geq \phi y_{0r}, \quad r = 1, \ldots, s, \\
\sum_{i=1}^{n} \lambda_i &= 1, \\
\lambda_i &\geq 0, \quad i = 1, \ldots, n.
\end{aligned}$$

The input-oriented radial measure (Banker et al. 1984):

$$\min_{\theta,\boldsymbol{\lambda}} \theta \tag{5.26}$$

subject to:

$$\begin{aligned}
\sum_{i=1}^{n} \lambda_i x_{ij} &\leq \theta x_{0j}, \quad j = 1, \ldots, m, \\
\sum_{i=1}^{n} \lambda_i \hat{y}_{ir} &\geq y_{0r}, \quad r = 1, \ldots, s, \\
\sum_{i=1}^{n} \lambda_i &= 1, \\
\lambda_i &\geq 0, \quad i = 1, \ldots, n.
\end{aligned}$$

The Directional Distance Function (Chambers et al. 1998):

$$\max_{\omega,\boldsymbol{\lambda}} \omega \tag{5.27}$$

subject to:

$$\begin{aligned}
\sum_{i=1}^{n} \lambda_i x_{ij} &\leq x_{0j} - \omega g_{x_j}, \quad j = 1, \ldots, m, \\
\sum_{i=1}^{n} \lambda_i \hat{y}_{ir} &\geq y_{0r} + \omega g_{y_r}, \quad r = 1, \ldots, s, \\
\sum_{i=1}^{n} \lambda_i &= 1, \\
\lambda_i &\geq 0, \quad i = 1, \ldots, n.
\end{aligned}$$

Here, $(\boldsymbol{g}_x, \boldsymbol{g}_y) = (g_{x_1}, \ldots, g_{x_m}, g_{y_1}, \ldots, g_{y_s})$ represents a directional projection vector to the frontier, where $(\boldsymbol{x}_0, \boldsymbol{y}_0) + \omega(\boldsymbol{g}_x, \boldsymbol{g}_y)$, with $\omega \geq 0$ intersects the frontier.

Additionally, other well-known efficiency measures can be determined in a similar manner. Additive models such as the Measure of Inefficiency Proportions (Cooper et al., 1999), the Normalized Weighted Additive Model (Lovell and Pastor, 1995), the Range Adjusted Measure (Cooper et al., 1999), and the Bounded Adjusted Measure (Cooper et al., 2011) can be applied. Likewise, input- and output-oriented non-radial Russell measures (Färe and Lovell, 1978; Färe et al., 1985) and the Enhanced Russell Graph Measure (Pastor et al., 1999) are also compatible with this approach, among others.

5.5 Some considerations about the smoothing procedures

In Section 4.3, we introduced two alternative procedures to smooth the function resulting from the adaptation of MARS to production function estimation. Although these procedures have not been explicitly reformulated within the methodological framework developed in this chapter, they remain conceptually applicable in the single-output setting when monotonicity and concavity constraints are enforced. In particular, two scenarios must be taken into account when applying these smoothing procedures in the current framework, as they affect both the feasibility and the validity of the resulting model:

- Interaction of variables. When interaction terms are included in the model, estimation is performed in an extended feature space $\mathbb{R}^{m_1+m_2}$, and the final production technology must be reconstructed in $\mathbb{R}^m$ using Equation (5.24). In this case, the resulting function is smooth in the transformed domain, where the spline BFs are defined. However, this smoothness does not necessarily carry over to the original input space $\mathbb{R}^m$, where the final model represents a different functional form. As a result, the production set obtained in the original space may not inherit the smoothness of the transformed domain and its regularity properties remain generally unknown.

- Model structure for smoothing. To ensure that the smoothing procedure preserves the theoretical properties of the estimated function, the model must adopt a structure compatible with the one introduced in the previous chapter. Specifically, the BF set must consist solely of reflected pairs and unpaired left-sided BFs, as defined in Equation (4.18). This structural constraint is crucial for guaranteeing that the shape restrictions imposed during estimation remain valid after smoothing. Consequently, coefficient estimation must be carried out using the LP introduced in Equation (4.15), rather than the formulation defined in Equation (5.10). Nevertheless, the objective function can still be modified to assign greater weight to observations near the production frontier, thereby prioritizing the accurate estimation of efficient units.

Regarding the second point, our proposed strategy applies the smoothing procedure to all candidate submodels that conform to this structural requirement. It is worth noting that the forward model always satisfies this condition, as it consists entirely of reflected BF pairs. For each admissible submodel, we recompute the GCV score after applying the smoothing transformation, allowing a fair comparison across candidates. The final smoothed model is then selected as the one minimizing the updated GCV. Importantly, the selected model is not necessarily a smoothed version of the submodel obtained via the backward algorithm, but may instead correspond to an entirely different candidate that better balances smoothness and compliance with shape constraints.

The following procedure describes how to apply the smoothing strategy under the single-output ACES framework, assuming that monotonicity and concavity are enforced in the first-stage estimation:

Step 1 [Select a candidate submodel]: Choose a submodel expressed in the format defined by Equation (4.18), ensuring that it satisfies the structural conditions for shape-constrained smoothing.

Step 2 [Construct knot triplets]: For each surviving central knot identified in Step 1, construct a triplet $\mathcal{T}(k) = \{k^-, k, k^+\}$, ensuring that the spacing conditions required to preserve monotonicity and concavity are met. See Sections 4.3.1 and 4.3.2 for implementation details.

Step 3 [Build smooth BFs]: Construct the corresponding set of smoothed BFs—either cubic $B_C^{\pm}$ or quintic $B_Q^{\pm}$—using the definitions provided in Equations (4.26)–(4.27) and (4.32)–(4.33), respectively.

Step 4 [Estimate ACES coefficients]: Solve the LP model defined in Equation (4.15) to estimate the ACES coefficients using the smoothed predictor defined in Equation (4.20).

Step 5 [Iterate over candidate submodels]: Repeat Steps 1–4 for all shape-conforming submodels generated by the backward selection algorithm.

Step 6 [Select the optimal smoothed model]: Choose the smoothed submodel that minimizes the updated GCV score defined in Equation (3.40).

Step 7 [Construct the final technology]: Use the smoothed model to generate output predictions for all units. Then, based on the refined dataset, construct the final ACES technology as defined in Equation (5.24).

5.6 Computational experiments and hyperparameter tuning

This section is divided into four subsections. The first introduces three distinct sets of computational experiments, based on the simulation frameworks proposed by Kuosmanen and Johnson (2010) (Experimental Design 1), Perelman and Santín (2009) (Experimental Design 2), and Färe et al. (1994) (Experimental Design 3). Each design addresses a different set of research questions and contributes to evaluating the methodological contributions of ACES under varied assumptions and data conditions. Experimental Design 1 focuses on a single-output setting and is specifically intended to examine the contribution of variable interactions and the impact of relaxing shape constraints during coefficient estimation. No benchmarking against alternative methods is performed in this case, as the goal is to conduct an internal evaluation of the modeling choices. By contrast, Experimental Designs 2 and 3 are employed to assess the model's goodness of fit in more realistic multi-output settings, where outputs may vary in scale, structure, and noise levels. These two designs are characterized by three key parameters: the sample size, the proportion of truly efficient units located on the production frontier,

and the presence or absence of random noise. In noisy scenarios, a subset of observations may lie above the true frontier, introducing stochastic variability that challenges the estimator's robustness. These experiments are used to evaluate the performance of the proposed ACES methodology relative to several well-established techniques, including DEA (Banker et al., 1984), Stochastic Non-Smooth Envelopment of Data (StoNED) (Kuosmanen and Johnson, 2017), Convexified Efficiency Analysis Trees (CEAT) (Esteve et al., 2020), and Bootstrap DEA (BDEA) (Simar and Wilson, 1998, 2000).

Building on the multi-output scenarios described above, the evaluation in Experimental Designs 2 and 3 involves estimating the true radial output score (ϕ). To this end, we employ the available data to estimate ϕ using different techniques, denoted as $\hat{\phi}^{\text{ACES}}$, $\hat{\phi}^{\text{DEA}}$, $\hat{\phi}^{\text{StoNED}}$, $\hat{\phi}^{\text{CEAT}}$, and $\hat{\phi}^{\text{BDEA}}$. The $\hat{\phi}^{\text{ACES}}$ score is obtained using the methodology outlined in this chapter, with Model (5.25) used as the final step. The $\hat{\phi}^{\text{DEA}}$ score is computed following the standard procedure described in Banker et al. (1984). The $\hat{\phi}^{\text{StoNED}}$ score is derived using the StoNED method, implemented via the Python package `pyStoNED` as described by Dai et al. (2024). Specifically, to perform StoNED, we use the directional projection vector to the frontier $(\boldsymbol{g}_x, \boldsymbol{g}_y) = (0, 0, \sigma_{y_1}, \sigma_{y_2})$, where σ_{y_1} and σ_{y_2} denote the standard deviation of outputs y_1 and y_2, respectively (see Kuosmanen and Johnson 2017). The $\hat{\phi}^{\text{CEAT}}$ score is calculated using the convexified version of EAT, which replaced the original stepwise frontier with a piecewise linear production function. To implement this, we rely on the `eat` R package (Esteve et al., 2022), using its default configuration of five units per terminal node to prevent excessive splitting. Finally, the $\hat{\phi}^{\text{BDEA}}$ score is obtained by applying the Bootstrap DEA method using the `Benchmarking` R package (Bogetoft et al., 2015) under VRS and 1,000 bootstrap replications.

The second subsection expands on the set of hyperparameters introduced in Section 4.4.2, providing a broader configuration space for the ACES methodology. It specifies the values adopted for each hyperparameter under the two multi-output experimental designs described previously. In addition, it offers a detailed procedure for configuring ACES in general, providing practical guidance on how to tune the method for any given application scenario.

The third subsection presents the results obtained from the simulations. For each combination of scenario, 100 independent trials were conducted to evaluate the relative performance of the methods. Three evaluation criteria are considered: MSE, bias, and computational time. The MSE captures the average magnitude of the estimation error by measuring the squared differences between the estimated and true radial output scores. Bias quantifies the average direction of the error, indicating whether a method systematically overestimates or underestimates the true frontier. Computational time, in turn, provides a practical assessment of the algorithm's efficiency, reflecting its applicability to large-scale or time-sensitive problems. Together, these metrics offer a comprehensive assessment of each model's accuracy, estimation behavior, and computational performance.

5.6.1 Experimental setup

Experimental Design 1 adopts the same sample sizes used in Section 4.4—namely, 50, 100, and 150 DMUs—with the aim of evaluating the methodological improvements introduced in this chapter for the case of single-output estimation. In contrast, Experimental Designs 2 and 3 are both structured around a two-input, two-output framework and share a unified design defined by three key parameters: (i) the sample size, (ii) the proportion of DMUs located on the true production frontier (0%, 5%, 10%, and 20%), and (iii) the presence or absence of random noise. In noise-free scenarios, observations may lie below the true frontier due to inefficiency, whereas in noisy scenarios, random deviations allow observations to exceed the true frontier. Experimental Design 2 considers sample sizes of 50, 100, 150, 200, and 300, and is based on the multi-output Translog production function proposed by Perelman and Santín (2009). Experimental Design 3, adapted from Färe et al. (1994), evaluates sample sizes of 25, 75, 125, 175, and 250, and employs a CET-CD specification in which returns to scale vary across producer groups. Full implementation details for both designs—including functional forms and procedures for simulating inefficiency and noise—are provided in Section 3.4.1.

5.6.2 Hyperparameter specification and tuning procedure

This section extends the set of hyperparameters previously introduced in Table 4.1, which includes core tuning parameters such as the maximum number of BFs (η) to include into the model, the minimum relative improvement required to retain a new BF (ξ), and the penalty factor in the backward phase (d), among others. Specifically, we introduce five additional hyperparameters that play a key role in configuring the ACES model under this extended framework:

Table 5.2: Additional hyperparameters to configure the ACES model.

Hyperparameter	**Description**
$q_{\max}$	Maximum degree of variable interaction during the forward algorithm. Higher values significantly increase the computational cost.
$\xi^{(q)}$	Minimum improvement ratio over the best 1-degree BF required to include a q-th degree BF in the model.
ψ	Threshold that determines whether the predicted output or the observed value is retained in the refinement step to construct the technology set defined in Equation (5.24).
`monotonicity`	Logical flag indicating whether the monotonicity constraint defined in Equation (5.18) is imposed during the first stage.
`concavity`	Logical flag indicating whether the concavity constraint defined in Equation (5.19) is imposed during the first stage.

For standard scenarios, the following hyperparameter values are recommended for testing. For η, multiples of 10 should be considered, up to the total number of observations in the dataset. Regarding ξ, typical values could range from 0, 0.005, 0.01, 0.02 or even 0.05 if the sample size is particularly large. For $q_{\max}$, it is advisable to test values of 1 or 2, and in cases where the number of features allows, 3 can be considered—although higher values may lead to excessive expansion of the feature space and increased computational burden. For $\xi^{(q)}$, values such as 0, 0.05, 0.10, 0.20, or 0.50 could be tested. However, when $\xi^{(q)}$ is close to 0.50, the model often defaults to a first-degree approximation. As a good practice, if the goal is to explore higher-degree BFs, the maximum model degree should be set greater than one, and at least one relatively high value of $\xi^{(q)}$ should be included to ensure that only sufficiently effective higher-degree terms are retained. For both `minspan` and `endspan`, small values between 1 and 10 are generally recommended to maximize flexibility during knot placement. Alternatively, values proposed by Friedman (1991) and Zhang (1994) can be used. As for the `grid`, unless computational issues arise, it is recommended to use the dataset's own values for the knots. Regarding, for the `LOF` criterion, it is suggested to keep the mean absolute error (MAE) or the MSE as the default, while for the penalty factor d, reasonable values to test are 0, 1, and 2.

All the aforementioned hyperparameters directly affect the estimation of the optimal output vector $\hat{\boldsymbol{y}}_i$, as they intervene during the model fitting process. In contrast, the threshold ψ, used in the refinement step defined in Equation (5.23), does not influence the estimation of $\hat{\boldsymbol{y}}_i$, but rather determines how the virtual dataset $\{(\boldsymbol{x}_i^*, \hat{\boldsymbol{y}}_i)\}_{i=1}^n$ is constructed by deciding when to replace a predicted output with its observed counterpart. Typical values include 0, 0.05, or 0.10, depending on how conservatively one wishes to correct potentially overestimated outputs. As a rule of thumb, $\psi = 0.05$ is a reasonable and robust choice across various scenarios. Importantly, evaluating different values of ψ does not incur additional computational cost, since it only modifies the post-processing of predictions and does not require refitting the underlying model.

Due to the large number of simulations carried out in this study, it was not feasible to perform an exhaustive search over the entire hyperparameter space. For this reason, some hyperparameters were fixed while others were varied. The maximum degree of variable interaction $q_{\max}$ was fixed at 2; the maximum number of BFs η was set to the number of DMUs in the analysis; the minimum error reduction rate ξ was set to 0.005; the knot penalty d was tested with values 1 and 2; the `minspan` followed Friedman's recommendation, while the `endspan` was not used; the knot grid was defined using the observed data points; and the LOF criterion was based on the MSE. Additionally, the threshold ψ used in the refinement step was set to 0.05. Finally, the hyperparameter $\xi^{(q)}$ was tested across a range of values specific to each simulation setting. In the Experimental Design 2 based on Perelman and Santín (2009), we used $\xi^{(2)} \in \{0.10, 0.20, 0.50\}$, while in the Experimental Design 3 based on Färe et al. (1994), we tested $\xi^{(2)} \in \{0, 0.05, 0.10\}$.

For optimal hyperparameter selection, techniques such as k-fold cross-validation (CV) are recommended. In our computational experiments, we adopt a 5-fold CV procedure, where the dataset is divided into five equal subsets. For each iteration, the model is trained on four folds and evaluated on the remaining one, ensuring that each fold serves as a test set exactly once. To better reflect the priorities of efficiency analysis, the predictive errors on the test fold are weighted using DEA-based output-oriented efficiency scores. Specifically, once the model is trained on the training folds and the refined technology is constructed, we estimate the output-oriented radial scores for the DMUs in the test fold by projecting them onto this estimated frontier. The inverse of each score is then used as a weight, assigning greater importance to highly efficient (out-of-sample) units. Differences between the predicted and observed outputs are computed and weighted accordingly, and the resulting fold-wise errors are averaged to compute the final CV score used for hyperparameter selection.

Beyond the hyperparameters described above, model performance is also influenced by the structural constraints imposed during the first stage of the method. These shape constraints—specifically, monotonicity and concavity—can be selectively applied depending on the application context. For instance, a fully flexible envelopment model is obtained by omitting both Constraints (5.18) and (5.19). If only non-decreasing monotonicity is desired, Constraint (5.18) should be enforced. Likewise, applying Constraint (5.19) enforces concavity. When both properties are required, both constraints must be simultaneously applied. This raises a natural question: since the final DEA-type technology, defined in Equation (5.24), inherently satisfies all shape axioms, is it necessary—or even beneficial—to enforce these constraints at the beginning? For instance, is it preferable to estimate a non-decreasing monotone model and defer the enforcement of concavity to the final DEA step—thus allowing greater flexibility in coefficient estimation—or is it more effective to impose both constraints from the outset? In principle, these shape constraints are treated as hyperparameter of the model, selected adaptively to best suit the data at hand. However, instead of adopting this strategy, we aim to offer practical guidance by analyzing the conditions under which each configuration of shape constraints performs best. To this end, we conduct a simulation study based on the Experimental Design 2 from Perelman and Santín (2009), evaluating the predictive performance of four constraint configurations across a range of scenarios that vary in both the number of DMUs (50, 100, and 150) and the level of noise present in the data.

The results are presented in Table 5.3 for noise-free scenarios and in Table 5.4 for scenarios with random noise. In both cases, 100 independent trials were performed for every combination of sample size (n) and proportion of units located on the true frontier (border). Four model configurations were considered: (i) both monotonicity and concavity imposed (approach 1); (ii) only monotonicity (approach 2); (iii) only concavity (approach 3); and (iv) no shape constraints (approach 4). In all settings, the envelopment condition was enforced to ensure that predicted outputs lie above the observed data.

The simulation results highlight clear guidelines for configuring the ACES estimator depending on the available sample size. For small sample sizes (approximately 50 units or fewer), imposing both monotonicity and concavity (Approach 1) typically yields the best results, as the limited data alone do not adequately capture the frontier's shape. With moderate sample sizes (around 100 units), imposing only one of the shape constraints—either monotonicity or concavity (Approach 2 or 3)—is preferable, offering a good trade-off between structure and flexibility. Finally, when the sample size reaches 150 units or more, it is most effective to impose only the envelopment condition (Approach 4). At this stage, the data itself sufficiently captures the true frontier, and additional shape restrictions tend to unnecessarily constrain model flexibility.

Table 5.3: Performance comparison of different approaches in the first stage of ACES in scenarios without random noise from Experimental Design 2.

border	n	**Approach 1**	**Approach 2**	**Approach 3**	**Approach 4**
0%	50	0.095	0.105	0.116	0.133
	100	0.066	0.058	0.061	0.067
	150	0.053	0.048	0.045	0.044
5%	50	0.073	0.085	0.098	0.104
	100	0.055	0.052	0.052	0.055
	150	0.057	0.045	0.046	0.041
10%	50	0.063	0.080	0.096	0.094
	100	0.055	0.049	0.046	0.046
	150	0.053	0.046	0.041	0.040
20%	50	0.063	0.069	0.064	0.064
	100	0.060	0.045	0.037	0.041
	150	0.061	0.053	0.049	0.037

While this section provides practical guidelines for setting the hyperparameters of the ACES model, in practice—and particularly when computational resources permit—it is strongly recommended to carry out a comprehensive and carefully designed CV procedure to determine the most suitable configuration. A grid search over plausible hyperparameter values is generally worthwhile, as the relationship between any given tuning parameter and the resulting quality of a non-parametric frontier can be highly non-linear and is rarely intuitive. This need for empirical tuning becomes even more pressing when there is no prior knowledge regarding the relative importance of the hyperparameters or their potential interaction effects—for example, how the complexity penalty d might interact with the maximum interaction degree $q_{\max}$.

Table 5.4: Performance comparison of different approaches in the first stage of ACES in scenarios with random noise from Experimental Design 2.

border	n	Approach 1	Approach 2	Approach 3	Approach 4
0%	50	0.115	0.128	0.131	0.163
	100	0.092	0.083	0.088	0.085
	150	0.100	0.079	0.081	0.067
5%	50	0.102	0.120	0.122	0.118
	100	0.095	0.074	0.074	0.081
	150	0.089	0.080	0.072	0.064
10%	50	0.101	0.101	0.101	0.133
	100	0.094	0.078	0.071	0.077
	150	0.089	0.080	0.074	0.067
20%	50	0.093	0.098	0.078	0.103
	100	0.092	0.085	0.069	0.064
	150	0.098	0.080	0.076	0.061

Importantly, enveloping models such as ACES exhibit a distinctive characteristic: the estimated frontier is shaped almost exclusively by a small subset of observations—those that sit very close to the boundary—whereas the many interior units exert little or no influence on its geometry. Consequently, even small changes in which boundary observations fall into a particular training fold can translate into noticeably different estimates of the technology. This "support-set sensitivity" can produce a surprisingly wide dispersion of validation errors from one data split to another, masking the true predictive capability of the model and potentially pushing the hyperparameter search toward configurations that perform well only under a lucky partition of the data.

To mitigate this variability and to obtain a validation statistic that is genuinely representative of out-of-sample performance, two complementary strategies are advisable:

- Repeated k-fold CV. Instead of relying on a single random partition, repeat the entire k-fold procedure dozens of times, each with a different random seed. Averaging the resulting validation errors reduces the variance induced by any one idiosyncratic split and yields a more stable risk estimate.

- Stratified CV targeted at frontier support. Whenever it is feasible to identify boundary candidates (e.g., by a quick preliminary DEA run or by selecting points with the highest predicted outputs in an initial ACES fit), enforce a stratification rule that distributes these high-leverage

units evenly across folds. This guarantees that each fold contains a balanced mix of frontier-defining and interior observations, thereby buffering the model against fold-to-fold swings driven by the accidental absence of key support points.

In practical terms, a 5×10 repeated (stratified) CV works well for medium-sized data sets: five folds provide a reasonable trade-off between training-set size and validation-set robustness, while ten repetitions typically shrink the standard error of the mean validation loss to an acceptably low value. For large-scale data sets, one can drop to three folds but increase the number of repetitions while for very small data sets, a leave-one-out approach with stratified resampling of "virtual" folds often gives the most reliable signal.

Finally, to perform CV, it is strongly recommended to use the MAE rather than the MSE as the LOF metric when validating frontier models. Since interior units—those far from the frontier—can produce disproportionately large residuals, MSE tends to overweight their contribution and may distort the selection process. In contrast, MAE treats all deviations linearly, yielding a more balanced validation signal that better reflects the geometry and purpose of the estimated frontier. In addition, it is advisable to weight validation errors according to each unit's proximity to the estimated frontier. Units closer to the boundary are more relevant for defining the shape and quality of the frontier estimate, and therefore their residuals should carry greater importance during model selection. One practical approach is to use DEA scores as weights. However, if the test set is small, DEA may label all units as efficient, providing no meaningful discrimination. In such cases, it is preferable to compute the efficiency scores using the model fitted on the training data and apply them as weights to the test residuals, ensuring consistent and informative weighting during CV.

Example 5.1. *We consider a dataset of 100 DMUs generated under the simulation framework proposed by Färe et al. (1994), and evaluate the sensitivity of the ACES methodology to its hyperparameters by applying a repeated 5-fold cross-validation procedure with five repetitions. In each repetition, the data is randomly split into five equally sized folds. For each configuration in the hyperparameter grid, the model is trained on four folds (80 DMUs) and validated on the remaining one (20 DMUs), rotating the validation fold across the five iterations. This entire process is repeated five times with different random partitions, ensuring a more robust estimation of the validation error.*

The minimum error reduction rate is fixed at $\xi = 0.01$*, while* `minspan` *and* `endspan` *are computed following Friedman's approach. The remaining hyperparameters form a grid that combines the following values:*

- *Interaction threshold* $\xi^{(q)} \in \{0, 0.05\}$.
- *Monotonicity constraint* `monotonicity` $\in \{$`TRUE`, `FALSE`$\}$.

- *Concavity constraint* $\texttt{concavity} \in \{\textit{TRUE}, \textit{FALSE}\}$.
- *GCV penalty parameter* $d \in \{1, 2\}$.
- *Regularization parameter for the second-stage envelopment* $\psi \in \{0.01, 0.02, 0.05, 0.10\}$.

This results in 64 different configurations. For each one, we compute the average validation error across folds. The total computation time using a Supermicro SYS-120GQ-TNRT node equipped with two Intel® Xeon® Silver 4316 CPUs operating at 2.30GHz was 4 hours and 05 minutes.

Ideally, the CV procedure should identify the hyperparameter configuration that most accurately estimates the true radial output scores. Since we operate in a simulated setting where the true efficiency scores are known (as in Example 5.1), we can directly assess the actual performance of each configuration. Table 5.5 reports the five best and five worst configurations based on the true mean squared error (MSE) between the estimated and actual scores. For each configuration, the table includes the tested hyperparameter values, the true MSE, the computational time (in seconds), the true ranking position based on actual performance (column `True Rank`), and the ranking position derived from CV (column `CV Rank`).

Table 5.5: Top five and bottom five hyperparameter configurations according to true performance.

True Rank	CV Rank	$\xi^{(q)}$	Monotonicity	Concavity	d	ψ	MSE	Time (s)
1	1	0.05	T	T	2	0.01	0.02	17.9
2	1	0.05	T	T	2	0.02	0.02	17.9
3	5	0.05	T	T	2	0.05	0.02	17.9
4	13	0.05	T	T	2	0.10	0.02	17.9
5	8	0.00	T	F	1	0.01	0.03	23.9
3	31	0.00	F	F	2	0.10	0.04	66.0
4	26	0.00	F	F	1	0.01	0.05	66.0
4	26	0.00	F	F	1	0.02	0.05	66.0
4	27	0.00	F	F	1	0.05	0.05	66.0
4	29	0.00	F	F	1	0.10	0.05	66.0

Lastly, it is worth emphasizing that hyperparameter choices strongly influence the computational cost of fitting the ACES model. In particular, configurations that do not impose monotonicity or concavity tend to produce slower convergence during the forward selection stage. Similarly, the value of the error reduction threshold ξ plays a key role: lower values of ξ lead to the inclusion of more BFs, thereby increasing the number of optimization problems that must be solved. As such, careful tuning of these parameters is not only important for model performance, but also critical for ensuring computational tractability in practice.

5.6.3 Results from Experimental Design 1

Table 5.6 reports the MSE obtained by each method under Experimental Design 1. The first two columns describe the scenarios in terms of input dimensionality and sample size. The next three columns present the results from the adapted MARS-based models introduced in Chapter 4, which are based on additive spline structures. In contrast, the last three columns correspond to the ACES-based methods proposed in this chapter, which estimate coefficients via Model (5.10) and incorporate input interaction terms. The results indicate that the ACES family consistently outperforms its MARS-based counterparts, particularly as dimensionality increases. In low-dimensional settings (Scenarios 1 and 2), all methods achieve comparably low error levels. However, as the complexity of the production function rises (Scenarios 3 to 6), the predictive gap becomes increasingly pronounced, highlighting the importance of capturing cross-variable effects in frontier estimation. While additive MARS variants remain competitive in simple environments, their accuracy deteriorates noticeably under stronger interactions, confirming the advantages of the more flexible ACES formulations.

Table 5.6: MSE for ACES methods in Experimental Design 1.

Scenario # inputs	Sample size	Linear MARS-based	Cubic MARS-based	Quintic MARS-based	**ACES**	**CS-ACES**	**QS-ACES**
1 (1)	50	0.007	0.005	0.007	0.006	0.006	0.005
	100	0.003	0.002	0.002	0.003	0.003	0.003
	150	0.002	0.001	0.002	0.002	0.002	0.001
2 (1)	50	0.007	0.006	0.009	0.007	0.006	0.011
	100	0.004	0.002	0.003	0.004	0.004	0.004
	150	0.002	0.001	0.002	0.002	0.002	0.002
3 (2)	50	0.018	0.024	0.022	0.014	0.012	0.011
	100	0.018	0.028	0.024	0.005	0.005	0.004
	150	0.018	0.026	0.024	0.004	0.003	0.003
4 (3)	50	0.018	0.021	0.020	0.021	0.025	0.028
	100	0.016	0.020	0.018	0.012	0.014	0.017
	150	0.015	0.022	0.018	0.008	0.012	0.018
5 (2)	50	0.008	0.006	0.007	0.008	0.010	0.009
	100	0.004	0.003	0.004	0.003	0.003	0.002
	150	0.003	0.003	0.004	0.002	0.002	0.001
6 (3)	50	0.012	0.010	0.011	0.012	0.012	0.015
	100	0.006	0.006	0.007	0.006	0.005	0.007
	150	0.004	0.005	0.006	0.003	0.003	0.004

5.6.4 Results from Experimental Designs 2 and 3

In this final section on computational experiments, we present a comparative analysis of the performance of ACES, DEA, StoNED, CEAT, and BDEA across the previously described scenarios. For the ACES model, and following the guidance provided in Tables 5.3 and 5.4, the results are reported using the following configurations: Approach 1 for sample sizes of 50 and 75; Approach 3 for sample sizes of 100 and 125; and Approach 4 for sample sizes of 150, 175, 200, 250, and 300. Accordingly, Tables 5.11 and 5.12 present the results corresponding to the Experimental Design 2 from Perelman and Santín (2009) while Tables 5.13 and 5.14 show those obtained under the framework of Experimental Design 3 from Färe et al. (1994).

The first two columns of Tables 5.11, 5.12, 5.13, and 5.14 indicate the proportion of DMUs located on the true frontier and the total number of DMUs evaluated in each scenario. The remaining columns compare the performance of the different methodologies in terms of estimation error (MSE), bias, and computation time. All percentage differences are expressed with respect to the competing methods. In this way, a negative value denotes that ACES outperforms the corresponding method (i.e., achieves lower error), while a positive value indicates a performance deficit. Finally, note that results for Experimental Design 3 and StoNED with a sample size of 25 are not reported, as they led to abnormally large values that distorted the true performance of the methods and would have undermined the comparability of the results.

We now detail the performance differences observed across the two experimental settings. We begin with Experimental Design 3 from Färe et al. (1994), where the results are particularly compelling. ACES outperforms all competing methods in every scenario tested—regardless of noise levels or the proportion of efficient units. Error reductions related to CEAT exceed 90% in nearly all cases, while improvements over DEA, BDEA, and StoNED range from 19.6% to 85.3%. Overall, no systematic performance differences are observed between noisy and noise-free settings; however, a clear trend emerges whereby the relative advantage of ACES increases with both sample size and the proportion of DMUs located on the true frontier. Tables 5.7 and 5.8 summarize these results.

Table 5.7: Aggregated results by number of DMUs on the true frontier from Experimental Design 3.

Border	Not noise				Noise			
	DEA	BDEA	CEAT	StoNED	DEA	BDEA	CEAT	StoNED
0%	-66.2%	-52.4%	-89.2%	-61.3%	-68.3%	-56.7%	-90.2%	-65.5%
5%	-72.2%	-61.9%	-92.6%	-66.2%	-72.6%	-62.6%	-93.2%	-70.2%
10%	-74.2%	-64.4%	-93.7%	-71.7%	-73.0%	-63.8%	-93.7%	-65.6%
20%	-75.6%	-67.4%	-95.5%	-73.0%	-75.1%	-66.8%	-95.2%	-71.1%

Table 5.8: Aggregated results by sample size from Experimental Design 3.

Sample size	Not noise				Noise			
	DEA	**BDEA**	**CEAT**	**StoNED**	**DEA**	**BDEA**	**CEAT**	**StoNED**
25	-49.0%	-32.3%	-78.0%	-67.0%	-46.2%	-28.8%	-77.6%	-67.0%
75	-72.2%	-61.1%	-94.5%	-76.3%	-74.3%	-64.9%	-95.0%	-72.5%
125	-79.8%	-71.5%	-96.3%	-71.6%	-81.5%	-74.6%	-97.2%	-73.6%
175	-77.4%	-68.1%	-96.9%	-60.5%	-77.9%	-69.6%	-97.3%	-63.8%
250	-81.9%	-74.5%	-98.2%	-63.7%	-81.3%	-74.5%	-98.4%	-62.4%

In the case of the Experimental Design 2 from Perelman and Santín (2009), ACES maintains strong performance relative to most benchmark methods. Notably, it consistently achieves error reductions exceeding 90% when compared to CEAT. ACES also outperforms StoNED across all configurations—regardless of noise levels, sample size, or the proportion of efficient units—with a larger margin of improvement observed in noise-free settings. In the presence of stochastic noise, however, the performance gap between StoNED and ACES narrows as both the sample size and the number of units on the true frontier increase, suggesting that StoNED benefits more from larger, noise-contaminated datasets where its stochastic structure becomes more effective. A similar pattern is observed when comparing ACES to standard DEA. In both noisy and noise-free settings, the performance of the two methods converges as the sample size and the proportion of efficient units increase. In particular, under noise-free conditions, DEA slightly outperforms ACES in scenarios with a sample size of 200 or more and when 20% of the units lie on the true frontier. This suggests that DEA may benefit from its nonparametric envelopment structure in large, well-populated datasets where the frontier is densely represented. Finally, BDEA consistently delivers superior results in this class of scenarios. Its advantage over ACES becomes more pronounced as the sample size increases and a larger proportion of DMUs lie on the true frontier. These findings reinforce the strengths of bootstrap-based bias correction in well-populated and frontier-dense datasets, where resampling techniques can more effectively capture the underlying efficiency structure. Tables 5.9 and 5.10 summarize these results.

Table 5.9: Aggregated results by number of DMUs on the true frontier from Experimental Design 2.

Border	Not noise				Noise			
	DEA	**BDEA**	**CEAT**	**StoNED**	**DEA**	**BDEA**	**CEAT**	**StoNED**
0%	-40.2%	+10.1%	-96.6%	-47.6%	-20.1%	+6.8%	-96.0%	-39.2%
5%	-35.2%	+14.4%	-97.5%	-50.5%	-9.6%	+14.0%	-96.3%	-32.6%
10%	-23.7%	+27.3%	-97.6%	-47.2%	-8.6%	+9.0%	-97.0%	-30.7%
20%	-13.9%	-24.5%	-98.2%	-40.1%	+1.1%	+14.5%	-97.2%	-22.6%

Table 5.10: Aggregated results by sample size from Experimental Design 2.

Sample size	Not noise				Noise			
	DEA	**BDEA**	**CEAT**	**StoNED**	**DEA**	**BDEA**	**CEAT**	**StoNED**
50	-43.3%	-9.3%	-95.0%	-43.8%	-35.8%	-6.9%	-93.6%	-39.2%
100	-48.2%	-14.4%	-97.5%	-47.0%	-33.4%	-6.2%	-96.6%	-39.9%
150	-28.4%	+20.5%	-97.8%	-45.0%	-14.3%	+9.2%	-97.1%	-31.5%
200	-16.1%	+42.2%	-97.8%	-44.0%	+2.4%	+19.4%	-97.6%	-26.2%
300	-5.3%	+56.5%	-98.9%	-51.9%	+34.7%	+39.9%	-98.1%	-19.5%

Table 5.11: Computational experiments in scenarios without random noise from Experimental Design 2.

Border	n	**MSE (ACES vs baseline model)**					**Bias**					**Computation time (in seconds)**				
		ACES	DEA	BDEA	CEAT	StoNED	ACES	DEA	BDEA	CEAT	StoNED	ACES	DEA	BDEA	CEAT	StoNED
0%	50	0.115	0.206 (-44.2%)	0.129 (-11.1%)	1.596 (-92.8%)	0.169 (-31.9%)	-0.194	-0.333	-0.200	0.560	-0.169	10	0	5	4	1
	100	0.071	0.137 (-48.2%)	0.078 (-9.7%)	1.962 (-96.4%)	0.132 (-46.3%)	-0.148	-0.261	-0.131	0.690	-0.213	34	0	13	11	1
	150	0.063	0.103 (-38.9%)	0.056 (+11.3%)	2.207 (-97.2%)	0.118 (-46.7%)	-0.082	-0.223	-0.102	0.778	-0.229	323	0	24	22	3
	200	0.051	0.083 (-38.2%)	0.043 (+18.3%)	2.320 (-97.8%)	0.105 (-51.3%)	-0.063	-0.197	-0.081	0.801	-0.210	989	0	35	38	16
	300	0.038	0.056 (-31.7%)	0.027 (+41.9%)	3.343 (-98.8%)	0.100 (-61.6%)	-0.021	-0.163	-0.056	1.071	-0.212	2920	0	93	95	21
5%	50	0.100	0.171 (-41.3%)	0.106 (-4.9%)	1.859 (-94.6%)	0.172 (-41.5%)	-0.142	-0.286	-0.146	0.647	-0.142	10	0	5	4	0
	100	0.056	0.109 (-48.6%)	0.064 (-11.8%)	2.512 (-97.8%)	0.100 (-43.9%)	-0.105	-0.214	-0.083	0.837	-0.184	40	0	12	10	1
	150	0.047	0.082 (-42.7%)	0.048 (-0.8%)	2.071 (-97.7%)	0.099 (-52.3%)	-0.041	-0.177	-0.051	0.739	-0.194	342	0	22	19	5
	200	0.043	0.056 (-23.6%)	0.030 (+45.0%)	2.557 (-98.3%)	0.092 (-53.5%)	-0.010	-0.150	-0.032	0.891	-0.189	909	0	34	31	12
	300	0.036	0.044 (-19.9%)	0.025 (+44.7%)	3.630 (-99.0%)	0.092 (-61.4%)	0.018	-0.122	-0.013	1.102	-0.201	3161	0	77	93	22
10%	50	0.082	0.149 (-44.7%)	0.092 (-10.5%)	1.846 (-95.5%)	0.144 (-42.7%)	-0.105	-0.260	-0.123	0.697	-0.070	15	0	8	5	0
	100	0.049	0.091 (-46.6%)	0.054 (-9.8%)	1.846 (-97.4%)	0.095 (-48.4%)	-0.075	-0.185	-0.052	0.747	-0.159	51	0	19	17	1
	150	0.044	0.061 (-28.6%)	0.035 (+23.8%)	2.161 (-97.9%)	0.085 (-48.6%)	-0.015	-0.146	-0.021	0.788	-0.165	390	0	25	23	3
	200	0.041	0.047 (-12.3%)	0.027 (+34.6%)	2.595 (-98.4%)	0.081 (-49.6%)	0.015	-0.123	-0.005	0.899	-0.162	967	0	35	33	8
	300	0.037	0.056 (-33.3%)	0.021 (+43.2%)	3.249 (-98.8%)	0.071 (-47.7%)	0.047	-0.091	0.015	1.051	-0.147	3041	0	75	71	29
20%	50	0.060	0.105 (-42.9%)	0.067 (-10.5%)	1.966 (-96.9%)	0.146 (-59.1%)	-0.052	-0.195	-0.057	0.757	-0.056	10	0	5	3	0
	100	0.035	0.070 (-49.5%)	0.048 (-26.2%)	2.288 (-98.5%)	0.070 (-49.5%)	-0.026	-0.134	-0.005	0.831	-0.100	37	0	12	10	1
	150	0.043	0.044 (-3.4%)	0.029 (+47.8%)	2.556 (-98.3%)	0.063 (-32.4%)	0.034	-0.103	0.014	0.905	-0.097	411	0	28	30	6
	200	0.042	0.038 (+9.5%)	0.027 (+52.7%)	2.626 (-98.4%)	0.054 (-22.7%)	0.053	-0.084	0.026	0.926	-0.098	883	0	34	37	6
	300	0.033	0.028 (+16.7%)	0.020 (+58.8%)	3.022 (-98.9%)	0.051 (-36.8%)	0.065	-0.065	0.033	1.022	-0.081	2522	0	77	73	33

Table 5.12: Computational experiments in scenarios with random noise from Experimental Design 2.

Border	n	**MSE (ACES vs baseline model)**					**Bias**					**Computation time (in seconds)**				
		ACES	DEA	BDEA	CEAT	StoNED	ACES	DEA	BDEA	CEAT	StoNED	ACES	DEA	BDEA	CEAT	StoNED
0%	50	0.143	0.233 (-38.8%)	0.158 (-9.4%)	1.785 (-92.0%)	0.234 (-39.0%)	-0.153	-0.319	-0.172	0.658	-0.137	15	0	8	5	0
	100	0.085	0.127 (-32.8%)	0.083 (+3.1%)	2.004 (-95.8%)	0.143 (-40.3%)	-0.094	-0.217	-0.077	0.729	-0.214	36	0	13	11	1
	150	0.087	0.116 (-24.3%)	0.083 (+5.8%)	2.510 (-96.5%)	0.141 (-38.0%)	-0.043	-0.174	-0.036	0.822	-0.222	321	0	24	22	3
	200	0.077	0.098 (-21.5%)	0.075 (+2.0%)	3.058 (-97.5%)	0.131 (-41.6%)	0.003	-0.138	-0.002	1.001	-0.212	1036	0	30	49	10
	300	0.081	0.069 (+17.0%)	0.061 (+32.3%)	4.159 (-98.1%)	0.128 (-37.1%)	0.066	-0.089	0.038	1.216	-0.217	3659	0	55	111	14
5%	50	0.130	0.188 (-30.7%)	0.125 (+4.4%)	1.748 (-92.6%)	0.198 (-34.2%)	-0.115	-0.277	-0.132	0.651	-0.137	14	0	4	5	0
	100	0.079	0.126 (-37.0%)	0.087 (-9.4%)	2.111 (-96.2%)	0.132 (-40.0%)	-0.082	-0.193	-0.052	0.741	-0.202	50	0	19	17	1
	150	0.074	0.085 (-12.7%)	0.065 (+14.4%)	2.582 (-97.1%)	0.115 (-35.4%)	0.000	-0.133	0.005	0.862	-0.186	434	0	34	34	3
	200	0.078	0.081 (-3.20%)	0.067 (+17.9%)	3.248 (-97.6%)	0.109 (-28.3%)	0.039	-0.108	0.024	1.027	-0.179	1203	0	56	59	9
	300	0.085	0.063 (+35.5%)	0.059 (+43.4%)	4.111 (-97.9%)	0.113 (-25.0%)	0.084	-0.069	0.053	1.203	-0.197	3941	0	102	109	15
10%	50	0.113	0.185 (-39.2%)	0.129 (-12.7%)	2.075 (-94.6%)	0.172 (-34.3%)	-0.095	-0.250	-0.102	0.711	-0.125	11	0	5	3	0
	100	0.069	0.102 (-32.7%)	0.074 (-7.0%)	2.317 (-97.0%)	0.125 (-45.0%)	-0.043	-0.161	-0.016	0.803	-0.150	38	0	12	10	2
	150	0.072	0.085 (-15.3%)	0.069 (+5.0%)	2.845 (-97.5%)	0.101 (-28.7%)	0.020	-0.116	0.019	0.948	-0.156	338	0	23	20	4
	200	0.073	0.071 (+2.6%)	0.062 (+17.9%)	3.504 (-97.9%)	0.103 (-29.2%)	0.050	-0.089	0.038	1.016	-0.170	994	0	36	33	8
	300	0.080	0.057 (+41.6%)	0.057 (+41.8%)	4.292 (-98.1%)	0.096 (-16.0%)	0.091	-0.052	0.065	1.250	-0.164	3508	0	71	68	14
20%	50	0.090	0.138 (-34.4%)	0.100 (-9.9%)	1.994 (-95.5%)	0.178 (-49.4%)	-0.036	-0.186	-0.045	0.766	-0.068	11	0	4	3	0
	100	0.055	0.079 (-31.2%)	0.062 (-11.5%)	2.204 (-97.5%)	0.083 (-34.2%)	-0.013	-0.121	0.012	0.846	-0.092	37	0	11	9	2
	150	0.063	0.066 (-5.1%)	0.056 (+11.6%)	2.403 (-97.4%)	0.082 (-23.9%)	0.048	-0.085	0.037	0.890	-0.119	317	0	22	19	5
	200	0.071	0.054 (+31.8%)	0.050 (+40.5%)	2.857 (-97.5%)	0.075 (-5.5%)	0.077	-0.061	0.055	0.987	-0.110	925	0	35	32	9
	300	0.067	0.047 (+44.6%)	0.047 (+42.1%)	3.726 (-98.2%)	0.067 (+0.1%)	0.100	-0.036	0.066	1.174	-0.083	2730	0	75	78	23

Table 5.13: Computational experiments in scenarios without random noise from Experimental Design 3.

Border	n	MSE (ACES vs baseline model)					Bias					Computation time (in seconds)				
		ACES	DEA	BDEA	CEAT	StoNED	ACES	DEA	BDEA	CEAT	StoNED	ACES	DEA	BDEA	CEAT	StoNED
0%	25	0.229	0.388 (-40.2%)	0.285 (-19.6%)	0.734 (-68.7%)		-0.336	-0.449	-0.324	0.298		6	0	2	1	
	75	0.096	0.278 (-65.5%)	0.194 (-50.7%)	1.094 (-91.2%)	0,379 (-74.7%)	-0.211	-0.353	-0.230	0.543	-0.066	36	0	7	9	2
	125	0.068	0.269 (-74.7%)	0.187 (-63.7%)	1.116 (-93.9%)	0.209 (-67.5%)	-0.175	-0.329	-0.207	0.567	-0.184	73	0	17	15	11
	175	0.070	0.255 (-72.7%)	0.178 (-60.8%)	1.412 (-95.1%)	0.136 (-48.9%)	-0.170	-0.302	-0.185	0.636	-0.184	867	0	32	30	12
	250	0.055	0.240 (-77.3%)	0.169 (-67.4%)	1.932 (-97.2%)	0.119 (-54.0%)	-0.142	-0.277	-0.165	0.825	-0.142	2542	0	58	56	16
5%	25	0.184	0.282 (-50.4%)	0.204 (-35.4%)	0.853 (-78.5%)		-0.272	-0.404	-0.283	0.377		9	0	4	2	0
	75	0.073	0.282 (-74.0%)	0.204 (-64.2%)	1.249 (-94.1%)	0.237 (-69.1%)	-0.165	-0.320	-0.192	0.595	-0.071	38	0	13	11	3
	125	0.049	0.228 (-78.7%)	0.161 (-69.7%)	1.174 (-95.8%)	0.158 (-69.3%)	-0.122	-0.269	-0.147	0.595	-0.139	98	0	19	25	22
	175	0.055	0.231 (-76.4%)	0.164 (-66.7%)	1.591 (-96.6%)	0.138 (-60.5%)	-0.131	-0.260	-0.143	0.747	-0.151	599	0	26	22	9
	250	0.041	0.221 (-81.4%)	0.156 (-73.5%)	2.121 (-98.1%)	0.121 (-66.0%)	-0.099	-0.238	-0.122	0.880	-0.122	1553	0	42	35	19
10%	25	0.147	0.324 (-54.6%)	0.239 (-29.2%)	0.721 (-79.6%)		-0.232	-0.366	-0.234	0.341	-0.006	6	0	2	1	0
	75	0.067	0.252 (-73.5%)	0.180 (-62.8%)	1.519 (-95.6%)	0.409 (-83.6%)	-0.136	-0.294	-0.164	0.693	0.006	27	0	7	6	3
	125	0.040	0.216 (-81.3%)	0.154 (-76.6%)	1.449 (-97.2%)	0.139 (-71.0%)	-0.092	-0.249	-0.125	0.691	-0.115	73	0	17	15	7
	175	0.047	0.214 (-69.0%)	0.151 (-69.0%)	1.978 (-97.6%)	0.136 (-65.7%)	-0.095	-0.231	-0.111	0.867	-0.115	587	0	34	33	7
	250	0.033	0.200 (-77.0%)	0.143 (-83.6%)	2.294 (-98.6%)	0.098 (-66.4%)	-0.067	-0.209	-0.095	0.953	-0.115	2280	0	52	51	27
20%	25	0.144	0.289 (-50.2%)	0.223 (-35.2%)	0.973 (-85.2%)		-0.198	-0.324	-0.200	0.478		6	0	2	1	
	75	0.048	0.199 (-75.7%)	0.146 (-66.8%)	1.559 (-96.9%)	0.218 (-77.8%)	-0.085	-0.231	-0.107	0.722	-0.046	29	0	8	6	3
	125	0.029	0.184 (-84.4%)	0.131 (-78.7%)	1.634 (-98.2%)	0.135 (-78.7%)	-0.048	-0.196	-0.073	0.785	-0.075	61	0	15	12	9
	175	0.031	0.178 (-82.3%)	0.131 (-76.0%)	1.696 (-98.1%)	0.095 (-67.5%)	-0.048	-0.183	-0.069	0.785	-0.103	522	0	22	18	9
	250	0.025	0.173 (-85.3%)	0.127 (-80.0%)	2.305 (-98.9%)	0.081 (-68.6%)	-0.031	-0.168	-0.059	0.953	-0.108	1486	0	43	35	18

Table 5.14: Computational experiments in scenarios with random noise from Experimental Design 3.

Border	n	**MSE (ACES vs baseline model)**					**Bias**					**Computation time (in seconds)**				
		ACES	DEA	BDEA	CEAT	StoNED	ACES	DEA	BDEA	CEAT	StoNED	ACES	DEA	BDEA	CEAT	StoNED
0%	25	0.243	0.406 (-40.2%)	0.308 (21.2%)	0.749 (-67.6%)		-0.327	-0.453	-0.330	0.321		9	0	2	2	
	75	0.094	0.310 (-69.7%)	0.227 (-58.5%)	1.407 (-93.3%)	0.242 (-61.2%)	-0.180	-0.337	-0.205	0.626	-0.164	40	0	13	11	1
	125	0.053	0.259 (-79.4%)	0.184 (-71.0%)	1.426 (-96.3%)	0.168 (-68.3%)	-0.114	-0.282	-0.148	0.675	-0.159	71	0	19	18	5
	175	0.062	0.236 (-73.6%)	0.167 (-62.7%)	1.097 (-96.1%)	0.233 (-73.2%)	-0.118	-0.251	-0.122	0.758	-0.130	834	0	26	39	5
	250	0.050	0.236 (-78.6%)	0.169 (-70.2%)	2.284 (-97.8%)	0.124 (-59.4%)	-0.088	-0.227	-0.102	0.966	-0.145	2360	0	75	78	19
5%	25	0.184	0.251 (-50.5%)	0.173 (-36.2%)	0.293 (78.3%)		-0.247	-0.386	-0.264	0.366		8	0	2	2	
	75	0.068	0.251 (-72.9%)	0.173 (-60.8%)	1.316 (-94.8%)	0.303 (-77.6%)	-0.131	-0.292	-0.159	0.653	-0.084	38	0	7	10	2
	125	0.047	0.242 (-80.8%)	0.161 (-73.2%)	1.472 (-96.8%)	0.245 (-81.0%)	-0.088	-0.255	-0.121	0.701	-0.107	97	0	25	26	3
	175	0.049	0.224 (-78.3%)	0.161 (-69.8%)	2.077 (-97.6%)	0.123 (-60.4%)	-0.086	-0.224	-0.098	0.878	-0.132	961	0	44	48	6
	250	0.042	0.214 (-80.4%)	0.155 (-72.9%)	2.671 (-98.4%)	0.110 (-61.8%)	-0.056	-0.199	-0.075	1.071	-0.111	2611	0	81	85	23
10%	25	0.168	0.308 (-45.4%)	0.230 (-23.0%)	0.872 (-80.9%)		-0.224	-0.351	-0.220	0.402		9	0	4	2	
	75	0.066	0.272 (-75.7%)	0.205 (-67.7%)	1.287 (-94.9%)	0.205 (-67.7%)	-0.097	-0.269	-0.134	0.641	-0.085	30	0	8	6	1
	125	0.039	0.227 (-82.7%)	0.168 (-76.5%)	1.508 (-97.4%)	0.136 (-71.1%)	-0.057	-0.226	-0.097	0.728	-0.129	76	0	18	15	3
	175	0.040	0.222 (-82.0%)	0.165 (-77.6%)	1.653 (-97.6%)	0.116 (-60.3%)	-0.068	-0.205	-0.078	0.793	-0.101	999	0	45	45	7
	250	0.039	0.216 (-82.0%)	0.161 (-75.8%)	2.514 (-98.4%)	0.106 (-63.2%)	-0.039	-0.187	-0.068	1.050	-0.099	2934	0	84	87	25
20%	25	0.126	0.245 (-48.7%)	0.182 (-30.8%)	0.766 (-83.6%)		-0.178	-0.299	-0.168	0.407		6	0	2	1	
	75	0.045	0.215 (-79.0%)	0.164 (-72.4%)	1.550 (-97.1%)	0.276 (-83.6%)	-0.058	-0.217	-0.088	0.729	-0.010	29	0	9	7	2
	125	0.031	0.185 (-83.3%)	0.139 (-77.7%)	1.883 (-98.4%)	0.120 (-74.2%)	-0.027	-0.185	-0.058	0.822	-0.084	74	0	17	15	3
	175	0.033	0.172 (-80.6%)	0.129 (-74.0%)	1.704 (-98.0%)	0.086 (-61.3%)	-0.031	-0.164	-0.046	0.818	-0.087	853	0	32	33	7
	250	0.028	0.177 (-84.2%)	0.134 (-79.0%)	2.260 (-98.8%)	0.080 (-65.2%)	-0.006	-0.150	-0.037	0.964	-0.091	2334	0	57	56	17

5.7 An empirical illustration

In this section, we apply a real dataset to demonstrate the performance of various technical efficiency measures using ACES. The dataset includes information on 31 Taiwanese banks for the year 2010, previously analyzed by Juo et al. (2015). The inputs considered are financial `FUNDS` (x_1), `LABOR` (x_2), and physical `CAPITAL` (x_3), while the outputs are financial `INVESTMENTS` (y_1) and `LOANS` (y_2). All monetary variables are measured in million TWD, with labor measured as the number of employees. To improve the numerical stability of the algorithm, all monetary variables were rescaled by dividing them by 1,000. A detailed discussion of the statistical sources and variable definitions is available in Juo et al. (2015).

Regarding the ACES configuration, we tuned three key hyperparameters: the error reduction threshold $\xi \in \{0.005, 0.01\}$; the required improvement of a 3-degree and 2-degree BF over the best 1-degree candidate, $\xi^{(2)} = \xi^{(3)} \in \{0.0, 0.05, 0.10\}$; and the penalty per knot $d \in \{1, 2\}$. Monotonicity and concavity are both imposed in the initial stage due to the sample size being lower than 50 DMUs. The two best results were obtained by the following configurations: $\xi = 0.005$, $\xi^{(2)} = \xi^{(3)} = 0.10$, and $d = 1$ (`ACES 1`); and (ii) $\xi = 0.010$, $\xi^{(2)} = \xi^{(3)} = 0.05$, and $d = 1$ (`ACES 2`).

Table 5.15 reports the results for the various efficiency measures considered in the empirical analysis. The first column identifies each assessed bank. The subsequent three blocks, each comprising three columns, correspond to the efficiency models employed: the output-oriented radial model defined by Equation (5.25), the input-oriented radial model introduced by Equation (5.26), and the Directional Distance Function (DDF) specified in Equation (5.27). For the DDF, the directional vector $(\boldsymbol{g}_x, \boldsymbol{g}_y) = (-x_{01}, -x_{02}, -x_{03}, y_{01}, y_{02})$ is used to evaluate each DMU. Each column within these blocks reports the results obtained using DEA, `ACES 1`, and `ACES 2`, respectively.

Table 5.15 demonstrates that DEA consistently produces more optimistic efficiency assessments across all evaluated cases. This outcome is primarily due to DEA's omission of the minimum extrapolation principle, which typically positions the production frontier as close as possible to the observed data points. Moreover, the results reveal that different configurations of ACES (model 1 and model 2) lead to notably different efficiency evaluations. For example, under the radial output approach, the mean score in ACES model 1 is 1.29, in ACES model 2 is 1.31, while in DEA it is 1.24. Consequently, the density of ACES scores tends to decrease around 1, indicating fewer units at the efficiency threshold, while it increases throughout the rest of the distribution, reflecting a more realistic estimation of efficiency.

Table 5.15: Efficiency measures obtained from the empirical example.

Bank	Output-oriented			Input-oriented			DDF		
	ACES 1	ACES 2	DEA	ACES 1	ACES 2	DEA	ACES 1	ACES 2	DEA
Bank SinoPac	1.14	1.17	1.11	0.86	0.83	0.89	0.07	0.09	0.05
Bank of Kaohsiung	1.38	1.40	1.36	0.63	0.62	0.65	0.19	0.19	0.18
Bank of Panshin	1.76	1.75	1.75	0.37	0.37	0.37	0.34	0.34	0.34
Bank of Taiwan	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00
Cathay United Bank	1.40	1.41	1.38	0.69	0.69	0.71	0.17	0.18	0.17
Chang Hwa Bank	1.07	1.11	1.00	0.93	0.90	0.94	0.04	0.05	0.03
China Development	1.00	1.18	1.00	1.00	0.56	1.00	0.00	0.13	0.00
China Trust Bank	0.98	1.00	1.00	1.00	1.00	1.00	0.01	0.00	0.00
Cooperative Bank	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00
Cosmos Bank	2.26	2.25	2.24	0.23	0.23	0.23	0.52	0.52	0.52
Cota Bank	1.75	1.61	1.49	0.48	0.48	0.49	0.33	0.34	0.28
E. Sun Bank	1.12	1.12	1.12	0.84	0.84	0.84	0.08	0.08	0.08
Entie Bank	1.45	1.55	1.19	0.66	0.61	0.81	0.20	0.24	0.10
Export-Import Bank	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00
Far Eastern Bank	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00
First Bank	1.00	1.00	1.00	0.94	0.91	0.94	0.03	0.04	0.03
Hua Nan Bank	1.06	1.08	1.05	0.34	0.33	0.35	0.36	0.36	0.35
Hwatai Bank	1.74	1.73	1.03	0.72	0.72	1.00	0.26	0.26	0.00
Industrial Bank of Taiwan	1.58	1.58	1.00	0.43	0.43	0.48	0.32	0.32	0.28
Jih Sun Bank	1.73	1.72	1.63	0.43	0.43	0.48	0.32	0.32	0.28
Land Bank	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00
Mega Bank	1.01	1.03	1.00	0.98	0.96	1.00	0.01	0.02	0.00
Shin Kong Bank	1.36	1.35	1.35	0.70	0.70	0.70	0.16	0.16	0.16
Sunny Bank	1.45	1.44	1.44	0.59	0.60	0.66	0.22	0.21	0.21
Ta Chong Bank	1.25	1.25	1.14	0.74	0.72	0.86	0.13	0.15	0.07
Taichung Bank	1.33	1.36	1.29	0.70	0.68	0.73	0.16	0.17	0.14
Taipei Fubon Bank	1.00	1.04	1.00	1.00	0.93	1.00	0.00	0.03	0.00
Taishin Bank	1.21	1.25	1.19	0.80	0.76	0.82	0.12	0.12	0.09
Taiwan Business Bank	1.05	1.05	1.00	0.95	0.95	0.97	0.03	0.03	0.02
The Shanghai Bank	1.15	1.18	1.11	0.84	0.80	0.89	0.07	0.09	0.07
Union Bank	1.71	1.70	1.69	0.50	0.50	0.50	0.30	0.29	0.29
Mean	1.29	1.31	1.24	0.77	0.75	0.80	0.13	0.14	0.11
Median	1.15	1.18	1.11	0.83	0.76	0.89	0.08	0.12	0.06

5.8 Discussion and final remarks

ACES introduces several methodological advancements that markedly enhance both its flexibility and practical applicability compared to the earlier MARS-based approach. First, it relaxes the enforcement of monotonicity and concavity by applying these shape constraints to the global estimator rather than to each individual BF. This strategy enables more flexible parameter estimation while retaining rigorous compliance with fundamental production axioms. Second, ACES incorporates interaction terms among inputs, capturing complex nonlinearities and joint effects between variables that exceed the representational capability of purely additive models. Third, the methodology extends seamlessly to multi-output contexts through a novel DEA-type construction: an initial shape-constrained ML model is fitted to the data, selectively refined by replacing poorly predicted outputs with actual observations, and subsequently employed to define a VRS technology.

This innovative framework not only facilitates conventional efficiency measurement but also creates a natural bridge to the extensive DEA-based literature on productivity analysis. By preserving a VRS structure in the final stage, the ACES methodology remains fully compatible with standard distance function formulations widely used in productivity measurement. In particular, the piecewise linear technologies estimated via ACES can be employed to compute dynamic indices such as the Malmquist Productivity Index and the Luenberger Productivity Indicator. This connection opens the door to a powerful synthesis of advanced ML techniques with classical productivity analysis. The ability to incorporate shape constraints, interactions among inputs, and multi-output structures within a unified estimation framework offers a theoretically grounded and practically relevant alternative to traditional DEA approaches. Moreover, because the final ACES frontier satisfies the axioms of convexity and free disposability, and can be reconstructed for each time period or unit under study, it serves as a valid and flexible reference technology for multi-period comparisons. Consequently, the ACES framework is not only a methodological contribution to frontier estimation, but also a foundation upon which dynamic productivity measures can be consistently built. This dual capacity—combining predictive learning with economic interpretability—positions ACES as a valuable tool for both applied research and policy-oriented analysis in contexts where performance tracking over time is essential.

As a concluding remark, it is important to acknowledge two key limitations of the ACES methodology, both of which are closely interrelated: computational cost and hyperparameter tuning. Although ACES demonstrates robust empirical performance, its computational burden grows rapidly with the number of DMUs. As evidenced in Tables 5.11–5.14, the runtime increases at an exponential rate as the sample size expands. This effect becomes critical when the number of DMUs exceeds 300, even under moderate dimensionality (e.g., four variables and second-degree interactions for inputs). The main computational bottleneck arises from the number of times Model (5.15) must be solved during

the forward selection stage, which is directly driven by the size of the candidate BF set defined in Equation (3.33). Consequently, a promising direction for improvement would be to design more efficient strategies to intelligently reduce the size of the candidate set—focusing only on potentially viable BFs—thereby improving scalability without compromising estimation quality. In addition, the high computational cost hinders thorough exploration of the hyperparameter space. Parameters such as the error reduction threshold or the maximum degree of interaction play a critical role in balancing model flexibility and overfitting risk. However, the ability to systematically evaluate different configurations through CV is severely limited by runtime constraints, particularly in large-scale settings. As such, the development of faster heuristics or adaptive tuning strategies would be essential to unlock the full potential of ACES in practice.

Beyond these computational considerations, another important limitation of ACES lies in its sensitivity to local variations in the data. As a model based on adaptive selection of BFs, ACES can produce different frontier shapes depending on which units act as local supports, particularly when the sample includes noise or outliers, or when multiple BFs offer similar improvements in training error. This sensitivity may lead to instability in the estimated efficiency scores. To address this issue, the next chapter introduces RF-ACES, an ensemble-based extension of the methodology designed to enhance robustness. By aggregating the results of multiple ACES models trained on different subsamples, RF-ACES improves the stability of the estimator without compromising the shape constraints or interpretability of the original model, making it better suited for complex empirical environments.

Chapter 6

Random Forest-Adaptive Constrained Enveloping Splines

In previous chapters, we introduced Adaptive Constrained Enveloping Splines (ACES) as a robust method for estimating production frontiers that comply with economic theory constraints such as monotonicity and concavity. While ACES significantly mitigates the overfitting issues characteristic of traditional Data Envelopment Analysis (DEA) models, estimating frontiers through a single adaptive spline-based regression may still exhibit high variability—particularly in noisy environments or when dealing with complex nonlinear relationships

To address these limitations, this chapter builds upon the work of España et al. (2025a) and introduces a novel and flexible methodology for the nonparametric estimation of production technologies by integrating advanced machine learning (ML) techniques—specifically, Random Forests (RF) and shape-constrained splines—into the ACES framework. The resulting approach yields three principal contributions that strengthen the robustness of the estimator. First, the use of bootstrap samples—central to the Bagging (Bootstrap Aggregation) technique introduced by LeBlanc and Tibshirani (1996)—connects predictors (inputs) to outputs through multiple estimators fitted on resampled datasets. In our case, each estimator employs shape-constrained splines instead of decision trees, following the approach detailed in Chapter 5. Moreover, RF extends Bagging by also resampling the set of predictors at each step of the process, thereby introducing a second layer of randomness. This double randomization effectively decorrelates individual models and, when aggregated, substantially reduces variance and improves generalization—particularly in high-dimensional or noisy data settings.

Second, unlike DEA, the RF-based model can provide reliable predictions even for input configurations not observed in the sample, a feature particularly advantageous in policy-oriented applications where extrapolation and counterfactual analysis are required. Third, the ensemble structure enables the computation of input importance rankings, offering a systematic and interpretable way to identify the most influential drivers of output variation. Although statistical significance is not formally defined in ML, variable importance measures provide an analogous mechanism to assess relevance. This is particularly valuable in applied domains—such as education—where identifying critical factors like teacher–student ratios or infrastructure quality can inform evidence-based policymaking and guide resource allocation toward high-impact interventions.

Several methodologies in the efficiency analysis literature share conceptual similarities with the principles underpinning RF techniques, particularly in their use of resampling, model aggregation, and strategies to mitigate overfitting. Bootstrapping, for instance, has become one of the most widely used tools. Introduced into the DEA context by Simar and Wilson (1998, 2000), it enables the construction of confidence intervals for efficiency scores, allowing analysts to quantify the uncertainty associated with nonparametric frontier estimates. In parallel, robust partial frontier methods—pioneered by Cazals et al. (2002) and extended by Daouia and Gijbels (2011)—estimate the production frontier based on a restricted subset of the data, improving robustness against noise and outliers. While these techniques are primarily focused on inference, more recent developments have aimed at combining predictive performance with theoretical coherence. Notably, Esteve et al. (2023) proposed an ensemble methodology based on the Efficiency Analysis Trees (EAT) framework (Esteve et al., 2020), generating forest-based estimators designed to reduce variance through model randomization. Although this approach improves generalization compared to individual trees, it does not incorporate convexity assumptions and thus remains closer in spirit to Free Disposal Hull (FDH) than to fully convex technologies traditionally assumed in microeconomic production theory.

The proposed RF-ACES approach produces a piecewise linear estimator that remains analogous to DEA in structure and interpretability. Importantly, it preserves fundamental economic assumptions such as free disposability and convexity, ensuring consistency with microeconomic production theory. Building on this, the framework also enables the definition of a new radial output-oriented efficiency measure, offering a complementary perspective to the one presented in Chapter 5. To illustrate its empirical applicability, the methodology can be deployed in diverse contexts—ranging from the evaluation of regional educational systems using data from the Programme for International Student Assessment (PISA) to the assessment of firm-level efficiency in the Spanish food industry. In both settings, RF-ACES proves capable not only of delineating production frontiers but also of identifying the most influential inputs driving output performance. The resulting efficiency scores and variable importance rankings offer valuable insights for evidence-based policymaking and strategic resource allocation.

6.1 A new proposal to yield polyhedral technologies

The convexity postulate in microeconomics is a fundamental principle for understanding and measuring efficiency, particularly when using DEA. This postulate asserts that, given two producible input-output bundles, any convex combination of them must also be feasible. In the DEA framework, convexity is essential for constructing the production possibility set and generating reference targets for decision-making units (DMUs). Consequently, convexity serves as a cornerstone for DEA models, enabling the assessment of relative efficiency and supporting informed improvements in resource allocation and operational strategies (Banker et al., 1984). When seeking to estimate production technologies using robust ML models such as RF, recent proposals—such as the one introduced by Esteve et al. (2023)—have gained traction. However, these approaches rely on ensembles of tree-based models, which result in step-shaped frontiers and do not incorporate convexity constraints. As a result, their performance is typically compared with FDH methods, and not with DEA, despite convexity being a more common and desirable assumption in efficiency analysis.

Accordingly, this chapter seeks to fill this gap in the literature by proposing a RF–based estimator of production technologies that satisfies the convexity postulate. The key idea is to use the ACES method—introduced in Chapter 5—as the base learner within the ensemble. Each individual model is fitted on a bootstrap sample of the data. During the forward selection process, a random subset of the input variables is selected, following the standard RF paradigm. In contrast to the original ACES procedure, the backward elimination phase is omitted, since the combined randomization over both data and inputs provides sufficient regularization to control overfitting and maintain model stability. As in standard RF, RF-ACES introduces stochasticity into the model-building process by generating multiple diverse estimators based on resampled observations and input subsets. These individual models are then aggregated to produce a final prediction. As demonstrated by Louppe and Geurts (2012) and Louppe (2014), when the number of base learners increases, the variance of the ensemble prediction converges to the product of the correlation between any two individual learners and the variance of each individual learner. Assuming independence in the random components, both the noise and bias terms in the generalization error decomposition remain unchanged relative to the base estimator. Thus, the strength of RF-ACES lies in reducing the correlation among models through double randomization, which leads to lower ensemble variance and improved predictive performance.

Initially, the ensemble size is defined by the parameter $L_{\max}$, which specifies the total number of base learners to be trained. Subsequently, $L_{\max}$ bootstrap samples are drawn from the original dataset using sampling with replacement. Each resampled dataset is denoted by $\mathcal{L}_l^*$, for $l = 1, \ldots, L_{\max}$. The methodology introduced in Chapter 5 is then applied independently to each sample, with two key modifications. First, instead of evaluating all possible combinations in the set defined by Equa-

tion (3.33) to introduce a new reflected pair during the forward selection phase, a randomly selected subset of input variables is considered at each iteration. This selection typically follows Breiman's rule of thumb, which recommends choosing one-third of the available predictors in regression problems. If variable interactions are included and the feature space is expanded accordingly, the same proportion is preserved relative to the original set of predictors to avoid overfitting due to high dimensionality. Second, the backward elimination phase is omitted. Accordingly, the forward selection stops when either the number of selected terms reaches η, or the reduction in training error from adding a new reflected pair falls below ξ.

From this point onward, we focus on the specific case of estimating multiple outputs simultaneously, without incorporating variable interactions, while imposing monotonicity and concavity constraints on each base learner. Additionally, no refinement phase, such as the one described in Equation (5.23), is performed. In this setting, after executing RF-ACES, we obtain a collection of $L_{\max}$ fitted (non-decreasing and concave) production functions for each output r, each constructed by applying the ACES algorithm to a distinct bootstrap sample. These models form an ensemble of spline-based estimators. For any given input vector $\boldsymbol{x} = (x_1, \ldots, x_m) \in \mathbb{R}^m_+$—which may correspond to an out-of-sample configuration—the output for dimension r predicted by the l-th model is denoted by

$$\hat{f}^{\text{ACES}}_{lr}\left(\boldsymbol{x}; \tau_0\left(\mathcal{L}^*_l\right), \boldsymbol{\alpha}\left(\mathcal{L}^*_l\right), \boldsymbol{\beta}\left(\mathcal{L}^*_l\right)\right), \quad r = 1, \ldots, s, \quad l = 1, \ldots, L_{\max}. \tag{6.1}$$

Each of these base learners corresponds to the ACES estimator defined in Equation (5.15), applied to a different resampled dataset. For notational simplicity, we write $\hat{f}^{\text{ACES}}_{lr}(\boldsymbol{x})$ in the remainder of the Section, with the dependence on the bootstrap sample $\mathcal{L}^*_l$ understood implicitly.

Once all $L_{\max}$ models have been trained, the RF-ACES prediction for a given input profile is obtained by averaging the outputs of the individual models component-wise. The resulting vector-valued estimator can be formally expressed as:

$$\hat{\boldsymbol{f}}^{\text{RF-ACES}}(\boldsymbol{x}) = \frac{1}{L_{\max}} \sum_{l=1}^{L_{\max}} \hat{\boldsymbol{f}}^{\text{ACES}}_{l}(\boldsymbol{x}) = \frac{1}{L_{\max}} \sum_{l=1}^{L_{\max}} \left[\hat{f}^{\text{ACES}}_{l1}(\boldsymbol{x}), \ \ldots, \ \hat{f}^{\text{ACES}}_{ls}(\boldsymbol{x})\right]. \tag{6.2}$$

From the last mathematical expression, we define a possible RF-ACES technology estimator as:

$$\hat{\varphi}_{\text{RF-ACES}} = \left\{(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}^{m+s}_+ \ : \ y_r \leq \frac{1}{L_{\max}} \sum_{l=1}^{L_{\max}} \hat{f}^{\text{ACES}}_{lr}(\boldsymbol{x}), \quad r = 1, \ldots, s\right\}. \tag{6.3}$$

This set will serve as the reference technology for efficiency measurement. The monotonicity and concavity imposed on each component $\hat{f}^{\text{ACES}}_{lr}(\boldsymbol{x})$ ensure that $\hat{\varphi}_{\text{RF-ACES}}$ satisfies free disposability and defines a convex production set.

Proposition 6.1. [Free disposability] Let $(\boldsymbol{x}', \boldsymbol{y}') \in \hat{\varphi}_{\text{RF-ACES}}$. Then, for any $\boldsymbol{x}'' \geq \boldsymbol{x}'$ and $\mathbf{0}_s \leq \boldsymbol{y}'' \leq \boldsymbol{y}'$, it holds that $(\boldsymbol{x}'', \boldsymbol{y}'') \in \hat{\varphi}_{\text{RF-ACES}}$.

Proof. Since $(\boldsymbol{x}', \boldsymbol{y}') \in \hat{\varphi}_{\text{RF-ACES}}$, by definition we have:

$$y'_r \leq \frac{1}{L_{\max}} \sum_{l=1}^{L_{\max}} \hat{f}_{lr}^{\text{ACES}}(\boldsymbol{x}'), \quad \text{for all } r = 1, \ldots, s.$$

Moreover, each function $\hat{f}_{lr}^{\text{ACES}}(\cdot)$ is monotonically non-decreasing, which implies:

$$\hat{f}_{lr}^{\text{ACES}}(\boldsymbol{x}') \leq \hat{f}_{lr}^{\text{ACES}}(\boldsymbol{x}''), \quad \text{for all } l = 1, \ldots, L_{\max}, \text{ and } r = 1, \ldots, s.$$

Since $\boldsymbol{y}'' \leq \boldsymbol{y}'$, it follows that

$$y''_r \leq y'_r \leq \frac{1}{L_{\max}} \sum_{l=1}^{L_{\max}} \hat{f}_{lr}^{\text{ACES}}(\boldsymbol{x}') \leq \frac{1}{L_{\max}} \sum_{l=1}^{L_{\max}} \hat{f}_{lr}^{\text{ACES}}(\boldsymbol{x}''), \quad \text{for all } r = 1, \ldots, s.$$

Therefore, $(\boldsymbol{x}'', \boldsymbol{y}'') \in \hat{\varphi}_{\text{RF-ACES}}$. □

Proposition 6.2. [Convexity] $\hat{\varphi}_{\text{RF-ACES}}$ is a convex set.

Proof. Let $(\boldsymbol{x}', \boldsymbol{y}'), (\boldsymbol{x}'', \boldsymbol{y}'') \in \hat{\varphi}_{\text{RF-ACES}}$, and define

$$(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}) = \lambda(\boldsymbol{x}', \boldsymbol{y}') + (1 - \lambda)(\boldsymbol{x}'', \boldsymbol{y}''), \quad \text{with } \lambda \in [0, 1].$$

Since $(\boldsymbol{x}', \boldsymbol{y}'), (\boldsymbol{x}'', \boldsymbol{y}'') \in \hat{\varphi}_{\text{RF-ACES}}$, it follows that

$$y'_r \leq \frac{1}{L_{\max}} \sum_{l=1}^{L_{\max}} \hat{f}_{lr}^{\text{ACES}}(\boldsymbol{x}'), \quad y''_r \leq \frac{1}{L_{\max}} \sum_{l=1}^{L_{\max}} \hat{f}_{lr}^{\text{ACES}}(\boldsymbol{x}''), \quad \text{for all } r = 1, \ldots, s.$$

Now define $\hat{y}_r = \lambda y'_r + (1 - \lambda) y''_r$. Then,

$$\begin{aligned} \hat{y}_r &\leq \lambda \left(\frac{1}{L_{\max}} \sum_{l=1}^{L_{\max}} \hat{f}_{lr}^{\text{ACES}}(\boldsymbol{x}') \right) + (1 - \lambda) \left(\frac{1}{L_{\max}} \sum_{l=1}^{L_{\max}} \hat{f}_{lr}^{\text{ACES}}(\boldsymbol{x}'') \right) \\ &= \frac{1}{L_{\max}} \sum_{l=1}^{L_{\max}} \left[\lambda \hat{f}_{lr}^{\text{ACES}}(\boldsymbol{x}') + (1 - \lambda) \hat{f}_{lr}^{\text{ACES}}(\boldsymbol{x}'') \right]. \end{aligned}$$

Since each $\hat{f}_{lr}^{\text{ACES}}(\cdot)$ is concave, it holds that

$$\hat{f}_{lr}^{\text{ACES}}(\lambda \boldsymbol{x}' + (1 - \lambda)\boldsymbol{x}'') \geq \lambda \hat{f}_{lr}^{\text{ACES}}(\boldsymbol{x}') + (1 - \lambda) \hat{f}_{lr}^{\text{ACES}}(\boldsymbol{x}''), \quad \text{for all } l, r, \lambda \in [0, 1].$$

Therefore,

$$\hat{y}_r \leq \frac{1}{L_{\max}} \sum_{l=1}^{L_{\max}} \hat{f}_{lr}^{\text{ACES}}(\hat{\boldsymbol{x}}), \quad \text{with } \hat{\boldsymbol{x}} = \lambda \boldsymbol{x}' + (1-\lambda)\boldsymbol{x}''.$$

This implies that $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}) \in \hat{\varphi}_{\text{RF-ACES}}$ for all $\lambda \in [0,1]$, so $\hat{\varphi}_{\text{RF-ACES}}$ is convex. □

Regarding technical efficiency measurement, given an input–output bundle, in the single-output case, it is very natural to gauge efficiency as the ratio between the value of the production function evaluated at the given input vector and the value of the observed output. However, when dealing with multiple outputs, the different existing possibilities to measure technical efficiency introduce additional complexities and considerations, warranting a careful examination of the most appropriate approach for the given context. Next, we show how to compute the output-oriented radial measure when using the estimated technology $\hat{\varphi}_{\text{RF-ACES}}$. It is enough to calculate s ratios and compare their values, where s is the total number of output variables considered in the production context under analysis.

Proposition 6.3. [Closed-form output-oriented efficiency score] Let $(\boldsymbol{x}', \boldsymbol{y}') \in \mathbb{R}_+^{m+s}$. The output-oriented efficiency score associated with the technology $\hat{\varphi}_{\text{RF-ACES}}$ is given by:

$$\hat{\phi}(\boldsymbol{x}', \boldsymbol{y}') = \min_{r=1,\ldots,s} \left\{ \frac{\hat{f}_r^{\text{RF-ACES}}(\boldsymbol{x}')}{y_r'} \right\}.$$

Proof. The input–output bundle $(\boldsymbol{x}', \hat{\phi}(\boldsymbol{x}', \boldsymbol{y}')\, \boldsymbol{y}')$ belongs to $\hat{\varphi}_{\text{RF-ACES}}$. We know that $\hat{\phi}(\boldsymbol{x}', \boldsymbol{y}')\, \boldsymbol{y}' \leq \hat{\boldsymbol{f}}^{\text{RF-ACES}}(\boldsymbol{x}')$, which is equivalent to:

$$\hat{\phi}(\boldsymbol{x}', \boldsymbol{y}') \leq \frac{\hat{f}_r^{\text{RF-ACES}}(\boldsymbol{x}')}{y_r'}, \quad \text{for all } r = 1, \ldots, s.$$

Therefore,

$$\hat{\phi}(\boldsymbol{x}', \boldsymbol{y}') \leq \min_{r=1,\ldots,s} \left\{ \frac{\hat{f}_r^{\text{RF-ACES}}(\boldsymbol{x}')}{y_r'} \right\}.$$

Suppose now that

$$\left(\boldsymbol{x}', \min_{r=1,\ldots,s} \left\{ \frac{\hat{f}_r^{\text{RF-ACES}}(\boldsymbol{x}')}{y_r'} \right\} \boldsymbol{y}' \right) \in \hat{\varphi}_{\text{RF-ACES}}.$$

Since it is not possible to proportionally increase the outputs further while keeping inputs constant, we must have equality:

$$\hat{\phi}(\boldsymbol{x}', \boldsymbol{y}') = \min_{r=1,\ldots,s} \left\{ \frac{\hat{f}_r^{\text{RF-ACES}}(\boldsymbol{x}')}{y_r'} \right\}.$$

To confirm this, note that:

$$\min_{r=1,\ldots,s} \left\{ \frac{\hat{f}_r^{\text{RF-ACES}}(\boldsymbol{x}')}{y_r'} \right\} y_r' \leq \hat{f}_r^{\text{RF-ACES}}(\boldsymbol{x}'), \quad \text{for all } r = 1, \ldots, s.$$

Thus,

$$\left(\boldsymbol{x}', \min_{r=1,\ldots,s}\left\{\frac{\hat{f}_r^{\text{RF-ACES}}(\boldsymbol{x}')}{y'_r}\right\}\boldsymbol{y}'\right) \in \hat{\varphi}_{\text{RF-ACES}},$$

which completes the proof. □

Therefore, as stated by the proposition above, to determine the output-oriented radial measure, it is necessary, given a vector of inputs, to determine the model prediction based on RF for each of its corresponding outputs. Then, each of these values is compared to the observed value in the outputs through the calculation of ratios. Finally, the efficiency score we seek corresponds to the minimum value associated with these ratios.

It is important to note that the output-oriented radial efficiency measure defined in Proposition 6.3 relies on the assumption that each output estimator $\hat{f}_{lr}^{\text{ACES}}(\boldsymbol{x})$ satisfies both monotonicity and concavity. However, when incorporating additional elements—such as variable interactions, the relaxation of shape constraints in the first-stage estimation, or the application of the refinement phase described in Equation (5.23)—these structural properties may no longer be guaranteed. As a consequence, the closed-form expression in Proposition 6.3 is no longer valid in such settings. Nevertheless, an alternative approach remains viable. A DEA-type technology can still be constructed by following the steps outlined in Chapter 5, using the original input observations and the predicted outputs obtained after the ensemble—where, in this case, the refinement step is applied to each individual learner. This allows for the estimation of any standard efficiency measure within a flexible and theoretically grounded framework, even when shape constraints are partially relaxed or the prediction procedure deviates from the conditions required for Proposition 6.3 to hold.

More formally, let $\hat{y}_{ir}^{(l)}$ denote the maximum producible output given by the base learner $\hat{f}_{lr}^{\text{ACES}}(\boldsymbol{x}_i)$ after the refinement phase proposed in Equation (5.23). Then, for each learner $l = 1, \ldots, L_{\max}$, a technology $\hat{\varphi}_{\text{ACES}}^{(l)}$ can be constructed by applying Equation (5.24) to the subset of input vectors $\{\boldsymbol{x}_i\}_{i\in\mathcal{L}_l^*}$ and their associated refined output vectors $\{\hat{\boldsymbol{y}}_i^{(l)}\}_{i\in\mathcal{L}_l^*}$, where $\hat{\boldsymbol{y}}_i^{(l)} = (\hat{y}_{i1}^{(l)}, \ldots, \hat{y}_{is}^{(l)})$ and $\mathcal{L}_l^*$ denotes the bootstrap sample used to train the l-th base learner.

The ensemble-based estimator then defines a final technology by first computing the average predicted outputs for each observation across all learners:

$$\hat{y}_{ir}^{\text{RF}} = \frac{1}{|\mathcal{L}_i|}\sum_{l\in\mathcal{L}_i}\hat{y}_{ir}^{(l)}, \quad \text{where } \mathcal{L}_i = \{l : i \in \mathcal{L}_l^*\},$$

and constructing the set

$$\hat{\varphi}_{\mathrm{RF}} = \left\{ (\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}_+^{m+s} : \begin{array}{ll} x_j \geq \sum_{i=1}^{n} \lambda_i x_{ij}, & j = 1, \ldots, m, \\ y_r \leq \sum_{i=1}^{n} \lambda_i \hat{y}_{ir}^{\mathrm{RF}}, & r = 1, \ldots, s, \\ \sum_{i=1}^{n} \lambda_i = 1, & \\ \lambda_i \geq 0, & i = 1, \ldots, n. \end{array} \right\}. \tag{6.4}$$

To introduce additional flexibility and robustness in the construction of the production technology, one can define a family of partial technologies based on the ensemble of refined predictions obtained from the base learners. Rather than averaging predictions across learners, it is possible to consider the percentile value of each predicted output, thereby controlling the location of the estimated boundary. Let $p \in (0, 1]$ denote a fixed percentile level. For each unit $i = 1, \ldots, n$ and output $r = 1, \ldots, s$, define the p-percentile prediction as

$$\hat{y}_{ir}^{(p)} = \mathrm{Percentile}_p \left(\left\{ \hat{y}_{ir}^{(l)} : i \in \mathcal{L}_l^* \right\} \right), \tag{6.5}$$

where each $\hat{y}_{ir}^{(l)}$ denotes the prediction obtained from the l-th base learner after the refinement step. These values can be interpreted as empirical percentiles over the subset of learners for which unit i was included in the training sample.

Given the values $\hat{y}_{ir}^{(p)}$, a corresponding percentile-based production technology is defined as:

$$\hat{\varphi}_{\mathrm{RF}}^{(p)} = \left\{ (\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}_+^{m+s} : \begin{array}{ll} x_j \geq \sum_{i=1}^{n} \lambda_i x_{ij}, & j = 1, \ldots, m, \\ y_r \leq \sum_{i=1}^{n} \lambda_i \hat{y}_{ir}^{(p)}, & r = 1, \ldots, s, \\ \sum_{i=1}^{n} \lambda_i = 1, & \\ \lambda_i \geq 0, & i = 1, \ldots, n. \end{array} \right\}. \tag{6.6}$$

In Equation (6.6), lower values of p lead to more optimistic estimates of the production possibility set, as the predicted outputs are closer to the lower envelope of the ensemble. Conversely, higher values of p produce more demanding frontiers, reflecting more conservative assessments of technical efficiency. This approach is conceptually related to the order-m and quantile frontiers developed in the literature on robust nonparametric efficiency analysis (Cazals et al., 2002; Daouia and Gijbels, 2011), and it

provides a way to control the influence of extreme predictions within the ensemble. The resulting sets $\hat{\varphi}_{\mathrm{RF}}^{(p)}$ satisfy free disposability and convexity by construction, and standard DEA-type efficiency measures can be computed by projecting each unit onto the corresponding percentile technology.

Figure 6.1 illustrates a production function scenario involving 100 DMUs and a single input. The true production frontier, denoted as DGP, is shown in black and lies between the extreme cases defined by the lowest and highest percentiles of the RF-ACES estimators. Specifically, the lower envelope (in green) corresponds to the minimum prediction across all base learners for each input level ($p = 0$), while the upper envelope (in pink) represents the maximum prediction ($p = 1$). The estimator based on the mean—defined in Equation (6.4)—is shown in purple.

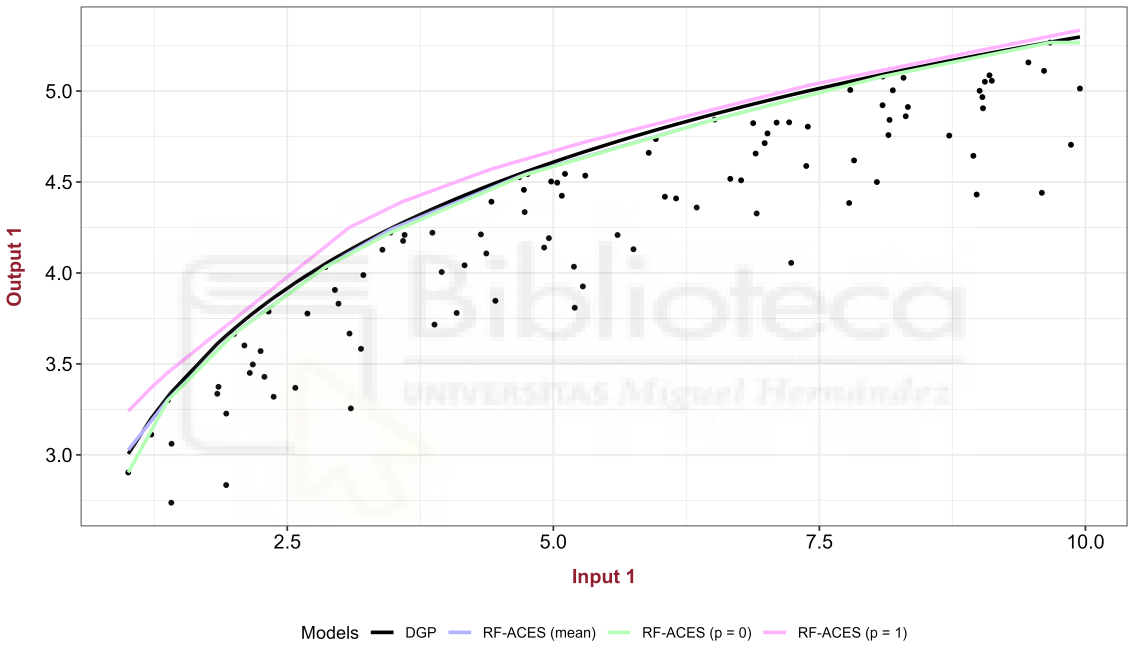

Figure 6.1: Comparison of different RF-ACES percentile-based technologies.

As shown, the lower and upper envelopes define a band within which the true DGP is fully contained. The mean-based estimator provides an accurate approximation to the true frontier. In this example, the RF-ACES mean estimator almost perfectly aligns with the DGP, underscoring its reliability as a central estimate of the underlying production technology. Additionally, the mean and minimum-based estimators are nearly indistinguishable, suggesting that the predictive distribution across base learners is strongly concentrated near the lower bound. The model was trained using $L_{\max} = 250$ base learners, each constructed from bootstrap resamples. Since the model includes only one input variable, the random selection of input subsets typically employed in RF methods was not applicable during the training phase.

Finally, the variable selection strategy described in Section 3.3 can be naturally incorporated into the RF-ACES framework. This method relies on comparing the out-of-bag (OOB) prediction error before and after permuting the values of a given input variable j. A significant increase in error after permutation indicates that the variable plays an important role in the prediction process. Although this approach offers a consistent mechanism to assess variable relevance, it is not computationally efficient, especially in high-dimensional settings.

In the upcoming sections, we demonstrate the application of the proposed technique on real-world datasets. Specifically, our aim is to highlight the usefulness of the method in generating predictions for output variables in production processes, estimating the underlying DGP inherent in the observed data, and even establishing a ranking of input variable importance within the given problem context. The case becomes particularly relevant when researchers work with a random sample drawn from an entire population of units. In such settings, the main focus shifts from assessing the efficiency of individual (and often anonymous) observations to understanding the production function of the broader sector—i.e., the DGP itself. To this end, we present two distinct empirical applications. The first focuses on the educational sector and analyzes data from Spanish schools participating in the PISA assessment, with the aim of modeling student performance. The second centers on the Spanish food industry and estimates the production technology of manufacturing firms based on financial and operational indicators. In both cases, the version of the algorithm used excludes variable interactions and the refinement phase, and estimates a separate production function for each output.

6.2 Empirical evaluation based on PISA data in Spain

In this part of the chapter, we will apply the newly developed algorithm to a real-world database from a public service and assess its usefulness in estimating the DGP, which is the unknown mechanism responsible for generating the sample, and making reliable output predictions for unobserved data. For demonstration purposes, we will use data gathered by PISA through the Organization for Economic Co-operation and Development (OECD). There has been a substantial increase in research initiatives, projects, and publications centered on utilizing rich public datasets over the past decades. Databases in education like PISA, TIMSS (Trends in Mathematics and Science Study), and PIRLS (Progress in International Reading Literacy Study) serve as foundations for numerous research efforts across various disciplines (see, for example, the works by Gustafsson, 2008; Kamens and McNeely, 2010; Johansson, 2016; Cordero et al., 2018; Steiner-Khamsi and Waldow 2018; Verger et al., 2019; Hernández-Torrano and Courtney, 2021). This growing trend underscores the importance placed upon comparative analyses and evidence-based approaches when addressing challenges faced by contemporary education systems.

The PISA program measures the ability of students close to finishing compulsory education to apply important academic skills and knowledge required for successful engagement in today's complex societies. Within our empirical study, we chose to analyze schools as the fundamental unit, following common practice in most practical applications concerning efficiency evaluation within the field of education (Johnes, 2015, Witte and López-Torres, 2017). By concentrating on schools, we ensure consistency with previous work and maintain relevance to current discourse surrounding educational institutions and their respective operational efficiencies. The specific dataset employed covers the year 2018, with a sample size of anonymous 1047 Spanish schools randomly chosen by the OECD. Estimating a specific technology for each autonomous community is essential due to the regional variability and distinct educational systems within Spain. This decentralized structure results in variations in educational policies and practices that can significantly impact student performance. By tailoring the technology to each region, we account for these differences and provide a more accurate and tailored analysis of educational efficiency. Moreover, policymakers in each region require precise, region-specific information to make informed decisions, as a one-size-fits-all approach could obscure critical regional differences and lead to ineffective or misinformed policies. Understanding these nuances is crucial for accurately interpreting our findings and developing targeted interventions to address specific challenges within Spain's diverse educational landscape.

The selection of input variables is a challenging decision, given the wide range of potential indicators available in the dataset. Most empirical studies aiming to measure school efficiency typically include variables reflecting human and capital resources (Worthington, 2001; Witte and López-Torres, 2017). Following this approach, our empirical study incorporates three main input variables. The first input variable, `EDUQUAL`, captures the quality of school resources. It is derived from school principals' responses concerning the availability of items such as computers, educational software, calculators, books, audiovisual materials, and laboratory equipment. The critical importance of educational quality in shaping student performance and educational outcomes is well-documented in the literature (see Hanushek and Woessmann, 2011). The second input, `TSRATIO`, measures the teacher–student ratio, computed as the number of teachers per hundred students. This variable reflects the human resources available in schools. Prior research has shown that smaller teacher–student ratios can enhance individual student support and improve learning outcomes (Finn and Achilles, 1999). The third input variable, `ESCS`, corresponds to the Economic, Social, and Cultural Status index, which captures the socio-economic background of students. It includes information about parents' occupations, educational resources, and cultural possessions at home. As students constitute the 'raw material' transformed through the educational process, including ESCS as an input is essential. Numerous studies have established a strong link between socio-economic background and academic achievement (e.g., Sirin, 2005). To ensure all input values are strictly positive, we applied a linear transformation to

ESCS by adding the minimum observed value of each variable across the sample. All three inputs have been previously used in the context of PISA data for school efficiency analysis. For instance, Aparicio et al. (2018) employed similar variables to assess efficiency across OECD countries using non-radial DEA methods.

Regarding outputs, we consider student performance in mathematics (`PVMATH`), reading (`PVREAD`), and science (`PVSCIE`) based on the standardized test scores provided by the PISA 2018 database. For each subject, we use the average of the ten plausible values available, in line with the recommendations of Aparicio et al. (2022a), who showed that such an approach yields reliable and robust efficiency estimates. To construct school-level variables from individual-level data, we aggregate student responses and performance metrics by computing weighted averages, where weights correspond to the student-level sampling weights provided by PISA. This procedure ensures that the aggregated values accurately represent the population, as recommended in the official technical documentation. Summary statistics—including means, standard deviations (in parentheses), and sample sizes per autonomous community—are presented in Table 6.1. Notably, the sample includes a greater number of schools from Madrid and the Basque Country. However, the use of the RF algorithm, which naturally incorporates subsampling techniques, mitigates the potential imbalance by generating diverse models from randomly drawn subsets, thus promoting robust and representative predictions.

The observed variability in both input and output variables across regions indicates significant disparities in educational resources and outcomes. Such differences highlight the need to investigate regional variations to gain a deeper understanding of the factors affecting educational performance in diverse contexts. Given that the dataset from PISA only contains a portion of the total population, our objective does not center around calculating the precise technical efficiencies of the observed schools. Rather, we aim to utilize the estimated education production function to predict outcomes for schools outside of the observed sample. In this sense, an intriguing situation for educational decision-makers could be optimizing the distribution of educational resources (`EDUQUAL`) and human resources (`TSRATIO`) so that schools can reach or even exceed certain thresholds in mathematics, reading and science scores. At this point, it is worth recognizing that the socioeconomic standing of students (`ESCS`), influenced primarily by the schools' locations, cannot be readily modified for this purpose.

Building upon the previous ideas, we estimate a separate technology for each autonomous community using the methodology outlined in this chapter. This strategy captures the unique characteristics, structural differences, and socio-educational complexities of each region, allowing for a more accurate and context-aware evaluation of educational efficiency. By tailoring the estimation to regional data, we avoid the limitations of assuming a homogeneous technology across territories, and instead acknowledge the heterogeneity inherent in educational systems shaped by diverse policies, resource allocations, and demographic factors.

Region	Size	EDUQUAL	TSRATIO	ESCS	PVMATH	PVREAD	PVSCIE
Andalusia	52	3.48	9.87	2.57	467.70	466.88	471.40
		(1.10)	(12.94)	(0.54)	(30.16)	(39.03)	(33.65)
Aragon	52	3.90	14.04	2.90	497.31	490.43	494.33
		(1.08)	(20.51)	(0.41)	(29.98)	(33.91)	(29.17)
Asturias	54	3.80	13.06	2.84	490.71	494.56	496.11
		(1.07)	(15.42)	(0.52)	(33.51)	(33.12)	(29.85)
Balearic Islands	53	3.67	14.78	2.76	482.01	478.83	482.28
		(0.97)	(18.10)	(0.49)	(29.43)	(31.13)	(28.80)
Canary Islands	53	3.51	9.05	2.51	460.41	471.90	469.61
		(1.09)	(3.91)	(0.50)	(33.28)	(34.80)	(32.51)
Cantabria	56	4.16	9.91	2.91	497.98	483.13	494.93
		(0.73)	(2.62)	(0.43)	(33.02)	(31.91)	(29.92)
Castile and Leon	58	3.95	14.47	2.85	500.67	494.61	499.54
		(1.06)	(20.07)	(0.40)	(34.13)	(33.17)	(29.83)
Castile-La Mancha	53	2.99	8.78	2.67	479.32	478.27	484.99
		(1.19)	(2.34)	(0.48)	(29.12)	(32.34)	(27.70)
Catalonia	50	3.97	15.15	2.99	488.47	483.12	487.10
		(1.12)	(19.33)	(0.50)	(35.90)	(39.70)	(37.09)
Extremadura	54	3.90	11.03	2.56	470.00	464.86	474.04
		(1.03)	(3.50)	(0.46)	(32.38)	(35.88)	(33.10)
Galicia	57	4.05	10.19	2.84	498.42	493.91	510.88
		(1.02)	(3.57)	(0.45)	(25.10)	(25.68)	(22.51)
La Rioja	44	3.99	8.45	2.73	492.64	461.38	484.84
		(1.00)	(2.77)	(0.45)	(37.21)	(42.61)	(35.25)
Madrid	142	4.10	11.21	3.25	497.19	483.29	495.98
		(0.94)	(15.96)	(0.63)	(39.75)	(46.35)	(36.76)
Murcia	52	3.63	8.89	2.52	474.97	480.26	480.63
		(1.04)	(4.92)	(0.50)	(32.86)	(38.23)	(35.74)
Navarre	49	4.11	11.30	2.92	502.26	471.97	492.02
		(0.97)	(10.06)	(0.45)	(34.62)	(42.54)	(35.55)
Basque Country	116	3.98	11.07	2.92	492.80	470.57	482.86
		(1.00)	(11.41)	(0.52)	(34.94)	(38.14)	(36.84)
Valencian Community	52	3.83	11.84	2.75	475.65	474.86	479.89
		(1.09)	(13.12)	(0.50)	(29.77)	(36.07)	(28.67)

Table 6.1: Summary statistics: averages with standard deviations in parentheses below.

Each individual ACES model is configured by setting the maximum number of terms in the output predictor expression, denoted by η, equal to the sample size. The forward selection process follows a stopping criterion whereby the addition of two new basis functions (BFs) is halted if the associated error reduction falls below 0.005. Additionally, the standard values for the `minspan` and `endspan` parameters proposed by Friedman are used. Regarding the RF configuration, the number of learners is limited to a maximum of 500. In the generation of BFs, one input variable—representing one-third of the total—is randomly selected at each iteration. This introduces additional variability and helps reduce model variance. The training process is terminated when the out-of-bag (OOB) error fails to improve by more than 1% relative to the average of the last 5, 10, and 25 OOB errors, on at least five separate occasions. This dynamic stopping rule ensures sufficient learning while preventing unnecessary computational overhead.

Figure 6.2 illustrates the progression of the OOB error for the model estimated using data from the Madrid region, complementing the earlier description of the stopping mechanism. As RF-ACES is computationally intensive—mainly due to the repeated constrained estimation steps performed within each learner—the stopping criterion is essential to balance model accuracy with efficiency. The curve reveals how predictive performance evolves as more learners are added, with the OOB error stabilizing after approximately 75 learners. This pattern supports the adequacy of the dynamic rule based on recent error trends.

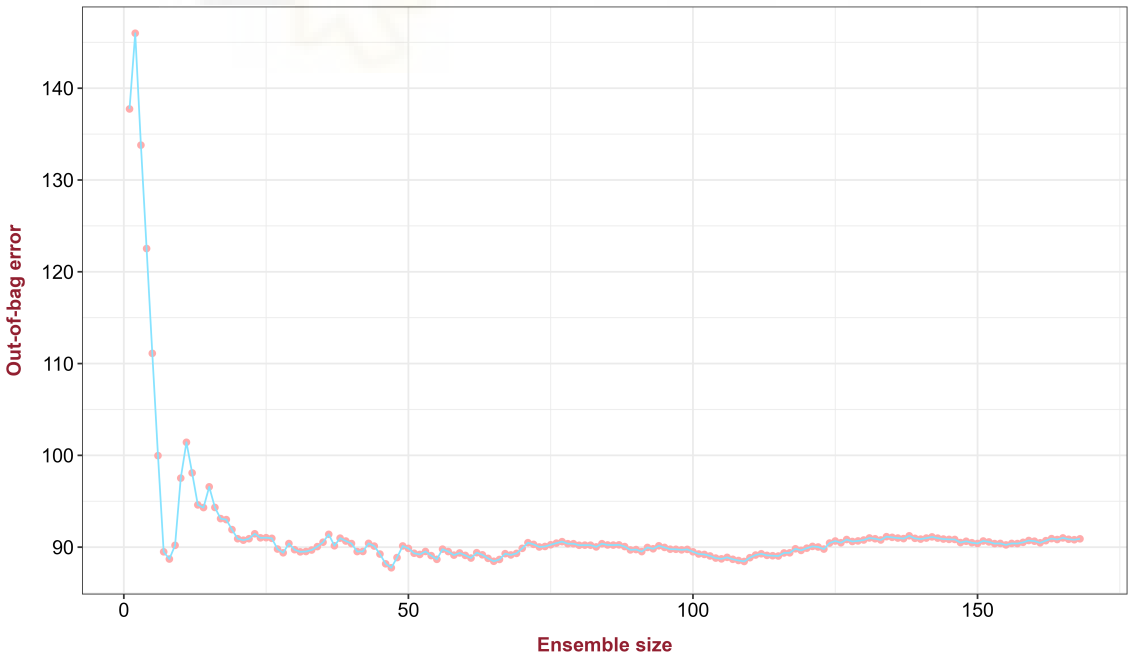

Figure 6.2: Evolution of the OOB error for the Madrid-based model.

Contrary to the traditional DEA approach, our methodology offers a significant advantage: it can predict outputs from a set of observed or even unseen inputs. To illustrate this, let us consider an unseen input profile: `EDUQUAL` = 3.86, `TSRATIO` = 11.35, and `ESCS` = 2.84, which are the average values across all schools in Spain. These values were chosen to construct a standardized example that highlights the model's predictive capabilities. It is important to emphasize that this particular input profile was not observed in the original dataset; it was selected solely to serve as a consistent benchmark for comparing predicted outputs across regions. Nonetheless, the methodology can be applied to any other input profile.

Region	**PVMATH**	**PVREAD**	**PVSCIE**
Andalusia	518.18	519.52	524.05
Aragon	537.93	535.49	535.91
Asturias	542.33	552.93	551.28
Balearic Islands	529.43	529.22	526.11
Canary Islands	518.40	536.03	528.88
Cantabria	536.05	528.39	534.13
Castile and Leon	549.77	546.08	544.38
Castile-La Mancha	523.94	525.12	522.93
Catalonia	540.42	539.84	541.11
Extremadura	529.23	536.84	532.94
Galicia	538.60	544.53	550.60
La Rioja	548.35	524.92	538.86
Madrid	532.77	541.88	537.54
Murcia	539.19	559.43	545.83
Navarre	543.18	527.39	535.58
Basque Country	545.92	533.51	531.60
Valencian Community	526.53	537.76	528.73

Table 6.2: Predicted outputs for an identical profile of inputs for each autonomous community.

From Table 6.2, several conclusions can be drawn. First, there is notable variability in educational performance across regions. For example, Castile and Leon, Asturias, and La Rioja exhibit relatively strong results across all three output variables (`PVMATH`, `PVREAD`, and `PVSCIE`), whereas regions such as Andalusia and the Canary Islands tend to show lower predicted scores, particularly in `PVMATH`. Other regions, like Madrid, Navarre, and the Basque Country, demonstrate strengths in specific domains: Madrid performs well in `PVREAD` and `PVSCIE`, while Navarre shows high values in `PVMATH` and `PVSCIE`. Importantly, it should be emphasized that neither traditional DEA nor bootstrapped DEA are capable of predicting outputs for a given input profile. This highlights a key advantage of the proposed methodology over conventional approaches.

Furthermore, we can investigate whether the efficiency score distributions obtained through DEA differ from those obtained by our methodology. In this scenario, we estimate the efficiency scores of schools in each region using their respective technology. Subsequently, we conduct the Li test (Simar and Zelenyuk, 2006) to compare the scores obtained through DEA with those derived from our methodology. All Li tests performed indicated significant differences between the distributions provided by DEA and the proposed methodology for all regions.

Figure 6.3 displays a density plot comparing the distributions of RF-ACES and DEA for the community of Madrid, as an example. In the figure, a shift towards the right is observed in the distribution of the efficiency score when using the new method compared to the standard method.

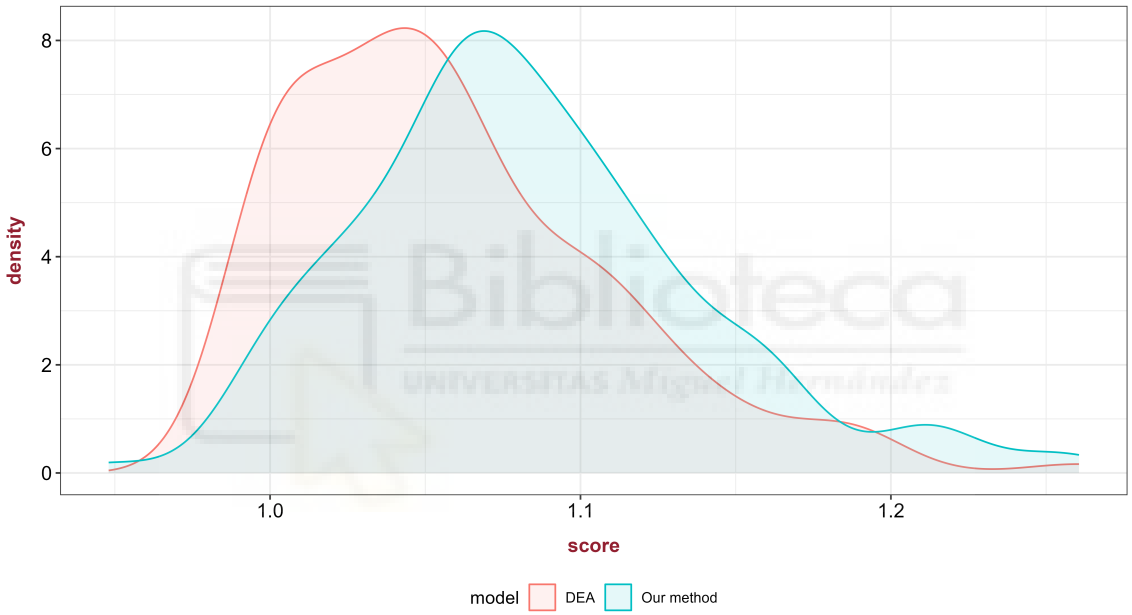

Figure 6.3: Density plot comparing efficiency scores between DEA and RF-ACES.

Finally, we compute input importance using our methodology. This step is fundamental for identifying the factors that most significantly influence the three output variables and for understanding the internal structure of the predictive models. By quantifying the relative contribution of each input, we provide insights into the drivers of educational performance. The results reveal a consistent pattern across regions: `ESCS` emerges as the most influential input in nearly all autonomous communities, highlighting the strong and pervasive impact of socio-economic background. `EDUQUAL` generally ranks second, reflecting the role of teacher qualification, while `TSRATIO` tends to be the least important. These findings reinforce the robustness of the models and offer a consistent interpretation framework for policymakers, emphasizing where interventions may yield the greatest improvements.

6.3 Empirical evaluation in the food industry

The food industry in Spain is a cornerstone of its economy, weaving together cultural heritage, economic health, and innovation. It encompasses a dynamic food processing and distribution sector that efficiently converts raw agricultural goods into a myriad of food products, ensuring availability both locally and globally. Characterized by a mix of traditional techniques and modern technologies, this sector supports a vast network of employment opportunities, ranging from agriculture and food processing to distribution and retail. This diverse ecosystem includes from small-scale farmers upholding age-old farming traditions to large-scale enterprises that leverage cutting-edge production technologies. Some applications of efficiency analysis in the study of the food industry across various regions include: Kumar and Basu (2008) in India, Flegl et al. (2022) in Mexico, Dadura and Lee (2011) in Taiwan or Machmud et al. (2019) in Indonesia.

Spain, as a nation, is divided into 17 autonomous communities, each with its own distinctive economic framework that significantly shapes the way companies operate. This regional heterogeneity introduces substantial complexity into the analysis, as economic policies, institutional settings, and market dynamics can vary widely across communities, directly affecting the performance and strategic behavior of businesses in the food sector. A thorough understanding of these regional economic nuances is therefore essential for accurately assessing the industry's structure and for designing targeted strategies to address the diverse challenges and opportunities present within Spain's economic landscape. The dataset used for this analysis comprises a sample of 917 companies in the food industry[1], each employing more than 50 workers, with data collected for the years 2022 and 2023[2].

The variables included in the dataset are designed to provide a comprehensive view of each company's financial and operational status. The output variable is represented by `operating income` (in millions of Euros), indicating the total income generated from the company's core business activities. On the input side, the dataset incorporates four key variables: `total assets` (in millions of Euros), reflecting the company's overall resource base utilized in its operations; `number of employees`, indicating the workforce size; tangible `fixed assets` (in millions of Euros), which includes physical assets like machinery, buildings, and equipment crucial for production; and `personnel expenses` (in millions of Euros), covering all costs related to staff, such as salaries, benefits, and training programs. These variables collectively provide a detailed snapshot of the companies' operational resources, workforce engagement, investment in physical infrastructure, and commitment to human capital, offering valuable insights into their operational efficiency, cost structure, and overall resource management within the food industry.

[1] Source: Iberian Balance Sheet Analysis System (SABI) https://sabi.bvdinfo.com.
[2] Only companies with complete records for both years were included to avoid missing data issues.

Table 6.3 presents the mean values for each variable under study, accompanied by the standard deviation (in brackets), and the number of observations for each autonomous community. Notably, the analysis has been conducted exclusively for the 10 autonomous communities that had at least 35 companies with more than 50 employees and recent data available, ensuring a relevant examination of the sector across regions with substantial industry presence.

Region	**Cases**	**Operating income** (in millions)	**Total assets** (in millions)	**Number of employees**	**Fixed assets** (in millions)	**Personnel expenses** (in millions)
Andalusia	85	89.52 (207.20)	61.24 (146.94)	159.18 (166.17)	14.04 (24.17)	5.64 (7.03)
Aragon	39	109.80 (174.77)	56.34 (70.23)	195.90 (257.82)	15.69 (22.18)	7.25 (10.40)
Castile-La Mancha	47	78.77 (159.11)	58.19 (129.40)	192.60 (285.74)	18.68 (28.28)	7.25 (12.30)
Castile and Leon	78	53.69 (66.43)	48.77 (63.21)	150.21 (159.91)	17.73 (31.99)	5.28 (5.86)
Catalonia	191	126.32 (313.03)	80.60 (207.68)	246.14 (516.17)	23.20 (68.10)	10.76 (28.70)
Galicia	75	89.12 (126.95)	51.78 (79.31)	177.96 (189.54)	12.94 (18.05)	5.79 (6.84)
Community of Madrid	75	74.65 (180.76)	71.69 (235.83)	211.65 (357.39)	16.17 (41.05)	8.75 (16.90)
Region of Murcia	59	86.10 (222.85)	71.13 (184.34)	263.36 (703.44)	22.87 (57.87)	9.20 (27.68)
Navarre	47	88.77 (114.66)	73.75 (125.21)	227.21 (301.03)	21.21 (37.74)	7.97 (11.92)
Valencian Community	97	62.31 (84.17)	41.03 (49.75)	191.44 (177.74)	15.28 (27.06)	6.76 (7.65)

Table 6.3: Summary of key descriptive statistics for the dataset.

The observed variability in both input and output variables across different regions suggests significant disparities in the economic health and operational scale of the food sector companies. Catalonia stands out with the highest number of companies (191), highlighting the relevance of the food industry (mainly meat) in this region. It also leads in operating income and total assets, indicating that these companies are not only numerous but also financially robust. The high average of fixed assets (23.20 million) further underscores the substantial investment in infrastructure and equipment, which is crucial for productivity and growth in this sector. Castile and Leon, Galicia and the Community of Madrid showcase a comparable business density within the food industry, yet they reveal diverse financial outcomes and operational efficiencies. Castile and Leon report the lowest operating income among the three, at 53.69 million euros, suggesting challenges in profitability and capital accumulation, which might be due to factors such as market access, cost structures, or business models. In contrast, Galicia and Madrid demonstrate higher operating incomes of 89.12 million euros and 74.65 million euros, respectively, suggesting more robust revenue streams. Galicia, with total assets of 51.78 million euros (similar to Castile and Leon), outperforms Madrid's operating income, despite Madrid having a more substantial asset base of 71.69 million euros. Madrid's scenario, with higher assets yet lower income than Galicia, points to variances in economic approaches or market conditions across these regions.

Galicia's ability to generate significant revenues with minimal investment in assets show indications of a model of operational excellence that offers valuable insights for optimizing the performance of the food industry. Similarly, Andalusia, the Region of Murcia and Galicia exhibit similar results in terms of operating income. Once again, companies in Galicia stands out for generating higher income with a lower level of resources. On the other hand, it is worth highlighting the situation in the region of Murcia, where companies with extensive assets and a large workforce do not seem to increase operating incomes, which underscores a significant efficiency gap when it comes to translating scale into revenue. Two other regions operating at the same income level are Castile-La Mancha and Navarre, with the former achieving higher income despite having fewer assets, underscoring its efficient resource utilization. Aragon presents a compelling case within the dataset, exhibiting one of the higher operating incomes at 109.80 million euros. This is achieved alongside a moderate asset base and a significant workforce, demonstrating the region's ability to outperform in terms of income, despite not having the largest asset base. Finally, the Valencian Community showcases businesses with relatively few assets yet a high number of employees per company. This configuration might be the reason behind the lower operating incomes observed in the region, suggesting a discrepancy between workforce size and asset leverage that could impact revenue generation efficiency.

Similar to the previous example, we estimate a separate technology for each autonomous community using the methodology outlined in this chapter. Specifically, we apply the RF-ACES approach, although in this case the input selection is not randomized at each iteration. This setting aligns with the original Bagging philosophy introduced by Breiman (1996), where the variance reduction is achieved through resampling alone, without injecting additional randomness in the model structure. For each individual learner, the maximum number of terms is set equal to the sample size, and a stopping rule is enforced when the error reduction from adding two new BFs falls below 0.01. The ensemble is limited to 300 learners. Training is terminated when the out-of-bag (OOB) error fails to decrease by more than 1% relative to the average of the last 5, 10, and 25 OOB errors on at least five separate occasions.

After fitting the model, we conduct a comparative analysis of the estimated technologies across the autonomous communities within the food industry sector. To this end, we compute efficiency scores for each DMU within its respective region using the RF-ACES model. Figure 6.4 offers a comprehensive view of DMU performance within each community, highlighting the heterogeneity of efficiency levels across regions. Specifically, this figure reveals that the Valencian Community and Navarre exhibit the lowest variability in DMU scores, indicating a consistent level of efficiency across all sampled units within these regions. Such uniformity contrasts with other areas, where company performance is more heterogeneous. In particular, regions such as Andalusia, Castile and Leon, and Catalonia display substantial variability in efficiency scores, reflecting divergent technological practices.

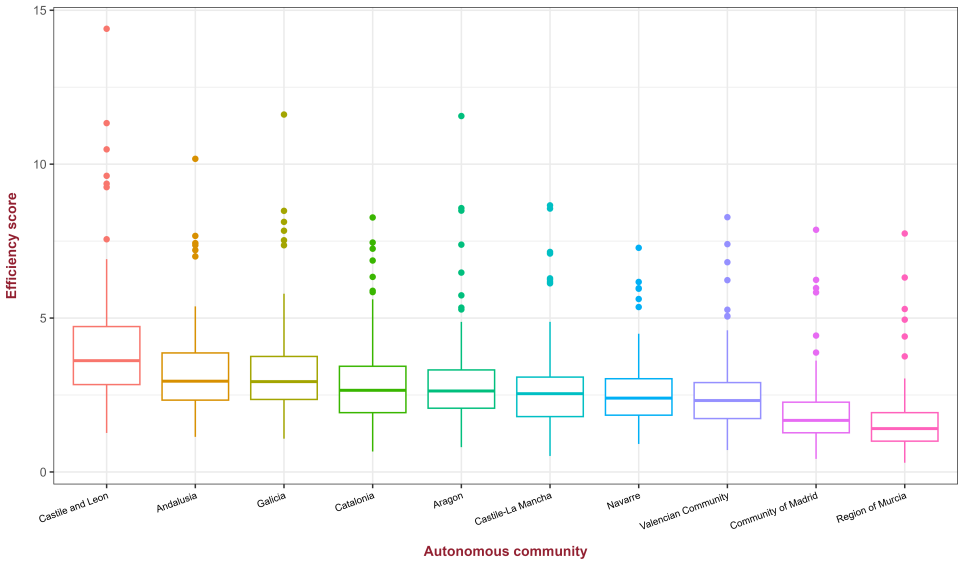

Figure 6.4: Boxplot comparing scores of each DMU projected against its own region's technology.

While this analysis highlights intra-regional trends in efficiency, it does not enable direct comparisons across autonomous communities. To address this limitation, we project a common set of DMUs onto the efficient frontiers estimated for each region, thereby allowing for the identification of inter-regional technological differences. Specifically, we calculate the efficiency score of each of the 917 DMUs using the RF-ACES model estimated for each of the 10 regions. This procedure yields a distance matrix that captures variations in efficiency scores across regional frontiers, offering valuable insights into the technological heterogeneity among regions.

Table 6.4: Distance matrix of vector of scores autonomous communities.

	AND	ARA	CLM	CAL	CAT	VLC	GAL	CMA	MUR	NAV
AND	0	26.65	39.56	40.55	66.83	32.88	18.91	32.75	55.41	63.07
ARA		0	41.33	34.55	63.97	35.62	31.37	30.24	43.90	50.57
CLM			0	16.92	92.11	32.06	39.69	28.11	44.59	54.93
CAL				0	87.11	34.49	41.80	26.06	39.45	49.21
CAT					0	92.18	75.59	88.65	101.54	105.78
VLC						0	33.04	15.90	32.33	39.42
GAL							0	31.93	58.26	65.35
CMA								0	32.92	39.51
MUR									0	13.96
NAV										0

The variation in efficiency scores across regions can be analyzed through the lens of company size. To this end, we compute the average distance between efficiency scores for companies with more than 250 employees (big companies), represented in the upper triangular matrix, and the average distance between efficiency scores for companies with fewer than 250 employees (medium-sized companies), shown in the lower triangular matrix. This distinction facilitates a detailed examination of how efficiency disparities between regions are influenced by the scale of the companies involved.

Table 6.5: Matrix of average distances of the score vectors between autonomous communities. Upper triangular matrix: big companies. Lower triangular matrix: medium-sized companies.

	AND	ARA	CLM	CAL	CAT	VLC	GAL	CMA	MUR	NAV
AND	0	0.35	0.75	0.92	0.60	0.61	0.21	0.77	1.03	1.41
ARA	0.52	0	0.55	0.70	0.79	0.44	0.40	0.59	0.85	1.24
CLM	0.79	0.73	0	0.24	1.28	0.34	0.66	0.34	0.46	0.95
CAL	0.80	0.67	0.29	0	1.47	0.42	0.79	0.34	0.30	0.61
CAT	1.04	1.38	1.58	1.65	0	1.21	0.72	1.37	1.62	2.01
VLC	0.80	0.55	0.70	0.72	1.83	0	0.55	0.24	0.45	0.81
GAL	0.28	0.51	0.90	0.94	1.01	0.83	0	0.74	0.95	1.30
CMA	0.69	0.44	0.54	0.48	1.71	0.38	0.72	0	0.32	0.65
MUR	1.03	1.17	1.24	1.11	2.01	0.92	0.90	0.52	0	0.39
NAV	1.74	1.37	1.45	1.31	2.75	1.00	1.80	1.09	0.26	0

This representation of the distances between efficiency scores reveals insightful patterns. For instance, the technological gap between the Valencian Community and the Community of Madrid appears to be largely driven by the performance of larger firms (0.38 vs. 0.24), whereas the discrepancy between the Community of Murcia and Andalusia is more pronounced among medium-sized companies (1.53 vs. 1.03). A particularly noteworthy case is Murcia itself, where the divergence from other regional technologies varies significantly depending on firm size, with average distances of 0.71 for large firms and 1.02 for medium-sized ones.

To delve into the observed technological differences between regions, we conduct a t-test at the 95% confidence level. This step aims to determine whether the mean differences in efficiency scores are statistically different from zero. This method assumes normality and homogeneity of variances in the data. Table 6.6 reports the average differences in efficiency scores between autonomous communities, while Table 6.7 shows the corresponding p-values obtained from the t-tests, confirming which differences are statistically significant.

Table 6.6: Mean differences in efficiency scores between autonomous communities.

	AND	ARA	CLM	CAL	CAT	VLC	GAL	CMA	MUR	NAV
AND	0	0.321	0.375	0.510	-0.953	0.764	0.002	0.696	1.431	1.665
ARA		0	0.054	0.189	-1.273	0.443	-0.319	0.375	1.110	1.345
CLM			0	0.135	-1.327	0.389	-0.373	0.321	1.056	1.291
CAL				0	-1.463	0.254	-0.508	0.186	0.921	1.156
CAT					0	1.717	0.955	1.649	2.383	2.618
VLC						0	-0.762	-0.068	0.667	0.902
GAL							0	0.694	1.429	1.663
CMA								0	0.735	0.969
MUR									0	0.235
NAV										0

Table 6.7: P-values from t-test comparing technological differences autonomous communities.

	AND	ARA	CLM	CAL	CAT	VLC	GAL	CMA	MUR	NAV
AND	1	0.000	0.000	0.000	0.000	0.000	0.920	0.000	0.000	0.000
ARA		1	0.231	0.000	0.000	0.000	0.000	0.000	0.000	0.000
CLM			1	0.000	0.000	0.000	0.000	0.000	0.000	0.000
CAL				1	0.000	0.000	0.000	0.000	0.000	0.000
CAT					1	0.000	0.000	0.000	0.000	0.000
VLC						1	0.000	0.000	0.000	0.000
GAL							1	0.000	0.000	0.000
CMA								1	0.000	0.000
MUR									1	0.000
NAV										1

The results of the t-test show significant differences in the scores obtained for nearly all pairs of technologies, with the sole exceptions being between Castile-La Mancha and Aragon, and between Galicia and Andalusia. Furthermore, the mean score differences suggest that the frontier estimated for companies in Catalonia dominates the others, indicating a higher degree of efficiency. Companies in Galicia and Andalusia also seem to perform well. Again, we can examine technological differences by company size:

Table 6.8: P-values from t-test comparing technological differences between autonomous communities. Upper triangular matrix: big companies. Lower triangular matrix: medium-sized companies.

	AND	ARA	CLM	CAL	CAT	VLC	GAL	CMA	MUR	NAV
AND	1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ARA	0.000	1	0.000	0.000	0.000	0.000	0.222	0.000	0.000	0.000
CLM	0.000	0.607	1	0.000	0.000	0.968	0.000	0.000	0.000	0.000
CAL	0.000	0.041	0.000	1	0.000	0.000	0.000	0.083	0.000	0.000
CAT	0.000	0.000	0.000	0.000	1	0.000	0.000	0.000	0.000	0.000
VLC	0.000	0.000	0.000	0.000	0.000	1	0.000	0.000	0.000	0.000
GAL	0.329	0.000	0.000	0.000	0.000	0.000	1	0.000	0.000	0.000
CMA	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1	0.000	0.000
MUR	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1	0.000
NAV	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1

These results suggest that the technological disparities between Galicia and Andalusia, as well as between Castile-La Mancha and Aragon, are largely driven by the performance of medium-sized companies. Conversely, for large companies, we observe minimal technological differences between regions such as Aragon and Galicia, the Valencian Community and Castile-La Mancha, or Castile and Leon and the Community of Madrid.

To complement the statistical analysis of regional technologies, we provide a visual representation of the average radial output score associated with each regional frontier. Figure 6.5 displays a heat map of Spain, where each autonomous community is colored according to the mean (radial output) efficiency score it assigns when projecting all 917 DMUs onto its own estimated technology. In particular, this map reveals clear disparities in the scoring behavior of regional technologies. Catalonia stands out as the most demanding region, with its frontier requiring, on average, the greatest proportional output expansion for DMUs to be deemed technically efficient. In contrast, Navarre and the Region of Murcia appear to define comparatively more accessible frontiers, assigning the lowest average scores.

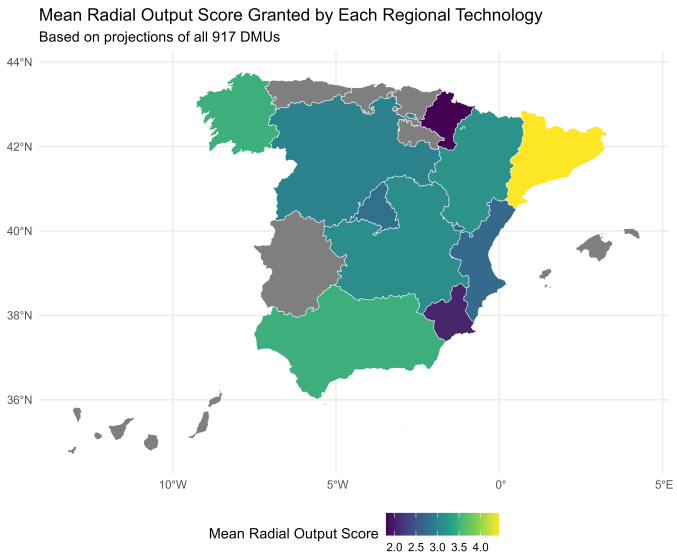

Figure 6.5: Mean radial output scores assigned by each regional technology.

Moreover, the RF-ACES methodology significantly improves prediction robustness by generating multiple optimal output predictions or efficiency score estimates for each unit. For instance, Figure 6.6 illustrates the density of estimates for a DMU belonging to the Valencian Community.

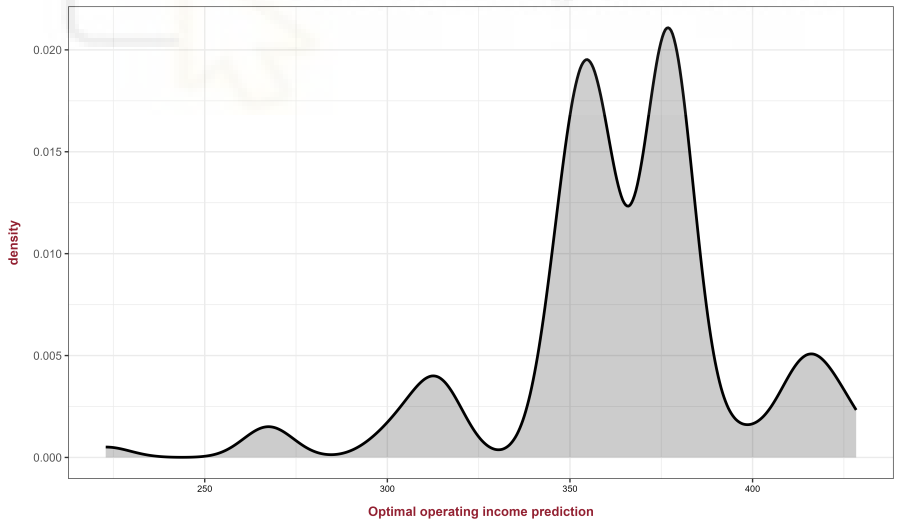

Figure 6.6: Density of estimates for a DMU belonging to the Valencian Community.

To gain deeper insights into the drivers of efficiency across regions, Table 6.9 reports the importance scores assigned to each input variable by the RF-ACES model. Alongside each score, we indicate the variable's rank within each region.

Region	Total assets	Number of employees	Fixed assets	Personnel expenses
Andalusia	30.65 (1^{st})	16.17 (2^{nd})	11.54 (3^{rd})	8.60 (4^{th})
Aragon	33.73 (1^{st})	20.10 (2^{nd})	14.36 (3^{rd})	10.67 (4^{th})
Castile-La Mancha	-3.07 (1^{st})	-7.42 (3^{rd})	-7.63 (4^{th})	-6.04 (2^{nd})
Castile and Leon	2.17 (3^{rd})	0.57 (4^{th})	2.57 (1^{st})	2.25 (2^{nd})
Catalonia	29.25 (1^{st})	17.80 (2^{nd})	13.76 (3^{rd})	10.80 (4^{th})
Valencian Community	30.49 (1^{st})	16.67 (2^{nd})	9.92 (3^{rd})	6.67 (4^{th})
Galicia	33.86 (1^{st})	20.04 (2^{nd})	12.95 (3^{rd})	9.26 (4^{th})
Community of Madrid	10.16 (1^{st})	3.96 (2^{nd})	1.71 (3^{rd})	0.43 (4^{th})
Murcia	4.10 (1^{st})	3.31 (2^{nd})	1.96 (3^{rd})	1.29 (4^{th})
Navarre	8.10 (1^{st})	4.89 (2^{nd})	3.73 (3^{rd})	3.13 (4^{th})

Table 6.9: Importance scores for each variable and region.

Results reveal a clear trend in the predominance of `Total assets` as the most important variable in most regions, consistently securing the top rank in importance. This is generally followed in sequence by `Number of employees`, `Fixed assets`, and `Personnel expenses`. Notably, the margin of importance between `Total assets` and the subsequent variable, `Number of employees`, is substantial, highlighting the critical role of asset management in reaching technical efficiency. However, an exception to this trend is observed in three autonomous communities—Castile and Leon, Murcia, and Navarre—where the importance scores of all four inputs cluster more closely together, suggesting a more balanced influence of these variables on efficiency.

6.4 Discussion and final remarks

This chapter has introduced RF-ACES, an ensemble-based extension of the ACES framework designed to improve the robustness and generalization of nonparametric frontier estimation. The method integrates bootstrap resampling and random input selection into the ACES procedure, generating a diverse set of shape-constrained estimators whose outputs are aggregated to produce a final, stabilized frontier. Each base model is estimated under monotonicity and concavity constraints, ensuring that the ensemble respects the economic structure of the production process. Unlike classical RFs, which rely on discontinuous decision trees, RF-ACES employs continuous, piecewise-linear spline models, yielding a frontier that satisfies the axioms of free disposability and convexity. By aggregating models that are individually consistent with production theory, the final estimator inherits both theoretical soundness while mitigating the impact of outliers and sampling variability, resulting in more stable efficiency estimates across diverse data configurations.

Beyond its predictive robustness, RF-ACES offers several methodological features that increase its practical value. Notably, it incorporates an internal mechanism for assessing variable relevance without relying on external feature selection procedures. By evaluating the impact of each input across the ensemble, the method yields empirical importance scores that capture both the stability and influence of each variable, thereby facilitating the identification of non-informative or redundant inputs. This functionality contributes to greater interpretability and dimensionality reduction—an important advantage in high-dimensional settings. In addition, RF-ACES introduces a percentile-based formulation of the estimated frontier, enabling users to shift the position of the boundary toward more optimistic or conservative configurations. This adjustable construction, conceptually related to order-m and quantile frontiers, provides added flexibility in scenarios where robustness to noise or model uncertainty is essential.

Despite these advantages, RF-ACES also introduces methodological complexities. Most notably, its computational cost increases significantly due to the need to estimate and aggregate multiple shape-constrained models, each involving a full forward selection process. To address this issue, the next chapter introduces Q-ACES, a simplified variant specifically designed to reduce computational time, offering a scalable alternative in scenarios where ensemble methods may be impractical. Finally, while the output of RF-ACES is smooth and interpretable at the aggregate level, the internal structure of the ensemble becomes opaque, limiting the traceability of individual BFs and knot configurations. Thus, RF-ACES trades some interpretability for improved robustness, generalization, and flexibility—positioning it as a powerful and adaptable tool for modern efficiency and productivity analysis.

Chapter 7

Quick-Adaptive Constrained Enveloping Splines

In Section 1.2 we have thoroughly discussed the critical issue of variable selection in Data Envelopment Analysis (DEA), emphasizing its significant influence on the robustness and interpretability of efficiency estimates. The risk of model misspecification, particularly in high-dimensional scenarios, was highlighted, along with its consequences such as inflated efficiency scores and diminished discriminatory power—commonly described as the curse of dimensionality. To mitigate these effects, the literature has proposed two broad types of responses: methodological strategies to reduce the number of variables and heuristic guidelines to ensure an adequate ratio between variables and DMUs. On the methodological side, several approaches aim to improve model specification and reduce dimensionality without compromising estimation accuracy. These include aggregation-based methods and statistical feature selection techniques, such as stepwise procedures, regression-based methods, and regularization approaches. However, these techniques may suffer from computational inefficiency and sensitivity to data structure. As a complementary response, various "rules of thumb" have been introduced to determine the minimum number of DMUs required given the number of inputs and outputs, thereby constraining dimensionality indirectly. Examples include the guidelines proposed by Boussofiane et al. (1991), Friedman and Sinuany-Stern (1998), Homburg (2001), Dyson et al. (2001), and Cooper et al. (2007). Yet, as noted by Wilson (2018), these rules, despite being widely referenced, lack a sound theoretical foundation and offer limited guidance when the goal is to identify a truly informative set of inputs.

To illustrate the impact of the curse of dimensionality in DEA, we present several simulations using a single output and varying numbers of inputs. Figure 7.1 compares the distribution of DEA efficiency scores with the true efficiency scores derived from the data generating process (DGP), calculated using the output-oriented radial measure defined by Equation (3.13). The DGP represents the maximum achievable output for a given resource configuration. In each scenario, the sample size is set to 20 times the number of input and output variables, surpassing the "rules of thumb" suggested in the literature. Nevertheless, Figure 7.1 clearly illustrates that DEA tends to produce overly optimistic efficiency scores as the number of variables increases relative to the sample size. This is particularly evident in the last three cases, where a significant portion of the DEA efficiency scores cluster at the lower bound of 1, incorrectly classifying many DMUs as fully efficient. In contrast, the DGP efficiency scores exhibit greater variability, reflecting the true performance of the DMUs. This discrepancy highlights the inherent bias and diminished discriminatory power of DEA when applied to high-dimensional datasets with inadequate sample sizes.

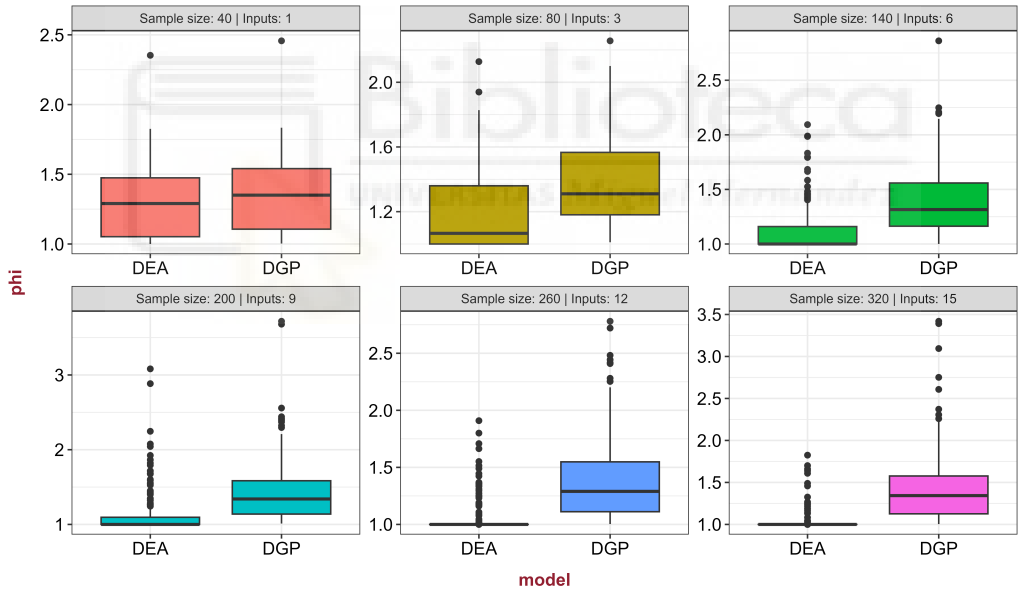

Figure 7.1: Comparison of DEA and true efficiency scores across different scenarios.

This chapter introduces Quick-Adaptive Constrained Enveloping Splines (Q-ACES) to address the limitations exposed in the preceding analysis. To that end, our new algorithm simultaneously mitigates overfitting and dimensionality by selecting the most relevant features while ensuring accurate output estimation. Our method begins by constructing a set of basis functions (BFs), each depending on one or more inputs. The relevance of each input is then assessed based on its contribution within these BFs, providing a systematic, data-driven criterion for feature selection. This enables decision-

makers to identify the inputs that most significantly influence efficiency outcomes. In addition, we incorporate strategies such as variable prioritization and knot reduction to improve scalability in high-dimensional settings and large datasets. Unlike the approaches described in previous chapters or standard Multivariate Adaptive Regression Splines (MARS) (Friedman, 1991), which greedily select the optimal BF at each iteration of the forward phase, our method reduces computational costs by restricting the number of candidate BFs considered at each step. This is achieved through an adaptive allocation strategy that prioritizes high-impact variables and assigns knots proportionally to their contribution in reducing estimation error.

7.1 Acceleration strategies

The forward step of MARS algorithm is computationally demanding, as it requires sequentially selecting optimal knot locations by exhaustively evaluating all candidate BFs at each iteration. As the number of candidate BFs grows exponentially with the number of inputs and sample size, the algorithm can quickly become computationally prohibitive (Friedman, 1991). Similarly, our approach selects a reflected pair from (3.33) and solves the optimization problem defined in (5.10) at each step. To highlight the computational demand, consider a scenario with 150 DMUs, 1 output, 5 inputs, and a degree of interaction set to 2, yielding 15 features. In this case, model (5.10) must be solved approximately 2,250 times per iteration. To mitigate this issue, two strategies can be applied: (i) reduce the number of variables, and/or (ii) limit the number of candidate knots per variable.

Feature selection methods are essential in regression analysis for excluding predictors based on their influence on the response variable. These methods prevent overfitting while improving interpretability and generalization by retaining only the most relevant predictors. In machine learning (ML), feature selection approaches are typically divided into three categories: filter, embedded, and wrapper methods. Each category has distinct ways of assessing relevance and integrating with the modeling process (Guyon and Elisseeff, 2003). Notice that these approaches mark a significant departure from traditional DEA practices, which typically evaluate inputs and outputs based on their impact on estimated efficiency scores. Filter methods rank variables independently of the model by using measures like correlation, mutual information, or statistical tests. Embedded methods incorporate feature selection into the model-building process. Examples include LASSO (Tibshirani, 1996) and MARS (Friedman, 1991). LASSO performs feature selection by adding an L_1-penalty to the objective function, shrinking the coefficients of less important predictors to zero. On the other hand, MARS assesses variable importance by quantifying each variable's contribution to reduce the residual sum of squares through the addition or removal of BFs. Finally, wrapper methods evaluate subsets of variables by iteratively building and testing models, achieving high accuracy but requiring substantial computational effort.

The following subsections present three complementary methods to alleviate the computational burden of ACES. The core idea behind these methods is to incorporate variable importance measures to prioritize predictors and guide the reduction of candidate BFs. First, we propose a correlation-based filtering technique to only consider variables strongly related to the output. Second, we introduce a neighborhood approach for identifying potential knot locations using the output-oriented radial measure defined in Equation (3.13). This creates a reduced set of candidate BFs, alternative to (3.33), by eliminating less relevant knots. Third, we describe an embedded method based on the percentage of residuals minimized for each variable to adaptively remove candidate knot locations for less significant variables. In practice a practitioner can apply any one or two of these stages, but whenever several are combined they should be executed in the exact order listed above, which defines the full Q-ACES pipeline.

7.1.1 Input reduction through a correlation-based approach

The first method employed to reduce the computational complexity of ACES focuses on variable filtering through correlation-based metrics. This approach uses Spearman's rank correlation (Spearman 1961) and Kendall's tau (Kendall 1938) to prioritize variables based on their relationship with the output. Unlike traditional regression techniques, which often rely on linear assumptions, Spearman and Kendall correlations are non-parametric, making them particularly suitable for assessing monotonic relationships in non-linear settings, such as the shape-restricted regression framework employed in ACES. By filtering out variables with weak associations to the response variable, this method significantly reduces the number of candidate BFs considered in each iteration, thus improving computational efficiency.

The procedure is outlined as follows. First, for each input (and interaction) variable in the dataset, we compute Spearman's rank correlation and Kendall's tau with the output. Next, we establish two thresholds to identify weak correlations: the smaller value between 0.1 and the 20th percentile of positive Spearman correlation values, and a similar threshold for Kendall's tau. This dual-threshold approach ensures flexibility: 0.1 serves as a conservative cut-off to retain moderately relevant variables, whereas the 20th percentile accommodates cases where all correlations are weak. Negative correlations, indicating non-increasing relationships between inputs and outputs, are directly flagged for removal. A variable is excluded if both its Spearman and Kendall correlation values fall below their respective thresholds. In the case of multiple outputs, it suffices that the variable meets the threshold for at least one output to be retained. Additionally, interactions between inputs are preserved even if the individual variables involved are excluded due to low correlation with the output. This systematic approach ensures that only genuinely non-informative variables are removed.

7.1.2 Neighborhood-based knot set reduction using DEA

The second method for reducing the computational complexity of ACES employs DEA as an oracle to narrow the search space for potential knot locations. This approach is based on the idea that efficient DMUs often highlight key changes in input-output relationships, similar to how knots in MARS identify shifts in data patterns within regression models. Therefore, these units naturally indicate regions where variations in the data's structure may be present, making them a logical focus for identifying potential knot locations.

Building on this premise, the method identifies neighborhoods around efficient DMUs, leveraging DEA's insights while maintaining independence from its results. This strategy addresses concerns raised in recent literature about the inferential limitations of DEA (see, for example, Esteve et al. 2020 and España et al. 2024). Hence, the objective is twofold: first, to identify promising regions for placing knots, and second, and more importantly, to exclude areas with low potential for capturing meaningful variations in data trends. This ensures computational efforts are concentrated on the most relevant areas. However, we acknowledge that this approach might be subject to dimensionality challenges; specifically, if DEA identifies many efficient DMUs, the reduction in the number of candidate knots will be minimal, potentially limiting computational gains.

Our method proceeds as follows. First, we compute efficiency scores $\phi_i \in [1, \infty)$, $i = 1, ..., n$, by using the DEA model defined by (3.13) in the original m-dimensional input space. Next, we define E as the set of DMUs achieving technical efficiency, i.e., $\phi_i = 1$. This subset represents technical efficient units and serves as the basis for the knot reduction. For each efficient DMU ($e \in E$), its corresponding input profile $\boldsymbol{x}_e$ is considered as a candidate for potential knot locations. Finally, to expand the knot locations set beyond DEA results, we can define a neighborhood around $\boldsymbol{x}_e$ as $\mathcal{N}_\omega(\boldsymbol{x}_e) = \left\{\boldsymbol{x} \in \mathbb{R}_+^{m_1} : \|\boldsymbol{x} - \boldsymbol{x}_e\|_2 \leq \omega\right\}$, where ω controls the radius of the neighborhood.

In addition to the previous method, we introduce an alternative approach that uses adjacent DMUs to define neighborhoods for each dimension independently. This method offers two key advantages. First, it eliminates the need for a user-defined parameter to control the neighborhood size (ω). Second, neighborhoods are constructed dimension by dimension, allowing the inclusion of DMUs that may be distant in the full-dimensional space but remain relevant within specific dimensions. As a result, knot locations are spread across the entire input space, reducing the reliance on DEA results.

Then, for each efficient DMU, a neighborhood in the m-dimensional input space can be defined as:

$$\mathcal{N}(x_{ej}) = \left[x_{ej}^-, x_{ej}, x_{ej}^+\right], \quad j = 1, \ldots, m, \tag{7.1}$$

where x_{ej}^- and x_{ej}^+ are the values of the nearest observed inputs in dimension j that are less than and

greater than x_{ej}, respectively. In this way, we can redefine (3.33) as follows:

$$\mathcal{C}^* = \{(x_j - k)_+, (k - x_j)_+\}_{k \in \bigcup_{e \in E} \mathcal{N}(x_{ej}),\ j=1,\ldots,m}. \tag{7.2}$$

Furthermore, this approach has been extended to include variable interactions. In this extension, inputs from efficient DMUs (in the original space) are multiplied based on the degree of interaction, and neighborhoods for these new variables are determined using the resulting values.

Finally, Figure 7.2 illustrates the process of knot location selection using a sample of 150 DMUs and two input variables: x_1 and x_2 (interactions are not considered here). In this example, the number of candidate knot locations is 300. The (original) complete set of potential knot locations set is depicted on lines 2 and 3. In contrast, lines 1 and 4 show the reduced set of candidate knots, obtained by applying the neighborhood-based reduction procedure described in this section.

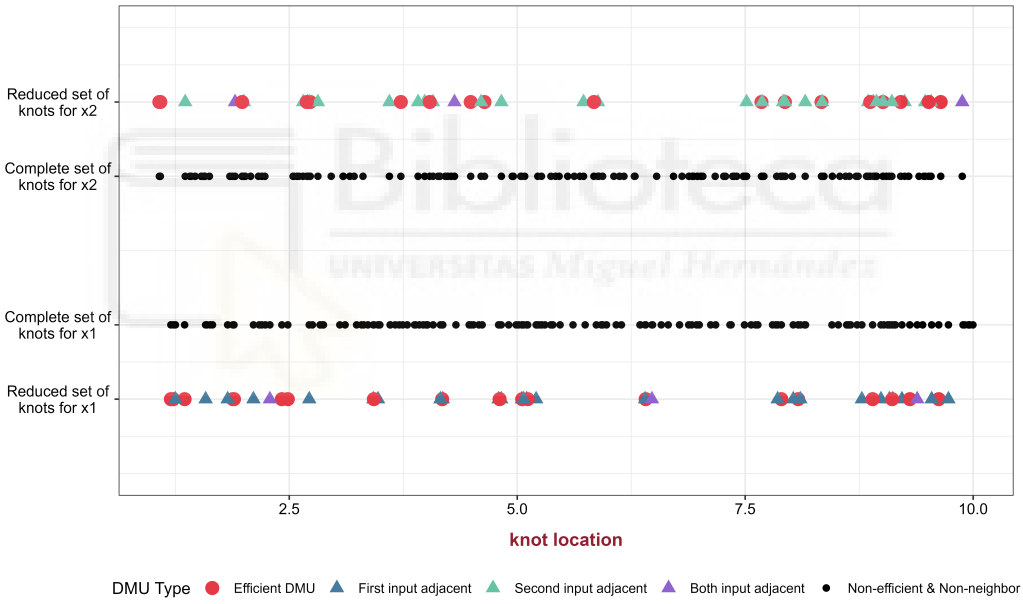

Figure 7.2: Visualization of efficient DMUs and their neighborhoods in the input space.

7.1.3 Adaptive knot set reduction based on feature performance

This section introduces an embedded variable importance approach to adaptively refine knot locations based on input relevance. Starting from the complete or reduced set of candidate knots defined in Equation (3.33) or (7.2), respectively, a random subset of knots is selected for each variable at every iteration. The subset size is dynamically adjusted to prioritize variables with greater potential to reduce residuals or the lack-of-fit (LOF) criterion.

Let $\hat{f}^{\text{ACES}}(\boldsymbol{x};\boldsymbol{\alpha},\boldsymbol{\beta},\mathcal{K})$ denote the ACES regression model at any forward step, parametrized by coefficients $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, and a set of selected knots $\mathcal{K}$, as defined in (3.42). To expand the model, consider a candidate knot $\overline{k} \in \Pi_j^{(l)}$ for variable j at iteration l, where $\Pi_j^{(l)}$ denotes the available knots for that variable. Let $\text{LOF}\left(\hat{f}^{\text{ACES}}\left(\boldsymbol{x};\boldsymbol{\alpha},\boldsymbol{\beta},\mathcal{K}\cup\overline{k}\right)\right)^{(l)}$ be the potential LOF achievable in iteration l by the model $\hat{f}^{\text{ACES}}(\boldsymbol{x};\boldsymbol{\alpha},\boldsymbol{\beta},\mathcal{K})$ when a new candidate knot $\overline{k}$ is included. Finally, let also $\text{LOF}^*\left(\hat{f}^{\text{ACES}}\left(\boldsymbol{x};\boldsymbol{\alpha},\boldsymbol{\beta},\mathcal{K}\right)\right)^{(l-1)}$ be the minimum LOF obtained by model $\hat{f}^{\text{ACES}}(\boldsymbol{x};\boldsymbol{\alpha},\boldsymbol{\beta},\mathcal{K})$ at iteration $l-1$.

Then, the relative reduction in the LOF achieved for input j at iteration l is defined as follows:

$$\Delta_j^{(l)} = 1 - \frac{\text{LOF}\left(\hat{f}^{\text{ACES}}\left(\boldsymbol{x};\boldsymbol{\alpha},\boldsymbol{\beta},\mathcal{K}\cup\overline{k}\right)\right)^{(l)}}{\text{LOF}^*\left(\hat{f}^{\text{ACES}}\left(\boldsymbol{x};\boldsymbol{\alpha},\boldsymbol{\beta},\mathcal{K}\right)\right)^{(l-1)}}. \tag{7.3}$$

To give more emphasis to recent iterations while still considering the influence of earlier ones, we compute a weighted reduction using an exponential decay factor. This approach ensures that recent reductions have a stronger influence on the overall importance score, while earlier iterations contribute with decreasing weight. The exponentially weighted reduction is defined as:

$$\overline{\Delta}_j^{(l)} = \frac{\sum_{b=1}^{l} \lambda^b \cdot \Delta_j^{(l+1-b)}}{\sum_{b=1}^{l} \lambda^b}, \qquad \text{with } \lambda = 0.95. \tag{7.4}$$

To facilitate comparisons across variables, we normalize the weighted reductions as follows:

$$\mathcal{S}_j^{(l)} = \frac{\overline{\Delta}_j^{(l)}}{\max_j\left(\overline{\Delta}_j^{(l)}\right)}. \tag{7.5}$$

Based on this normalized importance score, the variable j is activated in iteration l if:

$$\mathcal{S}_j^{(l)} \geq \frac{1}{2} \cdot \mathcal{S}_{(2)}^{(l)}, \tag{7.6}$$

where $\mathcal{S}_{(2)}^{(l)}$ denotes the second-largest value of $\mathcal{S}_j^{(l)}$ across all j. The use of the second-highest value rather than the maximum ensures that at least the top two variables remain active in each iteration, preventing overly aggressive pruning in early stages when importance scores may be concentrated on a few predictors.

Additionally, the set of candidate knots is randomly sampled from $\Pi_j^{(l)}$ with a proportional size of:

$$\mathcal{S}_j^{(l)} \cdot \left|\Pi_j^{(l)}\right|, \tag{7.7}$$

where $\left|\Pi_j^{(l)}\right|$ represents the cardinality of set $\Pi_j^{(l)}$. Thus, the better the performance of a variable in the most recent iterations (even if it has not been explicitly selected), the lower its likelihood of being discarded and the greater the proportion of knots assigned to it for further consideration.

In this way, the proposed methodology dynamically adjusts the activation of variables and the size of the candidate knot sets in each iteration. It exploits variable importance measures to concentrate computational efforts on the most significant predictors while minimizing the influence of less relevant ones. Furthermore, the criteria for deactivating a variable in a given iteration can be tailored to specific situations or fine-tuned using cross-validation techniques, offering greater flexibility and adaptability for a wide range of applications.

7.2 Integrating input importance in efficiency measurement

Let $x_1, \ldots, x_{m_1}$ denote the original input variables in $\mathbb{R}^{m_1}$. Consider a set of n DMUs, where each unit i is characterized by an observed input-output pair $(\boldsymbol{x}_i, \boldsymbol{y}_i) \in \mathbb{R}_+^{m_1+s}$, randomly drawn from a true population. Our goal is to identify the optimal input-output combinations $(\boldsymbol{x}_i^*, \boldsymbol{y}_i^*) \in \mathbb{R}_+^{m_{\text{opt}}+s}$, where $m_{\text{opt}} \leq m_1$, representing the true efficient frontier. This framework assumes that irrelevant inputs may be present in the observed data, with our method serving to exclude them.

Specifically, the proposed approach estimates the actual (theoretical) outputs $(\boldsymbol{y}_i^*)$ by computing approximations $(\hat{\boldsymbol{y}}_i)$ that more accurately reflect the true production capabilities of each DMU following the procedure described in Chapter 5. Simultaneously, it identifies a subset of relevant input variables $\{x_j^*\}_{j=1}^{m_{\text{opt}}} \subseteq \{x_j\}_{j=1}^{m_1}$, where the observed values are assumed to represent an efficient allocation of resources. This process results in a refined dataset $\{(\boldsymbol{x}_i^*, \hat{\boldsymbol{y}}_i)\}_{i=1}^n$, serving as the basis for subsequent efficiency analysis. We then introduce a novel methodology for measuring technical efficiency by applying any DEA model to a technology constructed from this reduced set of relevant inputs and the estimated output vectors.

Following the procedure described in Section 5.2, the input space is expanded from $\mathbb{R}^{m_1}$ to $\mathbb{R}^{m_1+m_2}$ by incorporating the interaction variables $\{z_1 = x_{m_1+1}, z_2 = x_{m_1+2}, \ldots, z_{m_2} = x_{m_1+m_2}\}$. At the end of the forward selection process, we obtain a set of selected knots $\mathcal{K}$ as defined in (3.42), with each K_j corresponding to a variable in the expanded space. Specifically, $K_1, \ldots, K_{m_1}$ are associated with original inputs, while $K_{m_1+1}, \ldots, K_{m_1+m_2}$ refer to interaction terms.

In this way, the inclusion criterion for variable x_j in the set of relevant variables is as follows:

$$x_j \in \{x_j^*\}_{j=1}^{m_{\text{opt}}} \quad \text{if} \quad (K_j \neq \varnothing) \ \vee \ (\exists j' > m_1 \ \text{ such that } \ K_{j'} \neq \varnothing \ \text{ and } \ x_{j'} \text{ involves } x_j)\,. \tag{7.8}$$

This inclusion criterion ensures that an input is selected either by contributing directly to the model or through its participation in at least one relevant interaction term.

Given the new set of data $\{(\boldsymbol{x}_i^*, \hat{\boldsymbol{y}}_i)\}_{i=1}^n$, Equation (3.9) is reformulated for the multi-output case as:

$$\hat{\varphi}_{\text{ACES-CRS}} = \left\{ (\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}_+^{m_{\text{opt}}+s} \;\middle|\; \begin{array}{ll} x_j^* \geq \displaystyle\sum_{i=1}^n \lambda_i x_{ij}^*, & j = 1, \ldots, m_{\text{opt}}, \\ y_r \leq \displaystyle\sum_{i=1}^n \lambda_i \hat{y}_{ir}, & r = 1, \ldots, s, \\ \lambda_i \geq 0, & i = 1, \ldots, n \end{array} \right\}. \tag{7.9}$$

where x_j^* is the j-th relevant input variable, and $\hat{y}_{ir}$ denotes the estimated maximum output for output r and DMU i. By adding the constraint $\sum_{i=1}^n \lambda_i = 1$ to Equation (7.9), the technology set is extended to allow for variable returns to scale (VRS), resulting in the formulation $\hat{\varphi}_{\text{ACES-VRS}}$. This version of the technology set is defined analogously to Equation (5.24), but restricted to the subset of relevant variables only. Accordingly, to assess the efficiency of a DMU with input-output bundle $(\boldsymbol{x}_0, \boldsymbol{y}_0)$ under $\hat{\varphi}_{\text{ACES-VRS}}$ using the output-oriented radial measure, the following Linear Programming model (LP) must be solved:

$$\max_{\phi, \boldsymbol{\lambda}} \phi \tag{7.10}$$

subject to:

$$\begin{aligned} \sum_{i=1}^n \lambda_i x_{ij} &\leq x_{0j}, \quad j = 1, \ldots, m_{\text{opt}}, \\ \sum_{i=1}^n \lambda_i \hat{y}_{ir} &\geq \phi y_{0r}, \quad r = 1, \ldots, s, \\ \sum_{i=1}^n \lambda_i &= 1, \\ \lambda_i &\geq 0, \quad i = 1, \ldots, n. \end{aligned}$$

Finally, Figure 7.3 illustrates the construction of two technological frontiers from the observed data sample (solid circles). Applying the ACES methodology, a transformed dataset is generated (square crosses), from which both CRS and VRS frontiers are estimated.

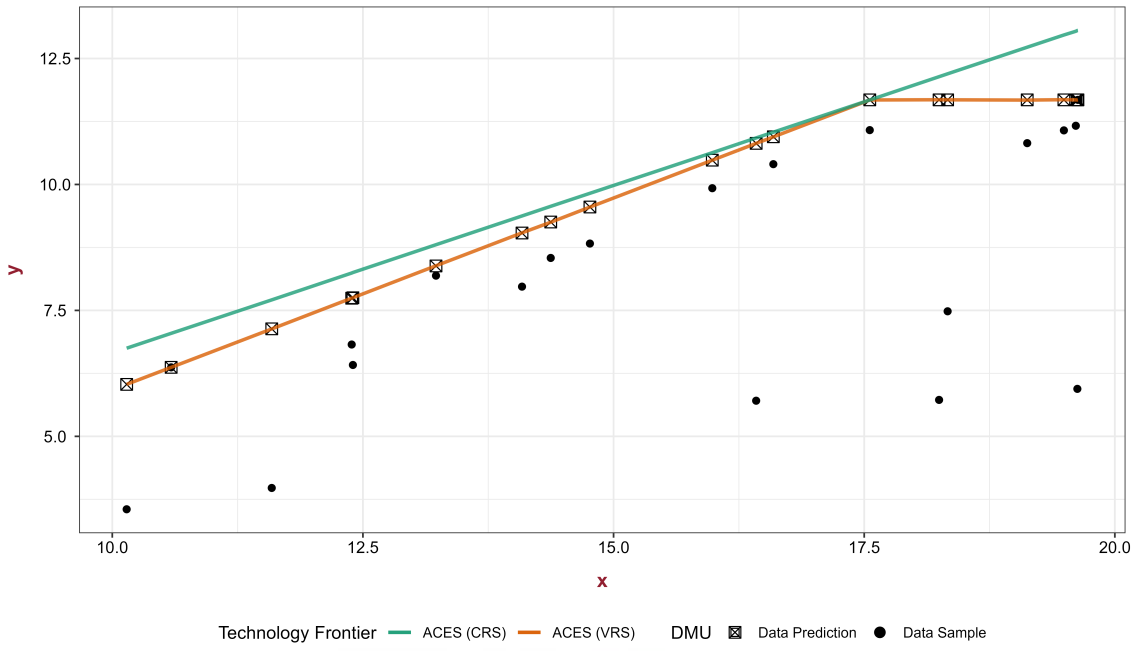

Figure 7.3: Comparison of technology frontiers using Adaptive Constrained Enveloping Splines.

7.3 Computational experiments

This section presents a series of computational experiments designed to assess the performance of the ACES and Q-ACES models. The experiments are organized into two groups: the first replicates scenarios proposed by Nataraja and Johnson (2011), while the second comprises custom-designed scenarios developed specifically for this study. Detailed descriptions of both experimental designs are provided in Section 3.4.1 (Experimental Design 4).

The simulations involve estimating the radial output score, defined as $\phi = \frac{y^*}{y}$ for the single-output production scenario. To achieve this, we use the available data to estimate ϕ through different methodologies, denoted as $\hat{\phi}^{\text{DEA}}$, $\hat{\phi}^{\text{ACES}}$, and $\hat{\phi}^{\text{Q-ACES}}$. The $\hat{\phi}^{\text{DEA}}$ score is computed following the standard procedure described by Banker et al. (1984), using only the relevant input variables. This ensures that surpassing DEA in terms of performance implies simultaneously outperforming other techniques that rely on DEA scores as benchmarks. The $\hat{\phi}^{\text{ACES}}$ and $\hat{\phi}^{\text{Q-ACES}}$ scores are derived using the methodologies outlined in Chapter 5 and 6, respectively. While both approaches account for variable interaction, only $\hat{\phi}^{\text{Q-ACES}}$ incorporates additional procedures aimed at accelerating the algorithm. Nonetheless, the strategy for optimizing input selection for efficiency estimation can also be applied to the standard ACES framework without introducing any acceleration techniques.

To evaluate the relative performance of the methods, 100 trials were conducted for each proposed scenario. Performance was assessed using two standard metrics: mean squared error (MSE) and linear correlation (ρ), as described in Nataraja and Johnson (2011). The MSE quantifies the average squared difference between the estimated scores ($\hat{\phi}$) and the true scores (ϕ) for a given DMU, offering a precise measure of estimation accuracy. In contrast, linear correlation measures the strength and direction of the relationship between the estimated and true scores, with higher values indicating a closer alignment between the estimated and theoretical results.

It is worth noting that all methods occasionally produce infeasible or highly atypical results that distort the true performance of the models. To address this, the average MSE, linear correlation and computation time per scenario were calculated exclusively using results whose MSE values fall within 20 times the interquartile range (IQR) of the data. Additionally, the number of excluded results is recorded for each case. For any variant of ACES, this issue could be resolved by fine-tuning the hyperparameters for the specific problem using k-fold cross-validation or a training and testing split. However, due to the high computational cost of the proposed simulations, a fixed set of hyperparameters has been established, as described in the following paragraph.

Our algorithms provide a broad range of hyperparameters that allow for substantial adaptability to the specific characteristics of the data (see Section 5.6.2 for details on the hyperparameter tuning procedure). In this study, the number of terms (η) was fixed at 50, and the minimum improvement threshold in the training error (ξ) was set at 0.5%. The `minspan` and `endspan` parameters were selected based on the guidelines proposed by Friedman (1991). Additionally, the complexity penalization hyperparameter (d) was set to 1, and the maximum interaction degree ($q_{\max}$) was limited to 2. To prevent high-degree BFs from deteriorating model performance, $\xi^{(2)}$ was set to 0.05, meaning that a second-degree BF must yield at least a 5% greater error reduction than the best first-degree term to be retained in the model. Finally, model selection was guided using the MSE as the LOF criterion.

7.3.1 Results from Experimental Design 4: Nataraja and Johnson (2011)

Table 7.1 presents four key metrics for each method: MSE, the linear correlation between theoretical and estimated efficiency scores (ρ), computation time in seconds (computation time is always below one second for DEA) and the number of excluded results due to infeasibility or atypicality based on a threshold of 20 times the interquartile range (denoted as #out). Table 7.2 complements these results by showing the percentage changes in MSE and linear correlation compared to a baseline model. Additionally, computation time is specifically compared between ACES and Q-ACES. In these comparisons, ACES serves as the baseline for the ACES vs. Q-ACES comparison, while DEA is used as the baseline for the DEA vs. Q-ACES.

The results presented in Table 7.2 highlight the conditions under which Q-ACES outperforms ACES. Regarding computation time, the proposed strategies for reducing knots in Q-ACES are highly effective, as the average execution time is halved compared to ACES. This is particularly notable in Scenario 10 (300 DMUs), where the execution time decreases by 75.5%. In terms of accuracy, Q-ACES outperforms ACES in 10 out of the 14 scenarios analyzed. This improvement is primarily due to (i) the focused search for knots in specific regions of the input space and, more significantly, (ii) the consideration of only those inputs that generate at least one BF for efficiency estimation. In these scenarios, the average MSE reduction is 20%. Conversely, ACES performs better in four scenarios: Scenario 3, where relevant inputs are highly correlated; Scenario 7, which present a high correlation between the irrelevant input and one of the relevant inputs, Scenario 9, characterized by a small sample size (25 DMUs); and Scenario 11, where decreasing returns to scale are present. When compared to DEA, the computational time for DEA remains significantly lower, as it requires solving only a single LP model. However, Q-ACES demonstrates substantial improvements in accuracy, with reductions in MSE observed across all scenarios. On average, MSE is reduced by 63.1%, yielding far superior results in terms of precision. Regarding linear correlation, all three methodologies deliver similar results, with values close to 1, indicating a strong relationship with the theoretical efficiency scores.

Table 7.1: Performance metrics of evaluated methods in scenarios from Nataraja and Johnson (2011).

set	ACES				Q-ACES				DEA		
	MSE	ρ	Time (s)	#out	MSE	ρ	Time (s)	#out	MSE	ρ	#out
1	0.003	0.99	44	0	0.002	0.99	24	0	0.008	0.99	0
2	0.002	0.99	59	0	0.002	0.99	29	0	0.005	0.99	0
3	0.002	0.99	54	0	0.003	0.99	29	0	0.004	1.00	0
4	0.003	0.99	69	0	0.002	0.99	34	0	0.007	0.99	0
5	0.003	0.99	66	0	0.002	0.99	32	0	0.007	0.99	0
6	0.003	0.99	68	0	0.002	0.99	32	0	0.007	0.99	0
7	0.002	1.00	62	0	0.002	0.99	43	0	0.007	0.99	0
8	0.002	0.99	60	0	0.002	0.99	42	0	0.007	0.99	0
9	0.020	0.96	13	0	0.024	0.94	7	0	0.030	0.97	0
10	0.001	1.00	515	0	0.000	1.00	126	0	0.002	1.00	0
11	0.008	0.98	63	0	0.012	0.97	33	0	0.043	0.90	0
12	0.005	0.99	87	0	0.004	0.99	44	0	0.014	0.99	0
13	0.001	1.00	28	1	0.001	1.00	13	1	0.002	1.00	0
14	0.000	1.00	60	0	0.000	1.00	31	0	0.002	0.99	0

Table 7.2: Relative performance of evaluated methods in scenarios from Nataraja and Johnson (2011).

set	ACES vs Q-ACES			DEA vs Q-ACES	
	MSE	ρ	Time (seconds)	MSE	ρ
1	-22.7%	+0.2%	-45.7%	-70.5%	+0.5%
2	-7.1%	-0.0%	-49.9%	-56.6%	+0.0%
3	+38.1%	-0.1%	-46.1%	-20.1%	-0.3%
4	-22.5%	+0.1%	-51.2%	-67.0%	+0.3%
5	-28.5%	+0.1%	-51.0%	-69.8%	+0.3%
6	-17.6%	+0.1%	-52.2%	-66.5%	+0.3%
7	+4.3%	-0.0%	-30.9%	-69.9%	+0.4%
8	-16.3%	+0.1%	-30.3%	-72.4%	+0.4%
9	+19.1%	-1.3%	-45.0%	-20.2%	-2.7%
10	-20.4%	+0.1%	-75.5%	-82.7%	+0.3%
11	+57.4%	-1.5%	-47.7%	-71.6%	+7.4%
12	-14.5%	+0.1%	-50.0%	-70.7%	+0.4%
13	-13.8%	+0.1%	-52.9%	-52.5%	+0.1%
14	-37.6%	+0.1%	-47.6%	-90.4%	+0.6%

7.3.2 Results from Experimental Design 4: custom-designed scenarios

The results of the experimental simulations are presented in Tables 7.7 and 7.8, which report the same four key performance metrics as in the previous experimental design: MSE, linear correlation (ρ), computation time (in seconds), and the number of excluded results (#out). These results are complemented by Tables 7.9 and 7.10, which highlight the relative differences in performance metrics across the evaluated methods. The relative differences focus on percentage changes in MSE and linear correlation compared to a baseline model, with computation time comparisons specifically drawn between ACES and Q-ACES. For the ACES vs. Q-ACES comparison, ACES is used as the baseline, while DEA serves as the baseline for the DEA vs. Q-ACES comparison. Tables 7.7 and 7.9 emphasize scenarios characterized by high correlation between irrelevant and relevant inputs, while Tables 7.8 and 7.10 address scenarios with low correlation between these inputs. Both sets of tables further capture varying levels of correlation among relevant inputs, including both high and low correlation cases.

In terms of accuracy, ACES consistently outperforms Q-ACES, achieving the lowest MSE values across the majority of scenarios. However, performance differences can be observed depending on the number of variables, sample sizes, and the correlations present between them, as highlighted in Tables 7.3 and 7.4.

Table 7.3: Impact of correlations and dimensionality on the relative MSE of Q-ACES vs ACES.

Number of variables	[High–High]	[Low–High]	[High–Low]	[Low–Low]
6	+49.5%	+39.7%	+19.4%	+22.9%
10	+29.2%	+25.4%	+13.6%	+1.9%
12	+22.6%	+14.9%	-5.4%	+12.7%
20	-35.4%	-40.7%	-13.2%	-19.2%

Table 7.4: Impact of sample size and dimensionality on the relative MSE of Q-ACES vs ACES.

Number of variables	50	150	300
6	+19.94%	+47.69%	+31.02%
10	+5.45%	+35.76%	+11.36%
12	+12.49%	+30.10%	-0.95%
20	+2.70%	-12.80%	-71.31%

The results in Table 7.3 show that ACES outperforms Q-ACES in most scenarios, particularly when the number of variables is small, and the correlation between relevant and irrelevant variables is high. The scenario with six variables and high correlations across all variables exhibits the greatest difference between ACES and Q-ACES. From that point onwards, as the number of variables increases and correlations weaken, the performance gap narrows, ultimately reversing in high-dimensional settings where Q-ACES begins to outperform ACES. On the other hand, Table 7.4 provides additional insights into the influence of sample size and dimensionality on the relative performance of Q-ACES compared to ACES. When the number of variables is below 10, ACES consistently outperforms Q-ACES across all sample sizes. However, as the sample size increases, Q-ACES shows progressive improvements relative to ACES, particularly when larger datasets are considered. Furthermore, as the number of variables grows, Q-ACES becomes increasingly competitive, eventually outperforming ACES in scenarios with higher dimensionality, regardless of the sample size.

In the comparison with DEA, the analysis focuses solely on the effect of the number of relevant variables and the sample size, since DEA computations are performed exclusively using the relevant variables. Results are displayed in Table 7.5.

Table 7.5: Impact of sample size and dimensionality on the relative MSE of Q-ACES vs DEA.

Number of relevant variables	50	150	300
4	-53.30%	-75.30%	-85.81%
8	-63.74%	-84.61%	-93.23%

Table 7.5 demonstrates that Q-ACES (and ACES) consistently offers a more accurate alternative to DEA for measuring technical efficiency. The results also reveal that as the sample size and the number of (relevant) variables increase, the relative advantage of Q-ACES over DEA becomes more pronounced.

Regarding linear correlation, the results are remarkably similar between ACES and Q-ACES, with both methods consistently achieving high values. Specifically, ACES exceeds a correlation of 0.9 in 37 out of 48 scenarios, while Q-ACES does so in 38 out of 48 cases, with minimum correlations for both methods remaining close to 0.8. In contrast, DEA performs significantly worse, surpassing the 0.9 threshold in only 12 scenarios, despite relying solely on the relevant variables. These results, detailed in Tables 7.7 and 7.8, highlight the robustness of our methods. Moreover, the consistently high correlation with the theoretical values suggests that, even if the methods were not perfectly precise in estimating technical efficiency (which they are), they would still be highly effective at ranking the units correctly. This implies that the methodologies can reliably distinguish better-performing units, even when exact efficiency scores may be slightly shifted.

In terms of the occurrence of highly atypical results (see Tables 7.7 and 7.8), Q-ACES demonstrates a clear advantage over both ACES and DEA. The number of scenarios with such extreme outcomes is significantly lower for Q-ACES compared to the other methods. This difference becomes particularly evident in high-dimensional settings with 20 variables and sample sizes of 150 and 300 DMUs. In these scenarios, ACES produces 159 cases with highly atypical results, while DEA exhibits 59 such instances. In contrast, Q-ACES remarkably reports only 2 scenarios with highly atypical results. Nevertheless, it is worth noting that for both Q-ACES and ACES, this issue can be effectively addressed by adjusting the available hyperparameters. Nonetheless, for both Q-ACES and ACES, this issue can be mitigated by tuning the available hyperparameters, allowing better adaptation to the dataset's characteristics.

Finally, from a computational standpoint, DEA remains unrivaled due to its simplicity, as it requires solving only a single LP model. However, the computational efficiency of Q-ACES compared to ACES has been significantly improved, demonstrating notable reductions in execution time while maintaining competitive accuracy. Results are shown in Table 7.6.

Table 7.6: Impact of sample size and dimensionality on the computation time of Q-ACES vs ACES.

Number of variables	50	150	300
6	-47.07%	-60.06%	-72.46%
10	-55.59%	-63.95%	-75.11%
12	-67.57%	-70.97%	-80.93%
20	-79.09%	-80.31%	-85.55%

As shown by the results, Q-ACES achieves a consistent and substantial reduction in computation time compared to ACES, with the percentage reduction becoming increasingly pronounced as both sample size and dimensionality increase. This efficiency gain is particularly evident in high-dimensional scenarios with a large number of DMUs. For instance, in cases involving 20 variables and 300 DMUs, the execution time for ACES in one single problem can easily exceed 10 hours, whereas Q-ACES typically completes the task in under 2 hours, demonstrating its practicality and computational advantage in complex settings.

In conclusion, both methodologies, ACES and Q-ACES, significantly outperform DEA in terms of precision and linear correlation, even when DEA acts as an oracle by having prior knowledge of the relevant variables. When comparing ACES and Q-ACES specifically, the latter emerges as an appealing alternative for problems involving a large number of DMUs or wide-data scenarios, where the number of variables exceeds the number of observations. Moreover, Q-ACES demonstrates remarkable computational efficiency, especially in high-dimensional settings, reducing execution times substantially compared to ACES while maintaining comparable or even superior performance in scenarios with large sample sizes or complex datasets.

Table 7.7: Performance metrics for scenarios with high correlation between irrelevant and relevant inputs.

set	# relevant	# irrelevant	Sample size	ρ (rel)	ρ (rel-irr)	**ACES**				**Q-ACES**				**DEA**		
						MSE	ρ	Time	#out	MSE	ρ	Time	#out	MSE	ρ	#out
1	4	2	50	High	High	0.023	0.95	50	0	0.028	0.93	37	0	0.060	0.88	0
2	4	2	150	High	High	0.005	0.99	180	0	0.008	0.98	85	0	0.034	0.91	1
3	4	2	300	High	High	0.002	0.99	692	0	0.004	0.99	211	0	0.024	0.95	0
4	4	6	50	High	High	0.034	0.92	205	0	0.040	0.90	127	0	0.075	0.84	0
5	4	6	150	High	High	0.007	0.98	572	2	0.011	0.97	274	0	0.037	0.90	1
6	4	6	300	High	High	0.004	0.99	2279	1	0.004	0.99	704	0	0.023	0.88	6
7	8	4	50	High	High	0.040	0.90	280	0	0.050	0.86	113	0	0.158	0.64	1
8	8	4	150	High	High	0.012	0.96	1718	2	0.015	0.96	578	0	0.113	0.72	4
9	8	4	300	High	High	0.006	0.96	6893	11	0.007	0.98	1730	0	0.080	0.79	3
10	8	12	50	High	High	0.070	0.83	2350	0	0.073	0.81	704	0	0.170	0.62	2
11	8	12	150	High	High	0.025	0.92	9131	21	0.020	0.95	2543	0	0.103	0.77	2
12	8	12	300	High	High	0.069	0.92	34243	31	0.007	0.98	6918	0	0.079	0.76	8
1	4	2	50	Low	High	0.025	0.95	56	0	0.032	0.92	33	0	0.096	0.80	1
2	4	2	150	Low	High	0.006	0.99	190	0	0.010	0.96	81	1	0.048	0.89	1
3	4	2	300	Low	High	0.003	0.99	739	0	0.004	0.99	228	0	0.030	0.93	2
4	4	6	50	Low	High	0.035	0.91	233	1	0.037	0.91	116	0	0.078	0.83	0
5	4	6	150	Low	High	0.008	0.98	637	0	0.013	0.97	272	0	0.051	0.89	2
6	4	6	300	Low	High	0.003	0.99	2445	1	0.004	0.99	747	0	0.030	0.93	3
7	8	4	50	Low	High	0.056	0.87	317	0	0.060	0.84	100	1	0.214	0.54	1
8	8	4	150	Low	High	0.014	0.96	1939	2	0.018	0.95	610	0	0.148	0.71	0
9	8	4	300	Low	High	0.007	0.97	11324	9	0.007	0.98	2024	0	0.113	0.75	3
10	8	12	50	Low	High	0.068	0.84	1583	1	0.075	0.80	349	0	0.204	0.57	0
11	8	12	150	Low	High	0.047	0.85	13466	20	0.024	0.93	2706	1	0.154	0.67	1
12	8	12	300	Low	High	0.048	0.91	46027	28	0.008	0.97	7023	1	0.114	0.75	7

Table 7.8: Performance metrics for scenarios with low correlation between irrelevant and relevant inputs.

set	# relevant	# irrelevant	Sample size	ρ (rel)	ρ (rel-irr)	**ACES**				**Q-ACES**				**DEA**		
						MSE	ρ	Time	#out	MSE	ρ	Time	#out	MSE	ρ	#out
1	4	2	50	High	Low	0.025	0.94	58	0	0.030	0.92	30	0	0.069	0.85	0
2	4	2	150	High	Low	0.006	0.99	193	0	0.008	0.98	70	0	0.035	0.90	1
3	4	2	300	High	Low	0.003	0.98	726	1	0.003	0.99	176	1	0.023	0.93	2
4	4	6	50	High	Low	0.045	0.88	267	0	0.047	0.88	96	0	0.075	0.82	1
5	4	6	150	High	Low	0.008	0.98	660	0	0.011	0.97	191	1	0.033	0.92	2
6	4	6	300	High	Low	0.003	0.98	2532	1	0.004	0.99	455	0	0.023	0.94	5
7	8	4	50	High	Low	0.053	0.85	303	2	0.056	0.85	98	0	0.164	0.63	3
8	8	4	150	High	Low	0.013	0.97	1838	0	0.016	0.96	490	0	0.114	0.72	5
9	8	4	300	High	Low	0.006	0.93	8834	6	0.005	0.98	1489	0	0.081	0.81	9
10	8	12	50	High	Low	0.081	0.81	3006	0	0.078	0.79	531	0	0.174	0.62	1
11	8	12	150	High	Low	0.020	0.86	14161	11	0.021	0.94	2104	0	0.110	0.74	11
12	8	12	300	High	Low	0.009	0.81	36616	19	0.005	0.98	4179	0	0.076	0.83	21
1	4	2	50	Low	Low	0.032	0.92	60	0	0.035	0.90	31	0	0.090	0.80	1
2	4	2	150	Low	Low	0.008	0.98	210	0	0.012	0.97	70	0	0.054	0.88	1
3	4	2	300	Low	Low	0.003	0.99	734	0	0.003	0.99	180	0	0.030	0.93	2
4	4	6	50	Low	Low	0.048	0.88	285	0	0.046	0.88	85	0	0.094	0.80	0
5	4	6	150	Low	Low	0.009	0.98	687	0	0.011	0.97	170	0	0.048	0.89	0
6	4	6	300	Low	Low	0.004	0.99	2403	0	0.004	0.99	484	0	0.030	0.93	1
7	8	4	50	Low	Low	0.057	0.87	330	0	0.064	0.83	84	0	0.209	0.56	0
8	8	4	150	Low	Low	0.014	0.97	1928	0	0.019	0.94	470	1	0.145	0.69	3
9	8	4	300	Low	Low	0.006	0.97	11558	3	0.006	0.98	1903	0	0.117	0.75	6
10	8	12	50	Low	Low	0.084	0.78	1998	1	0.084	0.77	279	0	0.207	0.53	0
11	8	12	150	Low	Low	0.020	0.91	13248	10	0.023	0.94	2111	0	0.145	0.69	3
12	8	12	300	Low	Low	0.019	0.89	48601	19	0.006	0.98	5312	0	0.107	0.78	6

Table 7.9: Q-ACES compared to ACES and DEA for scenarios with high correlation between irrelevant and relevant inputs.

set	# relevant	# irrelevant	Sample size	ρ (rel)	ρ (rel-irr)	**ACES vs Q-ACES**			**DEA vs Q-ACES**	
						MSE	ρ	Time	MSE	ρ
1	4	2	50	High	High	+23.6%	-1.7%	-26.3%	-52.9%	+6.5%
2	4	2	150	High	High	+60.5%	-1.1%	-52.9%	-76.3%	+6.1%
3	4	2	300	High	High	+64.4%	-0.6%	-69.5%	-82.8%	+4.4%
4	4	6	50	High	High	+19.1%	-2.8%	-38.3%	-46.5%	+6.4%
5	4	6	150	High	High	+45.0%	-1.2%	-52.2%	-71.5%	+6.2%
6	4	6	300	High	High	+23.5%	-0.3%	-69.1%	-81.5%	+4.8%
7	8	4	50	High	High	+24.4%	-4.2%	-59.7%	-68.3%	+32.0%
8	8	4	150	High	High	+32.5%	-0.7%	-66.4%	-86.4%	+26.9%
9	8	4	300	High	High	+10.9%	+2.5%	-74.9%	-91.8%	+19.2%
10	8	12	50	High	High	+3.8%	-3.1%	-70.0%	-57.0%	+27.7%
11	8	12	150	High	High	-20.4%	+2.7%	-72.1%	-80.4%	+20.6%
12	8	12	300	High	High	-89.6%	+6.1%	-79.8%	-91.0%	+17.9%
1	4	2	50	Low	High	+28.3%	-3.0%	-42.1%	-66.6%	+15.2%
2	4	2	150	Low	High	+58.6%	-1.3%	-57.1%	-79.2%	+8.7%
3	4	2	300	Low	High	+32.3%	-0.4%	-69.2%	-86.5%	+6.2%
4	4	6	50	Low	High	+4.2%	-0.6%	-50.0%	-52.9%	+8.9%
5	4	6	150	Low	High	+53.4%	-1.4%	-57.3%	-75.6%	+8.9%
6	4	6	300	Low	High	+18.6%	-0.3%	-69.5%	-87.3%	+6.5%
7	8	4	50	Low	High	+7.3%	-3.0%	-68.4%	-72.0%	+55.4%
8	8	4	150	Low	High	+30.2%	-1.2%	-68.6%	-87.6%	+34.4%
9	8	4	300	Low	High	+7.1%	+1.3%	-82.1%	-93.8%	+30.4%
10	8	12	50	Low	High	+10.2%	-4.7%	-77.9%	-63.4%	+40.1%
11	8	12	150	Low	High	-48.8%	+8.8%	-79.9%	-84.3%	+38.5%
12	8	12	300	Low	High	-83.4%	+7.6%	-84.7%	-93.0%	+29.4%

Table 7.10: Q-ACES compared to ACES and DEA for scenarios with low correlation between irrelevant and relevant inputs.

set	# relevant	# irrelevant	Sample size	ρ (rel)	ρ (rel-irr)	**ACES vs Q-ACES**			**DEA vs Q-ACES**	
						MSE	ρ	Time	MSE	ρ
1	4	2	50	High	Low	+18.1%	-2.1%	-47.9%	-56.6%	+7.9%
2	4	2	150	High	Low	+26.9%	-0.7%	-63.5%	-77.4%	+6.2%
3	4	2	300	High	Low	+13.3%	-0.3%	-75.7%	-87.2%	+4.6%
4	4	6	50	High	Low	+3.2%	-0.9%	-63.8%	-38.0%	+5.1%
5	4	6	150	High	Low	+24.4%	-1.0%	-71.1%	-67.5%	+4.7%
6	4	6	300	High	Low	+13.2%	+0.5%	-82.0%	-83.1%	+4.4%
7	8	4	50	High	Low	+5.8%	-2.7%	-67.5%	-65.6%	+31.9%
8	8	4	150	High	Low	+24.8%	-1.4%	-73.3%	-86.3%	+27.0%
9	8	4	300	High	Low	-14.4%	+1.5%	-83.1%	-93.6%	+18.9%
10	8	12	50	High	Low	-3.1%	-2.4%	-82.3%	-54.8%	+25.0%
11	8	12	150	High	Low	+7.1%	+1.4%	-85.1%	-80.7%	+22.6%
12	8	12	300	High	Low	-43.7%	+6.2%	-88.6%	-93.3%	+18.1%
1	4	2	50	Low	Low	+9.7%	-1.9%	-48.1%	-61.2%	+12.9%
2	4	2	150	Low	Low	+44.7%	-1.2%	-66.8%	-78.1%	+9.7%
3	4	2	300	Low	Low	+14.1%	-0.2%	-75.4%	-90.1%	+6.5%
4	4	6	50	Low	Low	-4.7%	+0.0%	-70.2%	-51.6%	+10.7%
5	4	6	150	Low	Low	+20.2%	-0.9%	-75.2%	-76.7%	+8.4%
6	4	6	300	Low	Low	-9.9%	-0.1%	-79.9%	-87.9%	+6.4%
7	8	4	50	Low	Low	+12.5%	-4.3%	-74.6%	-69.5%	+47.2%
8	8	4	150	Low	Low	+32.9%	-2.1%	-75.6%	-86.8%	+37.7%
9	8	4	300	Low	Low	-7.4%	+1.2%	-83.5%	-95.1%	+31.8%
10	8	12	50	Low	Low	-0.1%	-0.0%	-86.0%	-59.4%	+45.4%
11	8	12	150	Low	Low	+11.0%	+3.3%	-84.1%	-84.4%	+35.3%
12	8	12	300	Low	Low	-68.6%	+10.8%	-89.1%	-94.2%	+26.2%

7.4 Discussion and final remarks

This chapter introduces Q-ACES, a robust and efficient methodology designed to overcome the well-documented limitations of traditional DEA, such as the curse of dimensionality, overfitting, and the inclusion of irrelevant variables. Q-ACES combines correlation-based input filtering, adaptive candidate reduction based on variable importance, and residual-driven knot prioritization to guide the estimation process efficiently and selectively. These strategies are embedded within a shape-constrained regression framework, enabling Q-ACES to construct parsimonious yet flexible models that maintain interpretability while improving performance. As a result, Q-ACES offers a principled and scalable alternative to DEA and ACES for estimating technical efficiency in high-dimensional and complex data settings.

Although Q-ACES represents the most computationally efficient proposed in this thesis, it should be viewed as part of a broader methodological family that also includes ACES and RF-ACES. All three methods share a common foundation: they estimate nonparametric production frontiers through shape-constrained regression splines that enforce monotonicity and concavity, preserving the core axioms of production theory. ACES serves as the baseline model, providing a rigorous yet flexible framework via an adaptation of MARS. RF-ACES builds upon this foundation by integrating the Random Forest paradigm to reduce variance and enhance the model's generalization capacity. While Q-ACES has been explicitly developed as an extension of ACES, its key strategies can also be naturally incorporated into the RF-ACES framework. This underlines the conceptual coherence and modular adaptability of the overall family of methods.

A further contribution of this chapter is the introduction of a variable selection strategy that restricts the estimation of technical efficiency to inputs that are empirically relevant for explaining output levels. While this strategy is embedded natively into Q-ACES, it can also be applied to the standard ACES framework. However, in that case, the model fitting procedure does not explicitly prioritize the use of relevant variables during spline expansion, as Q-ACES does by design. In parallel, a similar strategy could be developed within the RF-ACES framework by using the RF structure and computing input relevance based on out-of-bag error contributions. Such an approach would allow for relevance-aware frontier estimation while preserving the ensemble's robustness and generalization properties, potentially yielding more stable results in the presence of noise or redundant predictors. Moreover, the reduction in execution time achieved by Q-ACES makes it a valuable tool for efficiently discarding poorly performing configurations during the hyperparameter tuning process.

Chapter 8

Conclusions

This thesis has introduced and thoroughly analyzed a family of novel methodologies for estimating production frontiers, specifically designed to address significant limitations associated with traditional Data Envelopment Analysis (DEA). DEA, despite its longstanding popularity in efficiency analysis due to its intuitive interpretation and flexibility, often struggles with issues such as overfitting, sensitivity to dimensionality, and the absence of systematic approaches for statistical inference and variable selection. The methodologies presented here—Adaptive Constrained Enveloping Splines (ACES), Random Forest-ACES (RF-ACES), and Quick-ACES (Q-ACES)—offer important progress toward overcoming these challenges.

At the core of the proposed framework, ACES integrates the Multivariate Adaptive Regression Splines (MARS) approach with explicit shape constraints—monotonicity and concavity—that align directly with established economic theory. This combination ensures theoretical consistency while maintaining sufficient flexibility to accurately model complex relationships within data. Unlike traditional DEA, the ACES method incorporates a structured pruning procedure based on generalized cross-validation, substantially reducing overfitting and improving the generalizability of the resulting frontier beyond the sample used for estimation.

Building on the foundation provided by ACES, the RF-ACES methodology incorporates ensemble learning concepts drawn from Random Forests. By aggregating multiple ACES estimators trained on bootstrap samples, RF-ACES achieves greater robustness and reduces sensitivity to local irregularities in the data. A key advantage of this method is its internal variable importance metric, which offers a data-driven approach for selecting relevant variables. However, this gain in robustness is accompanied by increased computational complexity, presenting a clear trade-off in practical applications.

To address the computational challenges identified in RF-ACES, Q-ACES introduces heuristic strategies aimed explicitly at reducing computational time and improving scalability for large datasets. These strategies include correlation-based input preselection, DEA-guided neighborhood knot reduction, and adaptive filtering of candidate basis functions. The result is a streamlined and computationally efficient approach that maintains high accuracy while adhering to the theoretical principles established by ACES.

Empirical validations conducted throughout multiple simulation experiments confirm the superiority of the ACES family relative to well-established methods such as DEA, the Corrected Concave Nonparametric Least Squares estimator, the Stochastic Non-Smooth Envelopment of Data approach, Bootstrap DEA, and the Convexified Efficiency Analysis Trees. The ACES methods consistently provided more accurate and stable efficiency estimates, especially in scenarios characterized by high dimensionality and the presence of irrelevant inputs, contexts where traditional methodologies often underperform significantly.

Additionally, from a practical perspective, this thesis has provided detailed guidance on how to configure and tune the proposed methods according to the characteristics of each dataset. These recommendations help analysts make well-informed decisions, taking into account factors such as dimensionality, sample size, and computational resources. By offering clear criteria for model selection and parameter adjustment, the thesis bridges the gap between methodological development and empirical application. In this way, the proposed methods not only strengthen theoretical foundations but also respond to real-world challenges commonly faced in efficiency analysis.

In summary, the methodologies introduced—ACES, RF-ACES, and Q-ACES—represent meaningful contributions to the field of technical efficiency analysis. They successfully address longstanding econometric and practical limitations, providing analysts and decision-makers with sophisticated yet interpretable tools. Consequently, these methodologies significantly increase the reliability and validity of efficiency evaluations across diverse application areas.

Future research can further extend the methodologies proposed in this thesis. First, incorporating explicit treatments of stochastic noise and formal statistical inference into the ACES family remains an important area of development. Second, investigating additional ensemble strategies or hybridizing ACES with other machine learning techniques could lead to further improvements in robustness and predictive accuracy. Third, applying these methodologies to richer, real-world datasets across various sectors will be essential for continuing to validate and refine their practical usefulness. Lastly, formalizing all the developed code into a well-documented, user-friendly software package would greatly facilitate adoption by other researchers and practitioners, promoting broader use and reproducibility of the proposed techniques.

Conclusiones

Esta tesis ha presentado y analizado en profundidad una familia de metodologías novedosas para la estimación de fronteras de producción, diseñadas específicamente para abordar limitaciones importantes de la técnica del Análisis Envolvente de Datos (DEA, por sus siglas en inglés). A pesar de su amplia aceptación en el análisis de eficiencia por su interpretación intuitiva y flexibilidad, DEA presenta debilidades bien conocidas, como el sobreajuste, la sensibilidad a la dimensionalidad y la falta de procedimientos sistemáticos para realizar inferencia estadística y selección de variables. Las metodologías propuestas aquí —Adaptive Constrained Enveloping Splines (ACES), Random Forest-ACES (RF-ACES) y Quick-ACES (Q-ACES)— suponen un avance importante en la superación de estos problemas.

En el núcleo del enfoque propuesto, ACES combina el modelo de Multivariate Adaptive Regression Splines (MARS) con restricciones de forma explícitas —monotonía y concavidad— inspiradas directamente en la teoría económica. Esta combinación garantiza consistencia teórica y, al mismo tiempo, mantiene la flexibilidad necesaria para modelar con precisión relaciones complejas en los datos. A diferencia del DEA tradicional, ACES incorpora un procedimiento estructurado de poda basado en validación cruzada generalizada, lo que permite reducir significativamente el sobreajuste y mejorar la capacidad del modelo para generalizar más allá de la muestra observada.

Sobre esta base, la metodología RF-ACES introduce principios de modelos agregados inspirados en los Bosques Aleatorios. Al combinar múltiples estimadores ACES entrenados sobre muestras bootstrap, RF-ACES incrementa la robustez del modelo y reduce su sensibilidad a irregularidades locales en los datos. Una de sus principales ventajas es la métrica interna de importancia de variables, que permite una selección sistemática y guiada por los datos, especialmente útil en contextos de alta dimensionalidad. No obstante, este aumento en robustez implica un coste computacional más elevado, lo que plantea un claro compromiso en aplicaciones prácticas.

Con el fin de resolver los desafíos computacionales asociados a RF-ACES, la versión Q-ACES introduce estrategias heurísticas pensadas específicamente para reducir los tiempos de ejecución y mejorar la escalabilidad en conjuntos de datos grandes. Estas estrategias incluyen una preselección de variables basada en correlaciones, una reducción de nodos guiada por DEA y un filtrado adaptativo de funciones base candidatas. El resultado es una versión más eficiente computacionalmente que conserva una alta precisión sin renunciar a los principios teóricos de ACES.

Las validaciones empíricas realizadas a lo largo de múltiples experimentos de simulación confirman la superioridad de la familia ACES frente a métodos ampliamente establecidos como DEA, Corrected Concave Nonparametric Least Squares, Stochastic Non-Smooth Envelopment of Data, Bootstrap DEA, y Convexified Efficiency Analysis Trees. Las metodologías ACES proporcionan estimaciones de eficiencia más precisas y estables, especialmente en escenarios con alta dimensionalidad o presencia de variables irrelevantes, contextos en los que los métodos tradicionales suelen mostrar un rendimiento significativamente inferior.

Desde una perspectiva aplicada, esta tesis también ha ofrecido recomendaciones detalladas sobre cómo configurar y ajustar los métodos propuestos en función de las características de cada conjunto de datos. Estas orientaciones permiten tomar decisiones informadas, teniendo en cuenta aspectos como la dimensionalidad, el tamaño muestral o las limitaciones computacionales. Al ofrecer criterios claros para la selección de modelos y parámetros, se estrecha la distancia entre el desarrollo metodológico y su aplicación empírica. Así, las propuestas no solo refuerzan el rigor teórico, sino que también aportan soluciones prácticas a los retos habituales del análisis de eficiencia.

En resumen, las metodologías introducidas —ACES, RF-ACES y Q-ACES— constituyen contribuciones relevantes al estudio de la eficiencia técnica. Abordan limitaciones econométricas y prácticas en grandes conjuntos de datos, y proporcionan herramientas sofisticadas pero interpretables tanto para analistas como para responsables de toma de decisiones. En consecuencia, permiten obtener evaluaciones de eficiencia más fiables y válidas en una amplia variedad de aplicaciones.

De cara al futuro, esta tesis abre varias líneas prometedoras de investigación. En primer lugar, incorporar de forma explícita el tratamiento del ruido estocástico y desarrollar procedimientos de inferencia estadística constituye un paso natural en la evolución de la familia ACES. En segundo lugar, explorar nuevas estrategias de modelos agregados o combinar ACES con otras técnicas de aprendizaje automático podría mejorar aún más la robustez y la capacidad predictiva de los modelos. En tercer lugar, aplicar estas metodologías a conjuntos de datos reales y variados será fundamental para seguir validando y perfeccionando su utilidad práctica. Por último, formalizar todo el código desarrollado en un paquete de software bien documentado y fácil de usar facilitará su adopción por parte de la comunidad investigadora y profesional, promoviendo un uso más amplio y reproducible de las técnicas propuestas.

Bibliography

Adler, N. and Golany, B. (2001). Evaluation of deregulated airline networks using data envelopment analysis combined with principal component analysis with an application to western europe. *European Journal of Operational Research*, 132(2):260–273.

Adler, N. and Golany, B. (2002). Including principal component weights to improve discrimination in data envelopment analysis. *Journal of the Operational Research Society*, 53(9):985–991.

Afriat, S. N. (1972). Efficiency estimation of production functions. *International economic review*, pages 568–598.

Aigner, D., Lovell, C. K., and Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. *Journal of econometrics*, 6(1):21–37.

Aigner, D. J. and Chu, S.-f. (1968). On estimating the industry production function. *The American economic review*, 58(4):826–839.

Antunes, J., Hadi-Vencheh, A., Jamshidi, A., Tan, Y., and Wanke, P. (2022). Bank efficiency estimation in china: Dea-renna approach. *Annals of Operations Research*, 315(2):1373–1398.

Aparicio, J., Cordero, J. M., Gonzalez, M., and Lopez-Espin, J. J. (2018). Using non-radial dea to assess school efficiency in a cross-country perspective: An empirical analysis of oecd countries. *Omega*, 79:9–20.

Aparicio, J., Cordero, J. M., and Ortiz, L. (2022a). Plausible values and their use in efficiency analyses with educational data. *Applied Economics*, 54(29):3340–3352.

Aparicio, J. and Esteve, M. (2023). How to peel a data envelopment analysis frontier: A cross-validation-based approach. *Journal of the Operational Research Society*, 74(12):2558–2572.

Aparicio, J., Perelman, S., and Santín, D. (2022b). Comparing the evolution of productivity and performance gaps in education systems through dea: an application to latin american countries. *Operational Research*, pages 1–35.

Aragon, Y., Daouia, A., and Thomas-Agnan, C. (2005). Nonparametric frontier estimation: a conditional quantile-based approach. *Econometric Theory*, 21(2):358–389.

Bakin, S., Hegland, M., and Osborne, M. R. (2000). Parallel mars algorithm based on b-splines. *Computational Statistics*, 15(4):463–484.

Balk, B. M. (2001). Scale efficiency and productivity change. *Journal of productivity analysis*, 15:159–183.

Banker, R. D. (1996). Hypothesis tests using data envelopment analysis. *Journal of productivity analysis*, 7:139–159.

Banker, R. D., Charnes, A., and Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management science*, 30(9):1078–1092.

Banker, R. D. et al. (1988). Stochastic data envelopment analysis. *Data envelopment analysis journal*, 5(2):281–309.

Banker, R. D. and Maindiratta, A. (1992). Maximum likelihood estimation of monotone and concave production frontiers. *Journal of Productivity Analysis*, 3(4):401–415.

Barros, C. P. (2005). Evaluating the efficiency of a small hotel chain with a malmquist productivity index. *International Journal of tourism research*, 7(3):173–184.

Battese, G. E. (1991). Frontier production functions and technical efficiency: A survey of empirical applications in agricultural economics.

Bauer, P. W. (1990). Recent developments in the econometric estimation of frontiers. *Journal of econometrics*, 46(1-2):39–56.

Benítez-Peña, S., Bogetoft, P., and Morales, D. R. (2020). Feature selection in data envelopment analysis: A mathematical optimization approach. *Omega*, 96:102068.

Benito, B., Faura, U., Guillamón, M.-D., and Ríos, A.-M. (2019). The efficiency of public services in small municipalities: The case of drinking water supply. *Cities*, 93:95–103.

Berk, R. A. et al. (2008). *Statistical learning from a regression perspective*, volume 14. Springer.

Bogetoft, P., Otto, L., and Otto, M. L. (2015). Package 'benchmarking'. *Retrieved January*, 4:2016.

Boussofiane, A., Dyson, R. G., and Thanassoulis, E. (1991). Applied data envelopment analysis. *European journal of operational research*, 52(1):1–15.

Breiman, L. (1993). Hinging hyperplanes for regression, classification, and function approximation. *IEEE Transactions on Information Theory*, 39(3):999–1013.

Breiman, L. (1996). Bagging predictors. *Machine learning*, 24:123–140.

Breiman, L. (2001). Random forests. *Machine learning*, 45:5–32.

Breiman, L., Friedman, J., Olshen, R. A., and Stone, C. J. (1984). *Classification and regression trees*. Routledge.

Briec, W. (1997). A graph-type extension of farrell technical efficiency measure. *Journal of Productivity Analysis*, 8(1):95–110.

Caves, D. W., Christensen, L. R., and Diewert, W. E. (1982). The economic theory of index numbers and the measurement of input, output, and productivity. *Econometrica: journal of the Econometric Society*, pages 1393–1414.

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

Chambers, R. G., Chung, Y., and Färe, R. (1996a). Benefit and distance functions. *Journal of economic theory*, 70(2):407–419.

Chambers, R. G., Chung, Y., and Färe, R. (1998). Profit, directional distance functions, and nerlovian efficiency. *Journal of optimization theory and applications*, 98:351–364.

Chambers, R. G., Fāure, R., and Grosskopf, S. (1996b). Productivity growth in apec countries. *Pacific Economic Review*, 1(3):181–190.

Chambers, R. G. and Pope, R. D. (1996). Aggregate productivity measures. *American Journal of Agricultural Economics*, 78(5):1360–1365.

Charles, V., Aparicio, J., and Zhu, J. (2019). The curse of dimensionality of decision-making units: A simple approach to increase the discriminatory power of data envelopment analysis. *European Journal of Operational Research*, 279(3):929–940.

Charnes, A. and Cooper, W. W. (1963). Programming with linear fractional functionals. *Naval Research Logistics Quarterly*, 10(1):273–274.

Charnes, A., Cooper, W. W., Golany, B., Seiford, L., and Stutz, J. (1985). Foundations of data envelopment analysis for pareto-koopmans efficient empirical production functions. *Journal of econometrics*, 30(1-2):91–107.

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

Charnes, A., Cooper, W. W., and Rhodes, E. (1981). Evaluating program and managerial efficiency: an application of data envelopment analysis to program follow through. *Management science*, 27(6):668–697.

Chen, V. C., Ruppert, D., and Shoemaker, C. A. (1999). Applying experimental design and regression splines to high-dimensional continuous-state stochastic dynamic programming. *Operations Research*, 47(1):38–53.

Chen, Y., Tsionas, M. G., and Zelenyuk, V. (2021). Lasso+ dea for small and big wide data. *Omega*, 102:102419.

Coelli, T. J. (1995). Recent developments in frontier modelling and efficiency measurement. *Australian Journal of agricultural economics*, 39(3):219–245.

Cooper, W. W., Park, K. S., and Pastor, J. T. (1999). Ram: a range adjusted measure of inefficiency for use with additive models, and relations to other models and measures in dea. *Journal of Productivity analysis*, 11:5–42.

Cooper, W. W., Pastor, J. T., Borras, F., Aparicio, J., and Pastor, D. (2011). Bam: a bounded adjusted measure of efficiency for use with bounded additive models. *Journal of Productivity analysis*, 35:85–94.

Cooper, W. W., Seiford, L. M., Tone, K., et al. (2007). *Data envelopment analysis: a comprehensive text with models, applications, references and DEA-solver software*, volume 2. Springer.

Cordero, J. M., Cristóbal, V., and Santín, D. (2018). Causal inference on education policies: A survey of empirical studies using pisa, timss and pirls. *Journal of Economic Surveys*, 32(3):878–915.

Croux, C., Gijbels, I., and Prosdocimi, I. (2012). Robust estimation of mean and dispersion functions in extended generalized additive models. *Biometrics*, 68(1):31–44.

Dadura, A. M. and Lee, T.-R. (2011). Measuring the innovation ability of taiwan's food industry using dea. *Innovation: The European Journal of Social Science Research*, 24(1-2):151–172.

Dai, S., Fang, Y.-H., Lee, C.-Y., and Kuosmanen, T. (2024). pystoned: A python package for convex regression and frontier estimation. *Journal of Statistical Software*, 111:1–43.

Daouia, A. and Gijbels, I. (2011). Robustness and inference in nonparametric partial frontier modeling. *Journal of Econometrics*, 161(2):147–165.

Daouia, A., Noh, H., and Park, B. U. (2016). Data envelope fitting with constrained polynomial splines. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 78(1):3–30.

Daouia, A. and Simar, L. (2007). Nonparametric efficiency analysis: A multivariate conditional quantile approach. *Journal of Econometrics*, 140(2):375–400.

Daraio, C. and Simar, L. (2007). *Advanced robust and nonparametric methods in efficiency analysis: Methodology and applications*. Springer Science & Business Media.

Debreu, G. (1951). The coefficient of resource utilization. *Econometrica: Journal of the Econometric Society*, pages 273–292.

Dellnitz, A. (2022). Big data efficiency analysis: Improved algorithms for data envelopment analysis involving large datasets. *Computers & Operations Research*, 137:105553.

Deprins, D., Simar, L., and Tulkens, H. (1984). Measuring labor-efficiency in post offices. In *Public goods, environmental externalities and fiscal competition*, pages 285–309. Springer.

Diewert, W. E. and Parkan, C. (1983). *Linear programming tests of regularity conditions for production functions*. Springer.

Drucker, H., Burges, C. J., Kaufman, L., Smola, A., and Vapnik, V. (1996). Support vector regression machines. *Advances in neural information processing systems*, 9.

Dyson, R. G., Allen, R., Camanho, A. S., Podinovski, V. V., Sarrico, C. S., and Shale, E. A. (2001). Pitfalls and protocols in dea. *European Journal of operational research*, 132(2):245–259.

Efron, B. (1992). Bootstrap methods: another look at the jackknife. In *Breakthroughs in statistics: Methodology and distribution*, pages 569–593. Springer.

Emrouznejad, A. and Anouze, A. L. (2010). Data envelopment analysis with classification and regression tree–a case of banking efficiency. *Expert Systems*, 27(4):231–246.

España, V. J., Aparicio, J., and Barber, X. (2025a). An adaptation of random forest to estimate convex non-parametric production technologies: an empirical illustration of efficiency measurement in education. *International Transactions in Operational Research*, 32(5):2523–2546.

España, V. J., Aparicio, J., and Barber, X. (2025b). Estimating production technologies using multi-output adaptive constrained enveloping splines. *Computers & Operations Research*, page 107242.

España, V. J., Aparicio, J., Barber, X., and Esteve, M. (2024). Estimating production functions through additive models based on regression splines. *European Journal of Operational Research*, 312(2):684–699.

Esteve, C. and Zuazua, E. (2020). The inverse problem for hamilton–jacobi equations and semiconcave envelopes. *SIAM Journal on Mathematical Analysis*, 52(6):5627–5657.

Esteve, M., Aparicio, J., Rabasa, A., and Rodriguez-Sala, J. J. (2020). Efficiency analysis trees: A new methodology for estimating production frontiers through decision trees. *Expert Systems with Applications*, 162:113783.

Esteve, M., Aparicio, J., Rodriguez-Sala, J. J., and Zhu, J. (2023). Random forests and the measurement of super-efficiency in the context of free disposal hull. *European Journal of Operational Research*, 304(2):729–744.

Esteve, M., España, V., Aparicio, J., and Barber, X. (2022). eat: An r package for fitting efficiency analysis trees. *R J.*, 14(3):249–281.

Esteve, M., Rodríguez-Sala, J. J., López-Espín, J. J., and Aparicio, J. (2021). Heuristic and backtracking algorithms for improving the performance of efficiency analysis trees. *IEEE Access*, 9:17421–17428.

Färe, R., Grosskopf, S., and Lovell, C. K. (1985). *The measurement of efficiency of production*, volume 6. Springer Science & Business Media.

Färe, R., Grosskopf, S., and Lovell, C. K. (1994). *Production frontiers*. Cambridge university press.

Färe, R. and Lovell, C. K. (1978). Measuring the technical efficiency of production. *Journal of Economic theory*, 19(1):150–162.

Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the royal statistical society series a: statistics in society*, 120(3):253–281.

Finn, J. D. and Achilles, C. M. (1999). Tennessee's class size study: Findings, implications, misconceptions. *Educational evaluation and policy analysis*, 21(2):97–109.

Flegl, M., Jiménez-Bandala, C. A., Sánchez-Juárez, I., and Matus, E. (2022). Analysis of production and investment efficiency in the mexican food industry: Application of two-stage dea. *Czech Journal of Food Sciences*, 40(2).

Førsund, F. R., Lovell, C. K., and Schmidt, P. (1980). A survey of frontier production functions and of their relationship to efficiency measurement. *Journal of econometrics*, 13(1):5–25.

Friedman, J. H. (1991). Multivariate adaptive regression splines. *The annals of statistics*, 19(1):1–67.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232.

Friedman, L. and Sinuany-Stern, Z. (1998). Combining ranking scales and selecting variables in the dea context: The case of industrial branches. *Computers & Operations Research*, 25(9):781–791.

Golany, B. and Roll, Y. (1989). An application procedure for dea. *Omega*, 17(3):237–250.

Golub, G. H., Heath, M., and Wahba, G. (1979). Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*, 21(2):215–223.

Greene, W. H. et al. (1993). The econometric approach to efficiency analysis. *The measurement of productive efficiency and productivity growth*, 1(1):92–250.

Grifell-Tatjé, E. and Lovell, C. K. (1995). A note on the malmquist productivity index. *Economics letters*, 47(2):169–175.

Griffin, J. E. and Steel, M. F. (2008). Flexible mixture modelling of stochastic frontiers. *Journal of Productivity Analysis*, 29:33–50.

Guerrero, N. M., Aparicio, J., and Valero-Carreras, D. (2022). Combining data envelopment analysis and machine learning. *Mathematics*, 10(6):909.

Guerrero, N. M., Moragues, R., Aparicio, J., and Valero-Carreras, D. (2024). Support vector frontiers with kernel splines. *Omega*, 128:103130.

Guillen, M. D., Aparicio, J., and Esteve, M. (2023a). Gradient tree boosting and the estimation of production frontiers. *Expert Systems with Applications*, 214:119134.

Guillen, M. D., Aparicio, J., and Esteve, M. (2023b). Performance evaluation of decision-making units through boosting methods in the context of free disposal hull: Some exact and heuristic algorithms. *International Journal of Information Technology & Decision Making*, pages 1–30.

Gustafsson, J.-E. (2008). Effects of international comparative studies on educational quality on the quality of educational research. *European Educational Research Journal*, 7(1):1–17.

Guyon, I. and Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of machine learning research*, 3(Mar):1157–1182.

Haas, D. J. (2003). Productive efficiency of english football teams—a data envelopment analysis approach. *Managerial and decision economics*, 24(5):403–410.

Hanushek, E. A. and Woessmann, L. (2011). The economics of international differences in educational achievement. *Handbook of the Economics of Education*, 3:89–200.

Hastie, T., Tibshirani, R., Friedman, J. H., and Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer.

Hernández-Torrano, D. and Courtney, M. G. (2021). Modern international large-scale assessment in education: An integrative review and mapping of the literature. *Large-Scale Assessments in Education*, 9(1):17.

Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67.

Homburg, C. (2001). Using data envelopment analysis to benchmark activities. *International journal of production economics*, 73(1):51–58.

Izadikhah, M., Azadi, M., Toloo, M., and Hussain, F. K. (2021). Sustainably resilient supply chains evaluation in public transport: A fuzzy chance-constrained two-stage dea approach. *Applied soft computing*, 113:107879.

Jain, S., Triantis, K. P., and Liu, S. (2011). Manufacturing performance measurement and target setting: A data envelopment analysis approach. *European Journal of Operational Research*, 214(3):616–626.

Jenkins, L. and Anderson, M. (2003). A multivariate statistical approach to reducing the number of variables in data envelopment analysis. *European journal of operational research*, 147(1):51–61.

Johansson, S. (2016). International large-scale assessments: What uses, what consequences? *Educational Research*, 58(2):139–148.

Johnes, J. (2015). Operational research in education. *European journal of operational research*, 243(3):683–696.

Juo, J.-C., Fu, T.-T., Yu, M.-M., and Lin, Y.-H. (2015). Profit-oriented productivity change. *Omega*, 57:176–187.

Kaffash, S., Azizi, R., Huang, Y., and Zhu, J. (2020). A survey of data envelopment analysis applications in the insurance industry 1993–2018. *European journal of operational research*, 284(3):801–813.

Kamens, D. H. and McNeely, C. L. (2010). Globalization and the growth of international educational testing and national assessment. *Comparative education review*, 54(1):5–25.

Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika*, 30(1-2):81–93.

Koc, E. K. and Bozdogan, H. (2015). Model selection in multivariate adaptive regression splines (mars) using information complexity as the fitness function. *Machine Learning*, 101:35–58.

Koc, E. K. and Iyigun, C. (2014). Restructuring forward step of mars algorithm using a new knot selection procedure based on a mapping approach. *Journal of Global Optimization*, 60:79–102.

Koopmans, T. C. (1951). An analysis of production as an efficient combination of activities. *Activity analysis of production and allocation*.

Korostelëv, A. P., Simar, L., and Tsybakov, A. B. (1995). Efficient estimation of monotone boundaries. *The Annals of Statistics*, pages 476–489.

Korostelev, A. P. and Tsybakov, A. B. (1993). *Minimax theory of image reconstruction*, volume 82. Springer Science & Business Media.

Kumar, M. and Basu, P. (2008). Perspectives of productivity growth in indian food industry: a data envelopment analysis. *International Journal of Productivity and Performance Management*, 57(7):503–522.

Kuosmanen, T. and Johnson, A. (2017). Modeling joint production of multiple outputs in stoned: Directional distance function approach. *European Journal of Operational Research*, 262(2):792–801.

Kuosmanen, T., Johnson, A., and Saastamoinen, A. (2015). Stochastic nonparametric approach to efficiency analysis: A unified framework. In *Data envelopment analysis: A handbook of models and methods*, pages 191–244. Springer.

Kuosmanen, T. and Johnson, A. L. (2010). Data envelopment analysis as nonparametric least-squares regression. *Operations Research*, 58(1):149–160.

Kuosmanen, T. and Kortelainen, M. (2012). Stochastic non-smooth envelopment of data: semi-parametric frontier estimation subject to shape constraints. *Journal of productivity analysis*, 38:11–28.

Kyrgiakos, L. S., Kleftodimos, G., Vlontzos, G., and Pardalos, P. M. (2023). A systematic literature review of data envelopment analysis implementation in agriculture under the prism of sustainability. *Operational Research*, 23(1):7.

LeBlanc, M. and Tibshirani, R. (1996). Combining estimates in regression and classification. *Journal of the American Statistical Association*, 91(436):1641–1650.

Lee, C.-Y. and Cai, J.-Y. (2020). Lasso variable selection in data envelopment analysis with small datasets. *Omega*, 91:102019.

Liu, J. S., Lu, L. Y., Lu, W.-M., and Lin, B. J. (2013). A survey of dea applications. *Omega*, 41(5):893–902.

Louppe, G. (2014). *Understanding random forests: From theory to practice*. PhD thesis, Universite de Liege (Belgium).

Louppe, G. and Geurts, P. (2012). Ensembles on random patches. In *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2012, Bristol, UK, September 24-28, 2012. Proceedings, Part I 23*, pages 346–361. Springer.

Lovell, C. K. (2003). The decomposition of malmquist productivity indexes. *Journal of productivity analysis*, 20:437–458.

Lovell, C. K. et al. (1993). Production frontiers and productive efficiency. *The measurement of productive efficiency: techniques and applications*, 3:67.

Lovell, C. K. and Pastor, J. T. (1995). Units invariant and translation invariant dea models. *Operations research letters*, 18(3):147–151.

Machmud, A., Ahman, E., Dirgantari, P. D., Waspada, I., and Nandiyanto, A. B. D. (2019). Data envelopment analysis: The efficiency study of food industry in indonesia. *Journal of Engineering Science and Technology*, 14(1):479–488.

Malmquist, S. (1953). Index numbers and indifference surfaces. *Trabajos de estadística*, 4(2):209–242.

Martinez, D. L., Shih, D. T., Chen, V. C., and Kim, S. B. (2015). A convex version of multivariate adaptive regression splines. *Computational statistics & data analysis*, 81:89–106.

Md Hamzah, N., Yu, M.-M., and See, K. F. (2021). Assessing the efficiency of malaysia health system in COVID-19 prevention and treatment response. *Health care management science*, 24:273–285.

Meeusen, W. and van Den Broeck, J. (1977). Efficiency estimation from cobb-douglas production functions with composed error. *International economic review*, pages 435–444.

Milborrow, S. (2014). Notes on the earth package. *Retrieved October*, 31:2017.

Milborrow, S., Hastie, T., Tibshirani, R., Miller, A., and Lumley, T. (2017). earth: Multivariate adaptive regression splines. *R package version*, 5(2).

Molinos-Senante, M., Maziotis, A., Sala-Garrido, R., and Mocholi-Arce, M. (2023). Assesing the influence of environmental variables on the performance of water companies: An efficiency analysis tree approach. *Expert Systems with Applications*, 212:118844.

Moragues, R., Aparicio, J., and Esteve, M. (2023a). Measuring technical efficiency for multi-input multi-output production processes through oneclass support vector machines: a finite-sample study. *Operational Research*, 23(3):47.

Moragues, R., Aparicio, J., and Esteve, M. (2023b). An unsupervised learning-based generalization of data envelopment analysis. *Operations Research Perspectives*, 11:100284.

Mosek, A. and Copenhagen, D. (2021). The mosek optimization toolbox for matlab manual. version 9.0., 2019. *URL http://docs. mosek. com/9.0/toolbox/index. html*.

Murat, N. (2023). Outlier detection in statistical modeling via multivariate adaptive regression splines. *Communications in Statistics-Simulation and Computation*, 52(7):3379–3390.

Nataraja, N. R. and Johnson, A. L. (2011). Guidelines for using variable selection techniques in data envelopment analysis. *European Journal of Operational Research*, 215(3):662–669.

Nelder, J. A. and Wedderburn, R. W. (1972). Generalized linear models. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 135(3):370–384.

Nunamaker, T. R. (1983). Measuring routine nursing service efficiency: a comparison of cost per patient day and data envelopment analysis models. *Health services research*, 18(2 Pt 1):183.

Nunamaker, T. R. (1985). Using data envelopment analysis to measure the efficiency of non-profit organizations: A critical evaluation. *Managerial and decision Economics*, 6(1):50–58.

Olesen, O. B. and Petersen, N. C. (2016). Stochastic data envelopment analysis—a review. *European journal of operational research*, 251(1):2–21.

Olesen, O. B. and Ruggiero, J. (2018). An improved afriat–diewert–parkan nonparametric production function estimator. *European Journal of Operational Research*, 264(3):1172–1188.

Olesen, O. B. and Ruggiero, J. (2022). The hinging hyperplanes: An alternative nonparametric representation of a production function. *European Journal of Operational Research*, 296(1):254–266.

Ouenniche, J. and Tone, K. (2017). An out-of-sample evaluation framework for dea with application in bankruptcy prediction. *Annals of Operations Research*, 254:235–250.

Özmen, A., Kropat, E., and Weber, G.-W. (2017). Robust optimization in spline regression models for multi-model regulatory networks under polyhedral uncertainty. *Optimization*, 66(12):2135–2155.

Özmen, A. and Weber, G. W. (2014). Rmars: robustification of multivariate adaptive regression spline under polyhedral uncertainty. *Journal of Computational and Applied Mathematics*, 259:914–924.

Özmen, A., Weber, G. W., Batmaz, İ., and Kropat, E. (2011). Rcmars: Robustification of cmars with different scenarios under polyhedral uncertainty set. *Communications in Nonlinear Science and Numerical Simulation*, 16(12):4780–4787.

Parmeter, C. F. and Racine, J. S. (2013). Smooth constrained frontier analysis. *Recent Advances and Future Directions in Causality, Prediction, and Specification Analysis: Essays in Honor of Halbert L. White Jr*, pages 463–488.

Pastor, J. T., Lovell, C. K., and Aparicio, J. (2012). Families of linear efficiency programs based on debreu's loss function. *Journal of Productivity Analysis*, 38:109–120.

Pastor, J. T., Ruiz, J. L., and Sirvent, I. (1999). An enhanced dea russell graph efficiency measure. *European journal of operational research*, 115(3):596–607.

Pastor, J. T., Ruiz, J. L., and Sirvent, I. (2002). A statistical test for nested radial dea models. *Operations Research*, 50(4):728–735.

Perelman, S. and Santín, D. (2009). How to generate regularly behaved production data? a monte carlo experimentation on dea scale efficiency measurement. *European Journal of Operational Research*, 199(1):303–310.

Peyrache, A., Rose, C., and Sicilia, G. (2020). Variable selection in data envelopment analysis. *European Journal of Operational Research*, 282(2):644–659.

Raab, R. L. and Lichty, R. W. (2002). Identifying subareas that comprise a greater metropolitan area: The criterion of county relative efficiency. *Journal of regional science*, 42(3):579–594.

Ray, S. C. and Desli, E. (1997). Productivity growth, technical progress, and efficiency change in industrialized countries: comment. *The American economic review*, 87(5):1033–1039.

Rebai, S., Yahia, F. B., and Essid, H. (2020). A graphically based machine learning approach to predict secondary schools performance in tunisia. *Socio-Economic Planning Sciences*, 70:100724.

Ruggiero, J. (2005). Impact assessment of input omission on dea. *International Journal of Information Technology & Decision Making*, 4(03):359–368.

Schefczyk, M. (1993). Operational performance of airlines: an extension of traditional measurement paradigms. *Strategic management journal*, 14(4):301–317.

Schmidt, P. (1985). Frontier production functions. *Econometric reviews*, 4(2):289–328.

Schölkopf, B. (2001). Statistical learning and kernel methods. In *Data fusion and perception*, pages 3–24. Springer.

Sexton, T. R., Silkman, R. H., and Hogan, A. J. (1986). Data envelopment analysis: Critique and extensions. *New directions for program evaluation*, 1986(32):73–105.

Shephard, R. W. (1953). Theory of cost and production functions.

Sherman, H. D. and Gold, F. (1985). Bank branch operating efficiency: Evaluation with data envelopment analysis. *Journal of banking & finance*, 9(2):297–315.

Simar, L. and Wilson, P. W. (1998). Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models. *Management science*, 44(1):49–61.

Simar, L. and Wilson, P. W. (2000). A general methodology for bootstrapping in non-parametric frontier models. *Journal of applied statistics*, 27(6):779–802.

Simar, L. and Wilson, P. W. (2001). Testing restrictions in nonparametric efficiency models. *Communications in Statistics-Simulation and Computation*, 30(1):159–184.

Simar, L. and Wilson, P. W. (2013). Estimation and inference in nonparametric frontier models: Recent developments and perspectives. *Foundations and trends® in econometrics*, 5(3–4):183–337.

Simar, L. and Wilson, P. W. (2015). Statistical approaches for non-parametric frontier models: a guided tour. *International Statistical Review*, 83(1):77–110.

Simar, L. and Zelenyuk, V. (2006). On testing equality of distributions of technical efficiency scores. *Econometric Reviews*, 25(4):497–522.

Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of educational research*, 75(3):417–453.

Smith, P. (1997). Model misspecification in data envelopment analysis. *Annals of Operations Research*, 73(0):233–252.

Spearman, C. (1961). The proof and measurement of association between two things.

Steiner-Khamsi, G. and Waldow, F. (2018). Pisa for scandalisation, pisa for projection: the use of international large-scale assessments in education policy making–an introduction. *Globalisation, Societies and Education*, 16(5):557–565.

Sueyoshi, T., Yuan, Y., and Goto, M. (2017). A literature study for dea applied to energy and environment. *Energy Economics*, 62:104–124.

Taylan, P., Weber, G.-W., Liu, L., and Yerlikaya-Özkurt, F. (2010). On the foundations of parameter estimation for generalized partial linear models with b-splines and continuous optimization. *Computers & Mathematics with Applications*, 60(1):134–143.

Theussl, S. (2009). R/gnu linear programming kit interface. *http://R-Forge. R-project. org/projects/rglp/*.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58(1):267–288.

Tone, K. (2001). A slacks-based measure of efficiency in data envelopment analysis. *European journal of operational research*, 130(3):498–509.

Tsai, J. C. and Chen, V. C. (2005). Flexible and robust implementations of multivariate adaptive regression splines within a wastewater treatment stochastic dynamic program. *Quality and Reliability Engineering International*, 21(7):689–699.

Tsionas, M. G. (2022). Efficiency estimation using probabilistic regression trees with an application to chilean manufacturing industries. *International Journal of Production Economics*, page 108492.

Tsolas, I. E., Charles, V., and Gherman, T. (2020). Supporting better practice benchmarking: A dea-ann approach to bank branch performance assessment. *Expert Systems with Applications*, 160:113599.

Ueda, T. and Hoshiai, Y. (1997). Application of principal component analysis for parsimonious summarization of dea inputs and/or outputs. *Journal of the Operations Research society of Japan*, 40(4):466–478.

Valero-Carreras, D., Aparicio, J., and Guerrero, N. M. (2021). Support vector frontiers: A new approach for estimating production functions through support vector machines. *Omega*, 104:102490.

Valero-Carreras, D., Aparicio, J., and Guerrero, N. M. (2022). Multi-output support vector frontiers. *Computers & Operations Research*, 143:105765.

Vapnik, V. N. (1999). An overview of statistical learning theory. *IEEE transactions on neural networks*, 10(5):988–999.

Verger, A., Parcerisa, L., and Fontdevila, C. (2019). The growth and spread of large-scale assessments and test-based accountabilities: A political sociology of global education reforms. *Educational Review*, 71(1):5–30.

Wagner, J. M. and Shimshak, D. G. (2007). Stepwise selection of variables in data envelopment analysis: Procedures and managerial perspectives. *European journal of operational research*, 180(1):57–67.

Wang, H.-J. and Schmidt, P. (2002). One-step and two-step estimation of the effects of exogenous variables on technical efficiency levels. *journal of Productivity Analysis*, 18:129–144.

Weber, G.-W., Batmaz, I., Köksal, G., Taylan, P., and Yerlikaya-Özkurt, F. (2012). Cmars: a new contribution to nonparametric regression with multivariate adaptive regression splines supported by continuous optimization. *Inverse Problems in Science and Engineering*, 20(3):371–400.

Wilson, P. W. (2018). Dimension reduction in nonparametric models of production. *European Journal of Operational Research*, 267(1):349–367.

Witte, K. D. and López-Torres, L. (2017). Efficiency in education: A review of literature and a way forward. *Journal of the operational research society*, 68(4):339–363.

Worthington, A. C. (2001). An empirical survey of frontier efficiency measurement techniques in education. *Education economics*, 9(3):245–268.

Yazıcı, C., Yerlikaya-Özkurt, F., and Batmaz, I. (2015). A computational approach to nonparametric regression: bootstrapping cmars method. *Machine Learning*, 101:211–230.

Zhang, H. (1994). Maximal correlation and adaptive splines. *Technometrics*, 36(2):196–201.

Zhang, H. and Singer, B. H. (2010). *Recursive partitioning and applications*. Springer Science & Business Media.

Zhou, H., Yang, Y., Chen, Y., and Zhu, J. (2018). Data envelopment analysis application in sustainability: The origins, development and future directions. *European journal of operational research*, 264(1):1–16.

Zhu, J. (2022). Dea under big data: Data enabled analytics and network data envelopment analysis. *Annals of Operations Research*, 309(2):761–783.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 67(2):301–320.